Label Consistent K-SVD for Sparse Micro-Doppler Classification

Fraser Coutts*, Domenico Gaglione*, Carmine Clemente*, Gang Li†, Ian K. Proudler‡, John J. Soraghan*
*University of Strathclyde, CeSIP, EEE, 204 George Street, G1 1XW, Glasgow, UK
E-mails: fraser.coutts.2013@uni.strath.ac.uk, domenico.gaglione@strath.ac.uk, carmine.clemente@strath.ac.uk, j.soraghan@strath.ac.uk
†Department of Electronic Engineering, Tsinghua University, Beijing, China
E-mail: gangli@tsinghua.edu.cn
‡School of Electronic, Electrical and Systems Engineering, Loughborough University, Leicestershire, UK
E-mail: i.k.proudler@lboro.ac.uk

***Abstract*—Secondary motions of targets observed by radar introduce non-stationary returns containing the so-called micro-Doppler information. This is characterizing information that can be exploited to enhance automatic target recognition systems. In this paper, the challenge of classifying the micro-Doppler return of helicopters is addressed. For this specific family of targets the received signal is characterized by a high degree of sparseness. A robust dictionary learning algorithm, Label Consistent K-SVD (LC-KSVD), is applied to identify effectively and efficiently helicopters. The effectiveness of the proposed algorithm is demonstrated on both synthetic and real radar data.**

I. Introduction

The Doppler effect [1] is commonly experienced when an object emitting a sound, such as a car, moves relative to an observer. As it approaches, the sound emitted by the object is observed as higher in frequency than the sound observed when the object moves away; this is due to the effective compression and stretching of the sound waves in front of and behind the object, respectively.

Similar Doppler shifts are observed when using Radio Detection and Ranging (RADAR) systems to investigate a moving object. The main motion of this object determines its predominant Doppler signature, which can be used to measure its bulk velocity, while any secondary motions, such as the rotation of rotor blades of a helicopter, contribute with features known as micro-Doppler (mD) signatures [2]. Such signatures appear superimposed on the object's main Doppler contribution, and can be used to identify (classify) different targets.

In the last decade several mD-based radar automatic target recognition techniques have been presented [3]. A common way to extract the mD information is to use time frequency analyis. In [4] the authors developed a mD features extraction approach which derives from the combination of the Short Time Fourier Transform (STFT) and the Wigner Distribution; in [5] the STFT was used in conjunction with the pseudo-Zernike moments in order to classify different human targets. A real time demonstrator was described in [6] where a simple STFT based feature was extracted. However, these methods present a relative high computational cost due to the computation of a time-frequency distribution and depend on the choice of the parameters of the distribution itself (i.e. window length), which, in turn, depend on the dynamic of the target.

The capability to classify a helicopter by analysing its mD properties was first investigated in [7], after that in [8] was demonstrated that the theoretical return signal from propeller blades depend on the number, the length and the rotation speed of the blades themselves. In [9] was demonstrated that even a passive bistatic radar (PBR) is able to record the mD signature of a helicopter. A mD features extraction algorithm from helicopter return signal was presented in [10], which is still based on the computation of the STFT, while in [11] a model-based classification algorithm was introduced, enabling high accuracy with low computational cost and exploiting the parameter estimation approach introduced in [12].

Dictionary learning for sparse representation has produced promising results in the field of image processing [13], [14], [15]. In this paper, a robust method for classifying the mD signatures of helicopters via dictionary learning is presented. The proposed technique utilises the Label Consistent K-SVD (LC-KSVD) algorithm [16] and does not need the computation of any time-frequency representation; thus, it is independent of the received signal, since no parameters have to be adapted to the input signal, such as the window length.

The reminder of the paper is organised as follows: Section II describes the LC-KSVD algorithm for dictionary learning and how it is applied in the mD automatic target recognition context, while in Section III the performance of the proposed algorithm on both simulated and real data are analysed. Section IV concludes the paper.

II. Label Consistent K-SVD for micro-Doppler Classification

A. Sparsity Constraint Using ℓ_0-Norm

Let $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_N] \in \mathbb{C}^{n \times N}$ be a set of N n-dimensional input signals. An over-complete reconstructive dictionary, $\mathbf{D} = [\mathbf{d}_1, \ldots, \mathbf{d}_K] \in \mathbb{C}^{n \times K}$ (where $K > n$), can be learned for the sparse reconstruction of $\mathbf{Y}$ by solving the following problem:

$$\langle \mathbf{D}, \mathbf{X} \rangle = \underset{\mathbf{D},\mathbf{X}}{\arg\min} \|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_2^2 \quad \text{s.t.} \quad \forall i, \|\mathbf{x}_i\|_0 \leq M \tag{1}$$

where $\mathbf{X} = [\mathbf{x}_1, \ldots\ \mathbf{x}_N] \in \mathbb{C}^{K \times N}$ are the sparse codes of the input signals $\mathbf{Y}$, and $\|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_2^2$ denotes the reconstruction error. M is a sparsity constraint factor; thus, each sparse code has fewer than M non-zero values. For a given $\mathbf{D}$, a sparse code $\mathbf{x}_i$ of $\mathbf{y}_i$ can be found by solving:

$$\mathbf{x}_i = \mathbf{x}^*(\mathbf{y}_i, \mathbf{D}) = \arg\min_{\mathbf{x}} \|\mathbf{y}_i - \mathbf{D}\mathbf{x}\|_2^2 \quad \text{s.t.} \quad \|\mathbf{x}\|_0 \leq M. \tag{2}$$

The K-SVD algorithm [17] comprises an iterative approach to efficiently minimise the energy in (1), while learning a reconstructive dictionary for the sparse decomposition of signals. The orthogonal matching pursuit algorithm (OMP) [18] or the pruned OMP (POM) [12] can be used to solve (2).

B. Sparsity Constraint Using ℓ_1-Norm

Very often, an alternative approach to (1) using ℓ_1-norm regularisation can be employed to enforce sparsity:

$$\langle \mathbf{D}, \mathbf{X} \rangle = \arg\min_{\mathbf{D},\mathbf{X}} \left[\|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_2^2 + \gamma \|\mathbf{X}\|_1 \right], \tag{3}$$

where γ is a parameter to balance the reconstruction error and sparsity. Similarly, given $\mathbf{D}$, a sparse code $\mathbf{x}_i$ of an input signal $\mathbf{y}_i$ can be established as follows:

$$\mathbf{x}_i = \mathbf{x}^*(\mathbf{y}_i, \mathbf{D}) = \arg\min_{\mathbf{x}} \left[\|\mathbf{y}_i - \mathbf{D}\mathbf{x}\|_2^2 + \gamma \|\mathbf{x}\|_1 \right], \tag{4}$$

which is suitably optimised by a number of efficient ℓ_1-optimisation approaches, such as [19], [20].

C. Label Consistent K-SVD

This algorithm aims to enforce the labels of the input signals to learn a reconstructive and discriminative dictionary. Each dictionary item $\mathbf{d}_k$ is chosen such that, in the ideal case, it represents a subset of the training signals from a single class; thus, each dictionary item can be associated with a particular label. The algorithm incorporates both a joint classification error and label consistency regularisation term into the objective function of (1) for the learning of a dictionary with reconstructive and discriminative capabilities.

In LC-KSVD, a linear predictive classifier $f(\mathbf{x}; \mathbf{W}) = \mathbf{W}\mathbf{x}$ is used. The discriminability of the input sparse codes $\mathbf{X}$ has a positive impact on the performance of such a linear classifier; therefore, encouraging discriminability during dictionary learning is of interest. Furthermore, incorporating the training of such a classifier into the objective function used for dictionary learning can simultaneously make the dictionary optimal for classification. The algorithm harnesses the above knowledge within an objective function for learning a dictionary $\mathbf{D}$ with both reconstructive and discriminative power; this is defined as follows:

$$\begin{aligned}\langle \mathbf{D}, \mathbf{W}, \mathbf{A}, \mathbf{X} \rangle = \arg\min_{\mathbf{D},\mathbf{W},\mathbf{A},\mathbf{X}} \Big[&\|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_2^2 + \alpha \|\mathbf{Q} - \mathbf{A}\mathbf{X}\|_2^2 + \\ &+ \beta \|\mathbf{H} - \mathbf{W}\mathbf{X}\|_2^2 \Big] \quad \text{s.t.} \quad \forall i, \|\mathbf{x}_i\|_0 \leq M,\end{aligned} \tag{5}$$

where the terms $\|\mathbf{Q} - \mathbf{A}\mathbf{X}\|_2^2$ and $\|\mathbf{H} - \mathbf{W}\mathbf{X}\|_2^2$ represent the discriminative sparse-code error and the classification error, respectively. The former ensures that the transformed sparse codes $\mathbf{A}\mathbf{X}$ (where $\mathbf{A} \in \mathbb{C}^{K \times K}$ is a linear transformation matrix) approximate the discriminative sparse codes $\mathbf{Q} = [\mathbf{q}_1, \ldots, \mathbf{q}_N] \in \mathbb{C}^{K \times N}$, while the latter trains the classifier parameters $\mathbf{W} \in \mathbb{C}^{L \times K}$, where L is the number of categories, to recover the class labels $\mathbf{H} = [\mathbf{h}_1, \ldots, \mathbf{h}_N] \in \mathbb{C}^{L \times N}$ of the input signals $\mathbf{Y}$. Thus, signals from the same class are encouraged to have very similar sparse representations, which results in good classification performance when using the trained linear classifier. Parameters α and β are the scalars controlling the relative contribution of the corresponding terms.

The vector $\mathbf{q}_i = \left[q_i^1, \ldots, q_i^K\right]^T \in \mathbb{R}^K$ is deemed a discriminative sparse representation of an input signal $\mathbf{y}_i$ if its non-zero values occur at those indices where the input signal $\mathbf{y}_i$ and the dictionary item $\mathbf{d}_k$ share the same label. In addition, $\mathbf{h}_i \in \mathbb{R}^L$ is a label vector corresponding to an input signal $\mathbf{y}_i$, where the class of $\mathbf{y}_i$ is defined by $\mathbf{h}_i$'s non-zero position.

The K-SVD algorithm is used to find the optimal solution for all parameters simultaneously. If equation (5) is written as follows:

$$\begin{aligned}\langle \mathbf{D}, \mathbf{W}, \mathbf{A}, \mathbf{X} \rangle = \arg\min_{\mathbf{D},\mathbf{W},\mathbf{A},\mathbf{X}} &\left\| \begin{pmatrix} \mathbf{Y} \\ \sqrt{\alpha}\mathbf{Q} \\ \sqrt{\beta}\mathbf{H} \end{pmatrix} - \begin{pmatrix} \mathbf{D} \\ \sqrt{\alpha}\mathbf{A} \\ \sqrt{\beta}\mathbf{W} \end{pmatrix} \mathbf{X} \right\|_2^2 \\ &\text{s.t.} \quad \forall i, \|\mathbf{x}_i\|_0 \leq M,\end{aligned} \tag{6}$$

defining $\mathbf{Y}_{new} = \left(\mathbf{Y}^T, \sqrt{\alpha}\mathbf{Q}^T, \sqrt{\beta}\mathbf{H}^T\right)^T$, and $\mathbf{D}_{new} = \left(\mathbf{D}^T, \sqrt{\alpha}\mathbf{A}^T, \sqrt{\beta}\mathbf{W}^T\right)^T$, where $\mathbf{D}_{new}$ is ℓ_2-normalised columnwise, allows for the optimisation of (6) to be rewritten as:

$$\begin{aligned}\langle \mathbf{D}_{new}, \mathbf{X} \rangle = \arg\min_{\mathbf{D}_{new},\mathbf{X}} &\|\mathbf{Y}_{new} - \mathbf{D}_{new}\mathbf{X}\|_2^2 \\ &\text{s.t.} \quad \forall i, \|\mathbf{x}_i\|_0 \leq M.\end{aligned} \tag{7}$$

This is exactly the problem that K-SVD [17] solves. Following the application of K-SVD, $\mathbf{d}_k$ and its corresponding coefficients, contained within the k-th row in $\mathbf{X}$, denoted as $\mathbf{x}_k^R$, are updated. Let

$$\mathbf{E}_k = \left(\mathbf{Y} - \sum_{j \neq k} \mathbf{d}_j \mathbf{x}_j^R \right), \tag{8}$$

and $\tilde{\mathbf{x}}_k^R$ and $\tilde{\mathbf{E}}_k$ denote the result of discarding the zero entries in $\mathbf{x}_k^R$ and $\mathbf{E}_k$, respectively. $\mathbf{d}_k$ and $\tilde{\mathbf{x}}_k^R$ can be computed by

$$\left\langle \mathbf{d}_k, \tilde{\mathbf{x}}_k^R \right\rangle = \arg\min_{\mathbf{d}_k, \tilde{\mathbf{x}}_k^R} \left\| \tilde{\mathbf{E}}_k - \mathbf{d}_k \tilde{\mathbf{x}}_k^R \right\|_2^2. \tag{9}$$

Following computation of the SVD for $\tilde{\mathbf{E}}_k$ (i.e., $\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T = \text{SVD}(\tilde{\mathbf{E}}_k)$), $\mathbf{d}_k$ and $\tilde{\mathbf{x}}_k^R$ are computed as

$$\mathbf{d}_k = \mathbf{U}\,(:,1)\,, \quad \tilde{\mathbf{x}}_k^R = \mathbf{\Sigma}(1,1)V(:,1) \tag{10}$$

Finally, the non-zero values in $\mathbf{x}_k^R$ are replaced by $\tilde{\mathbf{x}}_k^R$.

The ability of LC-KSVD to learn $\mathbf{D}$, $\mathbf{A}$, and $\mathbf{W}$ simultaneously is scalable to a large number of classes, and reduces the possibility of converging to a local minima. Furthermore, it enables the use of discriminative sparse-code error in the objective function; thus, LC-KSVD can produce a discriminative sparse representation irrespective of the size of the dictionary.

D. LC-KSVD Algorithm

As mentioned in the LC-KSVD algorithm description, the parameters $\mathbf{D}^{(0)}$, $\mathbf{A}^{(0)}$, and $\mathbf{W}^{(0)}$ must be initialised.

For the initial dictionary, $\mathbf{D}^{(0)}$, several iterations of K-SVD are applied within each class, and all of the outputs produced are combined. Each dictionary item $\mathbf{d}_k$ is then given a label based on the class it corresponds to; this label will remain fixed throughout the dictionary learning process, although $\mathbf{d}_k$ itself is updated during the learning process. While a distinct and fixed-class label is associated with each dictionary item, it is possible for an input signal of a specific class to use dictionary items corresponding to other classes. The algorithm uniformly allocates dictionary elements to each class; the number of elements allocated is proportional to the dictionary size.

Given the initialised dictionary, $\mathbf{D}^{(0)}$, the original K-SVD algorithm is employed to compute the sparse codes $\mathbf{X}$ of the training signals $\mathbf{Y}$.

To initialise $\mathbf{A}^{(0)}$, the technique of multivariate ridge regression [21] is used, with the quadratic loss and ℓ_2-norm regularisation, as follows:

$$\mathbf{A} = \arg\min_{\mathbf{A}} \|\mathbf{Q} - \mathbf{A}\mathbf{X}\|^2 + \lambda_1 \|\mathbf{A}\|_2^2. \tag{11}$$

This yields the following solution:

$$\mathbf{A} = \mathbf{Q}\mathbf{X}^T \left(\mathbf{X}\mathbf{X}^T + \lambda_1 \mathbf{I}\right)^{-1}. \tag{12}$$

Similarly, for $\mathbf{W}^{(0)}$, the ridge regression model is used to obtain the following solution:

$$\mathbf{W} = \mathbf{H}\mathbf{X}^T \left(\mathbf{X}\mathbf{X}^T + \lambda_2 \mathbf{I}\right)^{-1}. \tag{13}$$

where λ_1 and λ_2 are regularisation parameters.

The LC-KSVD algorithm is summarised in Algorithm 1.

Algorithm 1 Label Consistent K-SVD.

Input $\mathbf{Y}$, $\mathbf{Q}$, $\mathbf{H}$, α, β, M, K

Output $\mathbf{D}$, $\mathbf{A}$, $\mathbf{W}$

Compute $\mathbf{D}^{(0)}$, $\mathbf{A}^{(0)}$, $\mathbf{W}^{(0)}$: compute $\mathbf{D}^{(0)}$ by combining class-specific dictionary items for each class using original K-SVD [17]; compute the sparse codes $\mathbf{X}^{(0)}$ for $\mathbf{Y}$ by using (2); compute $\mathbf{A}^{(0)}$ and $\mathbf{W}^{(0)}$ using (12) and (13).

Initialise $\mathbf{Y}_{new} = \begin{pmatrix} \mathbf{Y} \\ \sqrt{\alpha}\mathbf{Q} \\ \sqrt{\beta}\mathbf{H} \end{pmatrix}$, $\mathbf{D}_{new} = \begin{pmatrix} \mathbf{D}^{(0)} \\ \sqrt{\alpha}\mathbf{A}^{(0)} \\ \sqrt{\beta}\mathbf{W}^{(0)} \end{pmatrix}$

Update $\mathbf{D}_{new}$ by solving (7) using original K-SVD [17];

Obtain $\mathbf{D}$, $\mathbf{A}$, $\mathbf{W}$ from $\mathbf{D}_{new}$ by using (14).

The parameters α and β are fixed for each data set and determined by n-fold cross validation on the training data. The feature descriptors used are random vectors, whereby each original radar return signal is projected onto an n-dimensional feature vector with a randomly generated matrix from a zero-mean normal distribution. Each row of the random matrix is ℓ_2-normalised.

E. Classification Approach

We obtain $\mathbf{D} = \{\mathbf{d}_1, \ldots, \mathbf{d}_K\}$, $\mathbf{A} = \{\mathbf{a}_1, \ldots, \mathbf{a}_K\}$ and $\mathbf{W} = \{\mathbf{w}_1, \ldots, \mathbf{w}_K\}$ from $\mathbf{D}_{new}$ by employing the K-SVD algorithm [17]. $\mathbf{D}$, $\mathbf{A}$ and $\mathbf{W}$ are ℓ_2-normalised in $\mathbf{D}_{new}$; i.e., $\forall k, \left\|\left(\mathbf{d}_k^T, \sqrt{\alpha}\mathbf{a}_k^T, \sqrt{\beta}\mathbf{w}_k^T\right)^T\right\|_2 = 1$. Thus, the desired dictionary $\hat{\mathbf{D}}$, transform parameters $\hat{\mathbf{A}}$, and classifier parameters $\hat{\mathbf{W}}$ are computed as follows:

$$\hat{\mathbf{D}} = \left\{\frac{\mathbf{d}_1}{\|\mathbf{d}_1\|_2}, \ldots, \frac{\mathbf{d}_K}{\|\mathbf{d}_K\|_2}\right\}, \quad \hat{\mathbf{A}} = \left\{\frac{\mathbf{a}_1}{\|\mathbf{a}_1\|_2}, \ldots, \frac{\mathbf{a}_K}{\|\mathbf{a}_K\|_2}\right\},$$
$$\hat{\mathbf{W}} = \left\{\frac{\mathbf{w}_1}{\|\mathbf{w}_1\|_2}, \ldots, \frac{\mathbf{w}_K}{\|\mathbf{w}_K\|_2}\right\}. \tag{14}$$

Hence, for an input signal $\mathbf{y}_i$, which has a discriminative code $\mathbf{q}_i$ and a label vector $\mathbf{h}_i$, the relationship between the desired $\left(\hat{\mathbf{D}}, \hat{\mathbf{A}}, \hat{\mathbf{W}}\right)$ and the learned $(\mathbf{D}, \mathbf{A}, \mathbf{W})$ parameters is established according to

$$\begin{aligned}\mathbf{y}_i \simeq \mathbf{D}\mathbf{x}_i = \sum_k \mathbf{x}_{k,i}\mathbf{d}_k = \sum_k \mathbf{x}_{k,i}\|\mathbf{d}_k\|_2 \frac{\mathbf{d}_k}{\|\mathbf{d}_k\|_2} = \\ \sum_k \hat{\mathbf{x}}_{k,i}\hat{\mathbf{d}}_k = \hat{\mathbf{D}}\hat{\mathbf{x}}_i\end{aligned} \tag{15}$$

$$\begin{aligned}\mathbf{q}_i \simeq \mathbf{A}\mathbf{x}_i = \sum_k \mathbf{x}_{k,i}\mathbf{a}_k = \sum_k \mathbf{x}_{k,i}\|\mathbf{d}_k\|_2 \frac{\mathbf{a}_k}{\|\mathbf{d}_k\|_2} = \\ \sum_k \hat{\mathbf{x}}_{k,i}\hat{\mathbf{a}}_k = \hat{\mathbf{A}}\hat{\mathbf{x}}_i\end{aligned} \tag{16}$$

$$\begin{aligned}\mathbf{h}_i \simeq \mathbf{W}\mathbf{x}_i = \sum_k \mathbf{x}_{k,i}\mathbf{w}_k = \sum_k \mathbf{x}_{k,i}\|\mathbf{d}_k\|_2 \frac{\mathbf{w}_k}{\|\mathbf{d}_k\|_2} = \\ \sum_k \hat{\mathbf{x}}_{k,i}\hat{\mathbf{w}}_k = \hat{\mathbf{W}}\hat{\mathbf{x}}_i\end{aligned} \tag{17}$$

where $\hat{\mathbf{d}}_k = \frac{\mathbf{d}_k}{\|\mathbf{d}_k\|_2}$, $\hat{\mathbf{a}}_k = \frac{\mathbf{a}_k}{\|\mathbf{d}_k\|_2}$ and $\hat{\mathbf{w}}_k = \frac{\mathbf{w}_k}{\|\mathbf{d}_k\|_2}$ are the k-th column of $\mathbf{D}$, $\mathbf{A}$ and $\mathbf{W}$, respectively.

Eventually, given a test signal $\mathbf{s}$, its sparse representation $\hat{\mathbf{x}}$ is obtained by using dictionary $\hat{\mathbf{D}}$ to solve (2) or (4). Subsequently, the trained linear predictive classifier is employed to estimate a label vector $\mathbf{p} = \hat{\mathbf{W}}\hat{\mathbf{x}}$. The resulting class which $\mathbf{s}$ belongs to, is given by the index of the largest element of $\mathbf{p}$.

III. Performance Analysis

The performance analysis is performed on both simulated and real data, for all the presented analysis in this section the results are obtained averaging over 5 runs, where each run used a newly trained dictionary, different randomly selected training/testing vectors, and a different sensing matrix for feature extraction.

The synthetic data were generated according to the model provided in [2]. In Table I the eight helicopters models and parameters are reported, these can be used alongside other variables such as the helicopter elevation angle and initial phase to generate synthetic mD signatures for each model. A carrier frequency of 1.5 GHz was used, and a signal duration of

Table I. Main rotor features of typical helicopters [2]

Name	# of Blades	Diameter (m)	Rotation Rate (r/s)
AH-1 Cobra	2	7.32	4.9
AH-64 Apache	4	7.32	4.8
UH-60 Black Hawk	4	8.18	4.3
CH-53 Stallion	7	12.04	2.9
MD-500E Defender	5	4.03	8.2
A109 Agusta	4	5.50	6.4
AS332 Super Puma	4	7.80	4.4
SA365 Dauphin	4	5.97	5.8

0.345 seconds was selected to accommodate for one revolution of slowest helicopter's rotor blades. The sampling rate was kept at 5 kHz to avoid aliasing in the mD signatures. In the generation of radar returns for training and testing purposes, the distance between the helicopter and receiver was assumed to be known and constant at 500 metres. The elevation was varied between zero and 90 degrees, and the initial phase was varied over the range $-\pi$ to π; both parameters were selected randomly. The rotation rate of rotor blades were randomly distributed over $\pm 1\%$ of values given in Table I, and blade lengths were as given in the same table. The used length of the feature vectors, n, was 1024, the sparsity threshold $M = 1$, $\sqrt{\alpha} = 0.1$ and $\sqrt{\beta} = 1$; 512 training vectors and 73 test vectors were used for each class. The analysis is carried out by varying dictionary size K, which is either 1024 or 4096. Figure 1 shows the proposed classification method's performance when classifying synthetic data on varying the SNR, assuming to be in a AWGN scenario. As expected, the

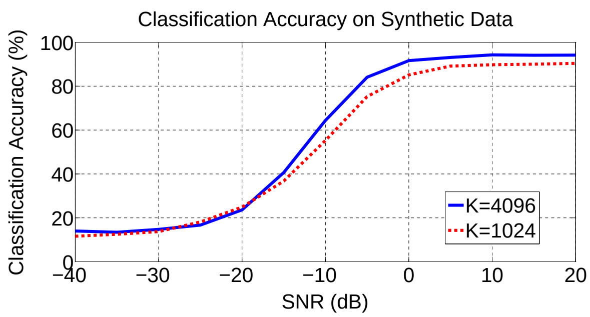

Figure 1. Classification accuracy on synthetic data on varying the SNR. The blue straight line refers to the case in which the dictionary size K is 4096, whereas for the red dashed line $K = 1024$.

algorithm performs increasingly well as the SNR increases; at SNR $= -5dB$ the percentage of correct classification is around 80% for both the dictionary sizes. It reaches 94.25% and 89.75% for $K = 4096$ and $K = 1024$, respectively, showing an overall better performance when the dictionary is larger.

The validity of the approach is proved with real data. Signals from the two-bladed helicopter scale model GAUI X3 are acquired with a 24 GHz radar in CW mode and sampling frequency of 22 kHz. Four rotating speed for the rotor of model helicopter have been chosen, whose actual values are reported in Table II. Three acquisitions of 30 seconds are made

Table II. SCALE MODEL'S ROTATION SPEED

Target	Average Speed
A	6.72 rps
B	9.12 rps
C	12.42 rps
D	13.32 rps

for each speed, at three different aspect angles, 0, 30 and 60 degrees. The duration of each sample considered for the analysis is of 0.1 seconds. Moreover, the algorithm was set up using feature vectors of length 64, a dictionary size of 896, sparsity threshold of 10, 400 training vectors, 400 test vectors, $\sqrt{\alpha} = 0.1$, and $\sqrt{\beta} = 0.1$. Figure 2 shows the confusion matrix of the proposed classification method on real data. For this case an average accuracy of 90.1875% is obtained with a maximum for target A of 93% and minimum for target D of 84.25%.

Last analysis investigates the influence of both dictionary

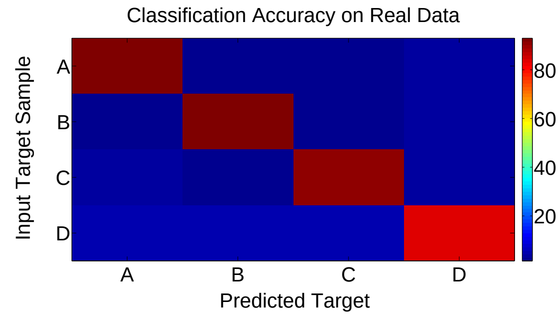

Figure 2. Confusion matrix obtained by applying the algorithm on real data.

size and feature vector length on the classification accuracy by varying these two parameters and providing the percentage of correct classification for the real data. This last results demon-

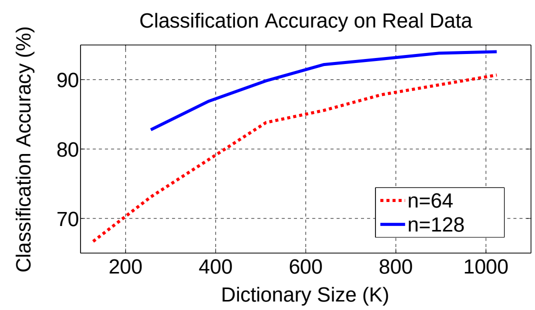

Figure 3. Classification accuracy on real data on varying the dictionary size, K, and the feature vector length, n.

strate that by increasing the number of features, better accuracy can be obtained. However, the drawback is an increase in the computational cost.

IV. CONCLUSION

In this paper a dictionary learning approach for automatic recognition of helicopters based on micro-Doppler information has been presented. The Label Consistent K-SVD algorithm has been used to learn the dictionary that is then used to classify the helicopters. The accuracy of the proposed method has been confirmed using both simulated and real data. Beside the high classification accuracy, the main advantage of the proposed algorithm is the low computational cost.

In the future, the developed techniques may be applied to non-rigid bodies micro-Doppler signals like those received from human targets.

ACKNOWLEDGMENT

This work was supported by the Engineering and PhysicalSciences Research Council (EPSRC) Grant number EP/K014307/1 and the MOD University Defence Research Collaboration in Signal Processing.

Gang Lis work was supported in part by the National Natural Science Foundation of China under Grants 61422110 and 41271011, and in part by the Tsinghua University Initiative Scientific Research Program.

REFERENCES

[1] T. P. Gill, *The Doppler Effect*, 1965.

[2] V. Chen, *The micro-Doppler effect in radar*. Artech House, 2011.

[3] C. Clemente, A. Balleri, K. Woodbridge, and 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, 2013.

[4] T. Thayaparan, L. Stankovi, and I. Djurovi, "Micro-Doppler-based target detection and feature extraction in indoor and outdoor environments," *Journal of the Franklin Institute*, vol. 345, no. 6, pp. 700 – 722, 2008, advances in Indoor Radar Imaging.

[5] L. Pallotta, C. Clemente, A. De Maio, J. Soraghan, and A. Farina, "Pseudo-Zernike moments based radar micro-Doppler classification," in *Radar Conference, 2014 IEEE*, May 2014, pp. 0850–0854.

[6] J. Zabalza, C. Clemente, G. Di Caterina, J. Ren, J. Soraghan, and S. Marshall, "Robust PCA micro-doppler classification using SVM on embedded systems," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 50, no. 3, pp. 2304–2310, July 2014.

[7] J. Misiurewicz, "Analysis of recorded helicopter echo," *IET Conference Proceedings*, pp. 449–453(4), January 1997.

[8] J. Martin and B. Mulgrew, "Analysis of the theoretical radar return signal form aircraft propeller blades," in *Radar Conference, 1990., Record of the IEEE 1990 International*, May 1990, pp. 569–572.

[9] C. Clemente and J. Soraghan, "GNSS-Based Passive Bistatic Radar for Micro-Doppler Analysis of Helicopter Rotor Blades," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 50, no. 1, pp. 491–500, January 2014.

[10] T. Thayaparan, "Analysis of radar micro-Doppler signatures from experimental helicopter and human data," *IET Radar, Sonar & Navigation*, vol. 1, pp. 289–299(10), August 2007.

[11] D. Gaglione, C. Clemente, F. Coutts, L. Gang, and J. Soraghan, "Model-Based Sparse Recovery Method for Automatic Classification of Helicopters," in *International Radar Conference, 2015 IEEE*, May 2015.

[12] G. Li and P. Varshney, "Micro-Doppler Parameter Estimation via Parametric Sparse Representation and Pruned Orthogonal Matching Pursuit," *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, vol. 7, no. 12, pp. 4937–4948, Dec 2014.

[13] J. Yang, K. Yu, Y. Gong, and T. Huang, "Linear spatial pyramid matching using sparse coding for image classification," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, June 2009, pp. 1794–1801.

[14] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma, "Robust Face Recognition via Sparse Representation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 31, no. 2, pp. 210–227, Feb 2009.

[15] D. M. Bradley and J. A. Bagnell, "Differential sparse coding," 2008.

[16] Z. Jiang, Z. Lin, and L. Davis, "Label Consistent K-SVD: Learning a Discriminative Dictionary for Recognition," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 35, no. 11, pp. 2651–2664, Nov 2013.

[17] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation," *Signal Processing, IEEE Transactions on*, vol. 54, no. 11, pp. 4311–4322, Nov 2006.

[18] J. Tropp and A. Gilbert, "Signal Recovery From Random Measurements Via Orthogonal Matching Pursuit," *Information Theory, IEEE Transactions on*, vol. 53, no. 12, pp. 4655–4666, Dec 2007.

[19] H. Lee, A. Battle, R. Raina, and A. Y. Ng, "Efficient sparse coding algorithms," in *Advances in neural information processing systems*, 2006, pp. 801–808.

[20] K. Gregor and Y. LeCun, "Learning fast approximations of sparse coding," in *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, 2010, pp. 399–406.

[21] G. H. Golub, P. C. Hansen, and D. P. O'Leary, "Tikhonov Regularization and Total Least Squares," *SIAM Journal on Matrix Analysis and Applications*, vol. 21, no. 1, pp. 185–194, 1999.