Proceedings of the Asia-Pacific Microwave Conference 2016

Classification of Multiple Targets Based on Disaggregation of Micro-Doppler Signatures

Shelly Vishwakarma and Shobha Sundar Ram
IIIT-Delhi, New Delhi, India, 110020
Email: {shellyv, shobha}@iiitd.ac.in

***Abstract*- Micro-Doppler signatures of dynamic indoor targets (such as humans and fans) serve as a useful tool for classification. However, all the current classification methods are limited by the assumption that only a single target is present in the channel. In this work, we propose a method to classify multiple targets that are simultaneously present in the channel on the basis of a single channel source separation technique. We apply sparse coding based dictionary learning (DL) algorithms for disaggregating micro-Doppler returns from multiple targets into its constituent signals. The classification is subsequently carried out on the disaggregated signals. We have tested the performance of the proposed algorithm on simulated human and fan data.**

I. Introduction

Radar backscatter from non-rigid dynamic targets frequency modulate the carrier signal from continuous wave radars and give rise to micro-Doppler features (mD) in the joint time-frequency space [1-3]. These mD signatures have been used for target classification for purposes such as law enforcement, surveillance and search and rescue missions [4-10]. In [4], authors used spectral analysis to discriminate mD signatures corresponding to different micro-motions. In [5], the time domain information of the backscattered signal was used as an input to dynamic time warping algorithms for classification purposes. Authors in [6-7] used heuristic methods while [8] employed principle component analysis (PCA) and independent component analysis (ICA) to extract features from time-frequency spectrograms for classification. Empirical mode decomposition (EMD) achieved classification accuracy up to 90% in [9]. All of these schemes used supervised learning methods that require domain specific knowledge. In [10], Kim deployed a deep convolution neural network (DCNN), an unsupervised feature learning method, for human activity classification.

All of the existing methods have however made one fundamental assumption – that the channel consists of only a single target during classification. But, in real world scenarios, backscattered signals, especially in indoor environments, may comprise of aggregated signals from multiple targets. For instance, a room may comprise of humans and other moving targets such as fans, speakers etc. all of which give rise to mD returns. Therefore, separating these returns into their constituent signals must be the preliminary step before classification. Otherwise, target detection and classification will be significantly distorted by the interference from multiple targets. This problem can be therefore framed as single channel source separation (SCSS) problem. SCSS has been previously used in medical imaging, energy disaggregation in smart grids, and speech processing applications [11-12].

In this paper, we have addressed this SCSS problem using dictionary learning (DL) based on sparse coding for disaggregation of mD returns from multiple targets. We simulate data from two classes of moving targets- walking humans and rotating fan. The superposed returns of the targets belonging to the two class are separated using the proposed method. The performance of the classification algorithms are compared for the aggregate returns and disaggregated returns. We demonstrate that in the case of multiple targets, the classification accuracy significantly improves provided the disaggregation is first carried out.

II. Theory

A. *Sparse Coding Based Dictionary Learning for Disaggregation and Classification*

A dictionary, B, is a set of vectors, which can be linearly combined to represent any signal X.

$$X = BA \tag{1}$$

The DL algorithm automatically learns discerning features and classification boundaries from the training dataset by optimizing a specified loss function in an unsupervised manner. Since the algorithm involved unsupervised learning, it overcomes the limitations associated with fixed representation matrices for signals such as wavelets, mathematical transforms and filter banks. A dictionary is an overcomplete representation of a signal that not only captures the discontinuities present in the signal but, also provides a sparse representation of the signal. Learning a dictionary from a training dataset $X \in \mathbb{R}^{M\times N}$, where M is the dimension of the signal vector and N is the number of signal vectors, is a two stage process as described in TABLE 1. At the first stage, a dictionary $B \in \mathbb{R}^{M\times P}$, an overcomplete representation having greater number of columns than the signal dimension M given by $P \gg M$, is initialized randomly and $A \in \mathbb{R}^{P\times N}$ is the corresponding sparse coefficient matrix. Sparse coefficients are obtained using matching pursuit algorithms such as orthogonal matching pursuit (OMP) [13], employing l_0 minimization technique. As solving l_0 is NP-hard, it can be relaxed to l_1 minimization using basis pursuit (BP) algorithms [14]. Once the sparse coefficients are extracted, the dictionary is updated using a simple least squares approach. Each column of dictionary is normalized to have a norm less than unity. There are several algorithms such as KSVD and method of optimized directions (MOD) for learning dictionaries

978-1-5090-1592-4/16/$31.00

TABLE 1. DICTIONARY LEARNING FRAMEWORK

Input: Training data matrix $\boldsymbol{X} \in \mathbb{R}^{\boldsymbol{M}\times\boldsymbol{N}}$, $\boldsymbol{\mu} \in \mathbb{R}$ is the regularization parameter and $\boldsymbol{\sigma}$ is a noise variance

Loop until convergence

Stage 1: Sparse coding

$$\{A\} = \min_{A} \|X - BA\|_F^2 \text{ s.t. } \|A\|_0 < \sigma \text{ using } l_0$$

Or

$$\{A\} = \min_{A} \|X - BA\|_F^2 + \mu\|A\|_1 \quad \text{using } l_1$$

Stage 2: Dictionary update

$$\{B\} = \min_{B} \|X - BA\|_F^2$$

$$\text{s.t } \|b_i\|_2^2 \leq 1 \ \ for \ \forall \ i = 1, 2, \ldots . . N$$

End

TABLE 2. DISAGGREGATION AND CLASSIFICATION ALGORITHM BASED ON SPARSE CODING

Input: Class specific dictionaries $\boldsymbol{B_k} \in \mathbb{R}^{\boldsymbol{M}\times\boldsymbol{N}}$,for $\boldsymbol{k} = \boldsymbol{1}, \boldsymbol{2}, \ldots, \boldsymbol{c}$, $\boldsymbol{c}$ is the number of classes, $\boldsymbol{\mu_1} \in \mathbb{R}$ is the regularization parameter

Stage 1: Disaggregation of $\boldsymbol{X_{mixed}}$ signal using learnt class specific dictionaries from $\boldsymbol{c}$ different classes concatenated together, $\boldsymbol{B_{1:c}} = [\boldsymbol{B_1}\ \boldsymbol{B_2} \ldots \boldsymbol{B_c}]$

(a) **Loop until convergence**

$$\{\widehat{A}_{1:c}\} = \min_{\widehat{A}_{1:c}} \left\|X_{agg} - B_{1:c}\widehat{A}_{1:c}\right\|_F^2 + \mu_1 \left\|\widehat{A}_{1:c}\right\|_1$$

end

(b) **Predict** $\widehat{\boldsymbol{X}}_{\boldsymbol{k}} = \boldsymbol{B_k}\widehat{\boldsymbol{A}}_{\boldsymbol{k}}$, for $\boldsymbol{k} = \boldsymbol{1}, \boldsymbol{2}, \ldots . . \boldsymbol{c}$

Stage 2: Classification using SRC

(a) Classification before disaggregation

Loop until convergence

$$c = \min_{k} \left\|X_{agg} - \ B_k\widehat{A}_k\right\|_2^2 \ \ for \ \forall \ k = 1, 2, \ldots c$$

end

(b) Classification after disaggregation

Loop until convergence

$$c = \min_{k} \left\|\widehat{X}_k - B_k\widehat{A}_k\right\|_2^2 \ \ for \ \forall \ k = 1, 2, \ldots c$$

End

from the datasets [15]. In this work, we use iterative least squares dictionary learning algorithm (ILS-DLA) for updating each column of matrix in every iteration [16]. Once the dictionaries are learnt, they can be directly used to classify test signals. Alternately, the DL framework can be further extended to solve SCSS (or disaggregation) problems. In this work, we compare the performance of direct classification using dictionary learning and the

Consider C diverse classes having training data matrices $X_k \in \mathbb{R}^{M\times N}$, for every class $k = 1,2, \ldots\ldots, C$, where column x_k^j represents j^{th} signal vector from k^{th} class. The objective of the method is to separate aggregated signals $X_{agg} = \sum_{k=1}^{C} X_k$ into their constituents $\hat{X}_1, \hat{X}_2, \ldots \hat{X}_k, .., \hat{X}_C$. The task of disaggregation is achieved using the algorithm illustrated in TABLE 2. The main intuition behind the proposed method is if B_k is trained to reconstruct X_k with least error of reconstruction, then it must reconstruct the k^{th} portion of aggregate signal X_{mixed} better than any other B_l for $l \neq k$. The proposed method assumes that the dictionaries learnt for different classes are discriminative, less coherent and must give significant amount of disaggregation. For classification of separated signals, sparse representation based classification (SRC) is used in TABLE 2 [17]. It simply selects the class having minimum representation error amongst all classes under test, using the concatenated dictionaries. This classification step can be applied on any test signal $\hat{X}_k$ either before or after disaggregation. In order to illustrate the benefits of disaggregation – we compare the performance of the classification algorithms on the aggregated (mixed) data from multiple targets with the performance on the disaggregated returns.

B. *Simulation of Walking Human and Fan*

To demonstrate the application of DL based disaggregation technique on mD signatures, we have simulated time domain returns from two classes of moving targets- walking human and a rotating ceiling fan. We have used a point scatterer model for modelling the targets backscatter [18-19]. The kinematic description of the human is derived from the human walking model developed by Boulic, et. al. [18], based on biomechanical experimental data. In this model, the dynamics of human motion are described using 12 time dependent trajectories that control the location of 17 reference points on the human body. These are the head, shoulders, the neck, two knees, spine base, hips, elbows, ankles and toes. These trajectories are analytic expressions and functions of two parameters - the height of the human and the relative velocity of the human. Thus by varying these two parameters, a corresponding variety of human motion descriptions can be derived. The aggregate received signal is composed of returns from these 17 points representing the human body as shown in Fig. 1(a). The joint time-frequency signature of the mD returns is derived using short time frequency transform (STFT) with a Gaussian window of size 0.02sec as shown in Fig. 1(c). The human walks along the XY plane and is upright along the Z axis.

The rotating fan is modelled using a 3 point scatterer model, where the point scatterers are situated mid-way on 3 rectangular blades as shown in Fig. 1(b). The scattered returns from the fan are a function of three parameters – the angular velocity of the motion, and the length and width of the blades. The fan is hung along the Z axis and the blades move circularly along the XY plane. The backscattered signal is the superposition of returns from all 3 point scatterers as shown in Fig. 1(d).

III. SIMULATION RESULTS

To validate the performance of the proposed algorithm, we simulated the radar returns of two targets- human walking towards the radar and the rotating fan, at a carrier frequency of 2.5 GHz. The duration of the signal is 1 second and a sampling frequency of 1 KHz is chosen to prevent aliasing. Therefore, each signal vector is of size [1000 x 1]. The geometry of the room, the location of the monostatic radar and fan as well as the path of the human considered in the problem are depicted in Fig. 2. We vary the human parameters (height and relative velocity) and the fan parameters (blade length, width and angular velocity) to simulate data for multiple unique cases. We vary the human heights from 1.6002m to 1.8288m and relative velocity ranging between 1 (H_t/sec) and 1.9 (H_t/sec), where H_t is the height from toe to hip as mentioned in [18]. Similarly, in the case of the fan, we vary the blade lengths between 0.60m to 0.80m, the blade widths between 0.09m to 0.12m and angular velocities between 300 RPM to 500 RPM. Overall, 300 unique human and fan cases were generated out of which we used 80% of the data for training and 20% for the testing purposes with five-fold cross validation. Thus the training matrix for each class was of size$[1000 \times 240]$. From this, we used the techniques described in TABLE. 1 to learn human and fan overcomplete dictionaries of size $[1000 \times 4000]$ each. We tested our algorithm for disaggregation and classification using test data of size $[1000 \times 60]$, comprising of aggregated returns from both human and fan. The aggregated data is derived by simply superposing the time domain data from both the individual cases.

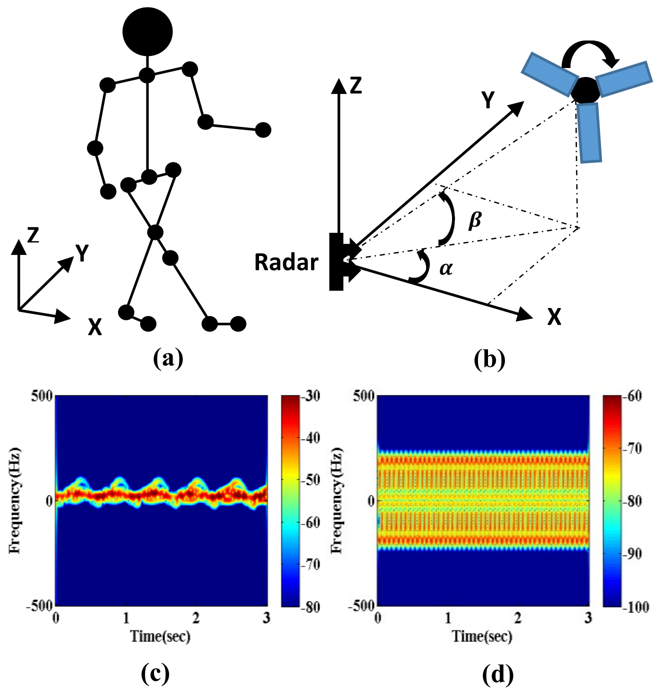

Fig. 1 (a) Point scatterer model of walking human, (b) Point scatterer model of rotating fan about y-axis, Spectrogram using STFT of (c) human walking towards radar, (d) rotating fan

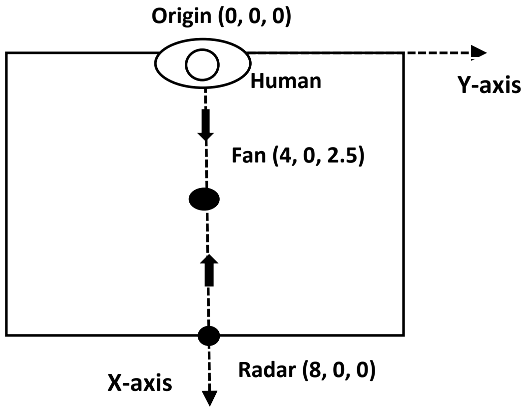

Fig. 2 Geometry of the room

Then for each test case, the dictionaries of both classes were concatenated. Using the first stage of the algorithm described in TABLE.2, we reconstructed the discrete components belonging to each class under test. The algorithm was allowed to run for maximally 500 iterations in MATLAB on a 2.4GHz Intel processor. Fig. 3(a) shows the spectrogram corresponding to the aggregate signal from one test case comprising both the human and fan returns. By comparing this figure to Fig. 1(c) and (d), it is evident that the spectrogram shows features belonging to both the targets. Hence, it is quite likely that the classification algorithms will not be capable of correctly classifying the aggregated signal. In other words, the classifier is likely to detect the target with the stronger radar cross-section. A thresholding criterion was applied on the sparse coefficients obtained after disaggregation. These coefficients correspond to each class and hence the constituent signals can be reconstructed using these coefficients. Therefore, a target is detected only if the corresponding coefficients are above the threshold. The spectrograms generated from the disaggregated signals corresponding to the human and the fan are shown in Fig. 3(b) and Fig. 3(c) respectively. Both of these figures show significant similarity to Figs. 1(c) and 1(d) respectively. There is error in the frequency regions of overlap between the two targets. However, some of the distinctive features (such as the frequency span and the periodicity) are retained in the disaggregated spectrograms. Therefore, the classification algorithms are likely to perform well on the disaggregated signals.

We classified the aggregated test signals before applying the proposed disaggregation algorithm and evaluated the performance of classifier. Next, we performed the disaggregation processing, and presented the separated constituents for target detection and classification. In both cases (before and after disaggregation), the SRC classifier is used. In TABLE 3, the classification accuracies for three scenarios (each with 60 test cases) have been presented. In the first scenario, only a single target, the human, is assumed to be present. In the second scenario, only the fan is assumed to be present and in the third scenario, both the human and fan are assumed to be present. The ground truths reflect these three scenarios in the table. Then the classification results for before disaggregation

(BD) and after disaggregation (AD) are presented. The results show that in the case of the single target, SRC classifier correctly identifies the targets using their corresponding dictionaries. However, in the case where two targets are present simultaneously, the algorithm detects the human but is not capable of identifying the fan possibly since the RCS of the humans are greater than the fan. However, after disaggregation, both the human and fan are correctly identified. From these results it becomes evident, that dictionary learning without disaggregation is successful in detecting single targets in a channel. However, an additional step of disaggregation is required for detecting and classifying multiple targets in the same channel.

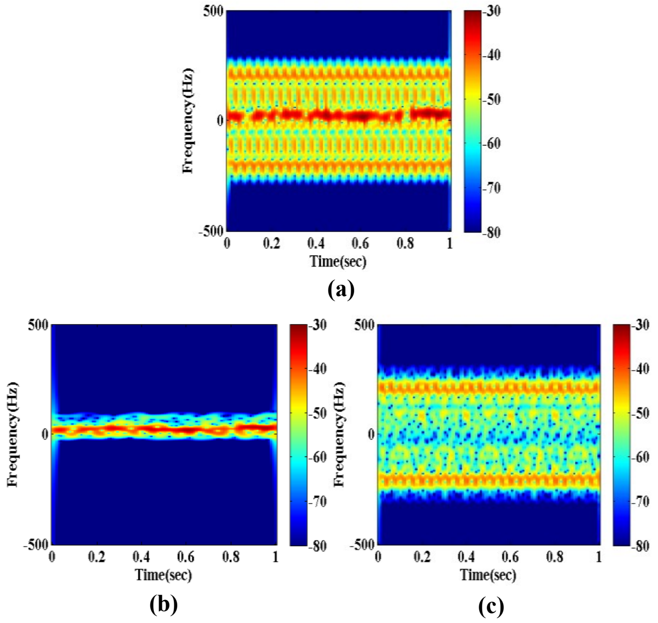

Fig. 3 (a) STFT of aggregate m-DR from human walking towards radar, and rotating fan, (b) STFT of disaggregated m-DR of human walking towards radar and (c) rotating fan

TABLE 3. CLASSIFICATION ACCURACY OF PROPOSED METHOD BEFORE AND AFTER DISAGGREGATION USING FIVE- FOLD CROSS VALIDATION

Cases	Classification	Human towards radar (%)	Rotating fan (%)
Single target-Walking Human	Ground truth	100	0
	BD	100	0
	AD	100	0
Single target-Rotating fan	Ground truth	0	100
	BD	0	100
	AD	0	100
Two targets Walking human and fan	Ground truth	100	100
	BD	100	0
	AD	100	100

IV. CONCLUSION

In this paper, we applied sparse coding based dictionary learning to characterize the mD from indoor dynamic targets for detection and classification purposes. The main advantage of the algorithm is that it learns features in an unsupervised manner by selecting an appropriate level of sparsity thus, relaxing the necessity of domain specific knowledge of the features. The algorithm performs single channel source separation to disaggregate the time domain mD from multiple targets. This enables the detection and classification of weak targets that would otherwise be missed.

REFERENCES

[1] V. C. Chen, F. Li, S.-S. Ho, and H. Wechsler, “Analysis of micro-Doppler signatures”, in *IEEE Proc. Radar, Sonar, Navig*., vol. 150, no. 4, pp. 271–276, Aug. 2003.

[2] V. C. Chen, "Analysis of radar micro-Doppler with time-frequency transform", in *Proc. of IEEE Workshop on Statistical Signal and Array Processing*, pp. 463-466, 2000.

[3] 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”, in *EURASIP Journal on Advances in Signal Processing*, no. 1, pp.1-18, 2013.

[4] A. G. Stove, and S. R. Sykes, “Doppler-based automatic target classifier for a battlefield surveillance radar,” in *Proc. Of IEEE Int. Radar Conf., Edinburgh, U.K.*, pp. 419–423, Oct. 15–17, 2002.

[5]G.E Smith, K. Woodbridge, and C. Baker, “Template based micro-Doppler signature classification”, in *IET seminar on High Resolution Imaging and Target Classification*, pp. 127-144, Nov. 2006.

[6] T. Thayaparan, L. Stanković, and I. Djurović. "Micro-Doppler-based target detection and feature extraction in indoor and outdoor environments." in *Journal of the Franklin Institute*, vol. 345, no. 6, pp. 700-722, 2008.

[7] Y. Kim and H. Ling, “Human activity classification based on micro-Doppler signatures using a support vector machine,” in *IEEE Trans. Geosci. Remote Sens*., vol. 47, no. 5, pp. 1328–1337, May 2009.

[8] V.C. Chen, "Spatial and temporal independent component analysis of micro-Doppler features", in *IEEE Int. Radar Conf.*, pp. 348-353, May 2005.

[9] D. Fairchild and R. Narayanan, “Classification of human motions using empirical mode decomposition of human micro-Doppler signatures,” in *IET Radar, Sonar, Navig*., vol. 8, no. 5, pp. 425–434, Jun. 2014.

[10] Y. Kim, and T. Moon. "Human Detection and Activity Classification Based on Micro-Doppler Signatures Using Deep Convolutional Neural Networks." In *IEEE LETTERS Geosci. Remote Sens*., vol. 13, no. 1, Jan. 2016.

[11] G. Hart. “Nonintrusive appliance load monitoring”, in *IEEE Proc.*, vol. 80, no. 12, 1992.

[12] D. Barry, E. Coyle, D. Fitzgerald, and B. Lawlor, "Single channel source separation using short-time independent component analysis" in *Audio Engineering Society Convention 119, Audio Engineering Society*, 2005.

[13] T.T. Cai, and L. Wang, "Orthogonal matching pursuit for sparse signal recovery with noise" in *IEEE Trans. on Information Theory*, vol. 57, no. 7, pp. 4680-4688, 2011.

[14] S. Chen, and D. Donoho, "Basis pursuit." in *IEEE Asilomar Conf. on Signals, Systems and Computers*, vol. 1, pp. 41-44, 1994.

[15] R. Rubinstein, T. Peleg, and M. Elad, “Analysis K-SVD: A dictionary-learning algorithm for the analysis sparse model” in *IEEE Trans. on Signal Processing*, vol. 61, no. 3, pp. 661-677, Feb. 2013.

[16] K. Engan, K. Skretting, and J. H. Husøy, “Family of iterative LS-based dictionary learning algorithms, ILS-DLA, for sparse signal representation” in *Digital Signal Processing*, vol. 17, no. 1, pp. 32-49, 2007.

[17] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma. “Robust face recognition via sparse representation”, *in IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 31, 2, pp. 210–227, 2009.

[18] R. Boulic, N. M Thalmann, and D. Thalmann, "A global human walking model with real-time kinematic personification." in *visual computer*, vol. 6, no. 6, pp. 344-358, 1990.

[19] V. C. Chen, “Doppler signatures of radar backscattering from objects with micro-motions”, in *IET Signal Process*., vol. 2, no. 3, pp. 291-300, 2008.