Radar Data Cube Processing for Human Activity Recognition Using Multisubspace Learning

BARIS EROL [ID] , Student Member, IEEE
MOENESS G. AMIN [ID] , Fellow, IEEE
Villanova University, Villanova, USA

In recent years, radar has been employed as a fall detector because of its effective sensing capabilities and penetration through walls. In this paper, we introduce a multilinear subspace human activity recognition scheme that exploits the three radar signal variables: slow-time, fast-time, and Doppler frequency. The proposed approach attempts to find the optimum subspaces that minimize the reconstruction error for different modes of the radar data cube. A comprehensive analysis of the optimization considerations is performed, such as initialization, number of projections, and convergence of the algorithms. Finally, a boosting scheme is proposed combining the unsupervised multilinear principal component analysis (PCA) with the supervised methods of linear discriminant analysis and shallow neural networks. Experimental results based on real radar data obtained from multiple subjects, different locations, and aspect angles (0°, 30°, 45°, 60°, and 90°) demonstrate that the proposed algorithm yields the highest overall classification accuracy among spectrogram-based methods including predefined physical features, one- and two-dimensional PCA and convolutional neural networks.

Authors' address: B. Erol and M. G. Amin are with the Center for Advanced Communications, Department of Electrical and Computer Engineering, Villanova University, Villanova, PA 19085 USA, E-mail: (berol@villanova.edu; moeness.amin@villanova.edu). (Corresponding author: Baris Erol.)

I. INTRODUCTION

Recent studies have revealed that falls were the leading cause of fatal and nonfatal injuries for people aged 65 and over [1]. As a result, fall detection systems have been identified as a major innovation opportunity to improve the quality of life for elderly. Much attention has been recently given to fall detection using the radio frequency sensing modality. The rising interest in using radar for fall detection is a response to aging population requiring effective elderly care and assisted living [2]–[4]. It is driven by the ability to preserve privacy, and is propelled by advances in machine learning and hardware-software integration.

Various contributions to radar fall detection have employed different features, both predefined and automatically learned, to classify micro-Doppler signatures [5]. The features can be speech inspired [6], derived from principal component analysis (PCA) [7], [8], based on specific kinematics [9], or drawn from transform-domains. Examples of transform-based features include discrete cosine transform coefficients [10], and features derived from the cadence velocity diagram [11], such as pseudo-Zernike moments [12] and shape spectrum features [13]. Recent human motion classification efforts have applied deep neural networks (DNNs) [14], [15]. DNNs, however, lack the availability of large data size for proper training and performance validations. Moreover, DNNs are only applied on time-frequency (TF) representations and their performance has not been fully investigated on a challenging database, such as ours with multiple subjects, locations, and aspect angles.

Radar backscattering signals from range-Doppler (RD) radar, such as frequency modulated continuous wave (FMCW), provide target information along the three variables of fast-time, slow-time, and Doppler frequency. Accordingly, two-dimensional (2-D) joint-variable signal representations can be constructed, depicting the received data in the TF domain, the RD domain, and the range versus slow-time (range-map) domain. Compared to a one-dimensional (1-D) single-variable domain, the 2-D joint-variable representations have shown the ability to reveal intricate properties of the target complex motion, specifically the time-dependency of target velocity. Each 2-D motion data representation provides distinct and valuable information that might not be present or difficult to extract in other 2-D domains.

PCA is an unsupervised dimensionality reduction method proposed to solve problems related to high dimensionality. It reduces the dimensionality of an input data set consisting of correlated variables into an output set of linearly uncorrelated variables, referred to as principal components. The output set, with reduced dimensionality, attempts to capture the variation present in the input data. PCA can be seen as a smart compression technique, however, since the principal components are uncorrelated, it can also be categorized as an automatic feature learning method, which has proven effective, offering high fall classification rates [16].

Conventional PCA-based human motion classification methods deal with one joint-variable representation of the

radar data, whether it is TF, range-map, or their fusion. In each case, we vectorize the 2-D data representation and use the result to determine the principal components. Generalized PCA, on the other hand, does not require a vectorization step in the beginning and tries to find two projection subspaces that operate on different modes of the 2-D joint variable representation [17]. Variations of this approach has been used in radar applications, more specifically for activity recognition [18] and gait abnormality detection [19]. In this paper, we apply multilinear PCA (MPCA) to the radar data cube (RDC), rather than lower-dimension processing.

This paper marks the first attempt to use three-dimensional (3-D) joint-variable signal representation to exploit the underlying dependency and correlations among the three radar signal variables. Encouraged by the classification results of 1-D and 2-D PCA, and recognizing possibilities for improvements, we pursue multilinear dimensionality reduction methods using a tensor analysis [20]. A comparison between the baseline performances of predefined physical features, 1-D PCA, 2-D PCA, a 12-layer convolutional neural network (CNN) (all operate on TF domain) and MPCA is provided to show the merits of using RDC. We use a database that contains multiple aspect angles, ($0°$, $30°$, $45°$, $60°$, and $90°$), a large variance within class samples because of performing each motion with different speeds, and collecting data from different locations and subjects. Because of these data challenges, we propose a boosting scheme for RDC-based processing where we combine MPCA with two supervised methods, linear discriminant analysis (LDA) and shallow neural networks (SNN), so as to increase the classification performance. We also examine the underlying optimization parameters and their effects on performance and convergence.

The paper is organized as follows. In Section II, the radar signal model and different joint variable domains are presented. In Section III, multilinear subspace learning methods are introduced. In Section IV, the design and optimization considerations of multilinear subspace learning methods are discussed in terms of iterative projections, initialization, termination criteria, convergence, and number of projections. In Section V, we introduce a boosting scheme of MPCA by combining it with two supervised methods followed by decision fusion. Section VI presents a comparison of specific subspace learning methods under different RDC dimensions. Finally, conclusions are provided in Section VII.

II. RADAR SIGNAL MODEL

Commercially available FMCW radars are in a compact form and have the ability to provide range, time, and Doppler information. The radar system used in this paper is SDRKIT 2500B, which is developed by Ancortek, Inc. Operating parameters of this radar system are: transmitting frequency 25 GHz, sampling frequency 1 kHz, and bandwidth of 2 GHz, which provides 0.075-m range resolution. The radar can switch between continuous wave and FMCW

modes and operates in a monostatic configuration, which consists of one transmitting and one receiving antennas.

For an FMCW radar, the backscattering signal from a target located at a distance R can be expressed as

$$s_{rx}(t) = A_{rx} \cos \left(2\pi j \left(f_0(t - \tau) + \frac{r}{2}(t - \tau)^2 \right) + \phi_{rx} \right) \tag{1}$$

where τ is the round trip time delay, f_0 is the carrier frequency, r is the chirp rate, ϕ_{rx} is the phase of the received signal, and A_{rx} is the amplitude of the received signal computed from the range radar equation as

$$A_{rx} = \frac{G\lambda\sqrt{P\sigma}}{(4\pi)^{1.5}R^2\sqrt{L_s}\sqrt{L_a}}. \tag{2}$$

Here, G is the antenna gain, P is the transmitter power, σ is the target radar cross section (RCS), and L_s and L_a represent the system and atmospheric losses, respectively.

The received signal is then demodulated by the I/Q demodulator, providing the in-phase and quadrature-phase components of the complex baseband signal, expressed as

$$s(t) = I(t) + jQ(t) = A \exp \psi(t) \tag{3}$$

where ψ is the phase of the signal. The sampled I/Q signal can be transformed into a 2-D matrix representation where the columns and rows, respectively, represent fast and slow time variables. The range map can be computed by taking the discrete Fourier transform along the matrix columns. The range maps for different human motion activities are depicted in Fig. 1(a) through (e). In Fig. 1(a), the human subject falls toward the radar and, as a result, we observe a high range spread, which is determined by the subject's height. Bending, kneeling, and sitting [see Fig. 1(b), (d), (e)] have small range spreads because of the limited translational motion of the body in range. In the case of a walking, subject starts walking toward the radar around 3.5 m away and stops around 1 m, as it can be clearly seen in Fig. 1(c).

The Doppler information of the human motion activities are often discriminative and descriptive of the nature of the signal, its source, and the mechanism of its generation. Raw slow time data, $x(n)$, where n is the slow time, can be computed by agglomeration of the range map along its range axis (columns). The resulting data from the human (individual being tracked) is nonstationary with time-varying frequencies, which are associated with velocity, acceleration, and higher order terms of different human body parts. Spectrograms are the simplest and most commonly used TF distribution. It is the energetic form of the short-time Fourier transform, which is obtained by splitting the time-domain signal into many overlapping or disjoint consecutive segments, and then taking the fast Fourier transform of each segment. This provides the signal's local frequency behavior. The spectrogram is mathematically defined as

$$S(n, k) = \left| \sum_{m=0}^{N-1} h(m)x(n - m)e^{-j2\pi km/N} \right|^2 \tag{4}$$

where $h(m)$ is a window function, which can affect both the time and frequency resolutions. The TF representation of
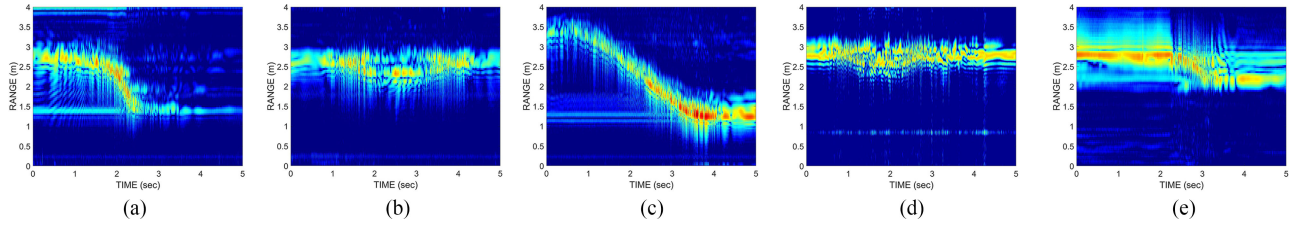
Fig. 1. Range maps of different motions. (a) Falling. (b) Bending. (c) Walking. (d) Kneeling. (e) Sitting.
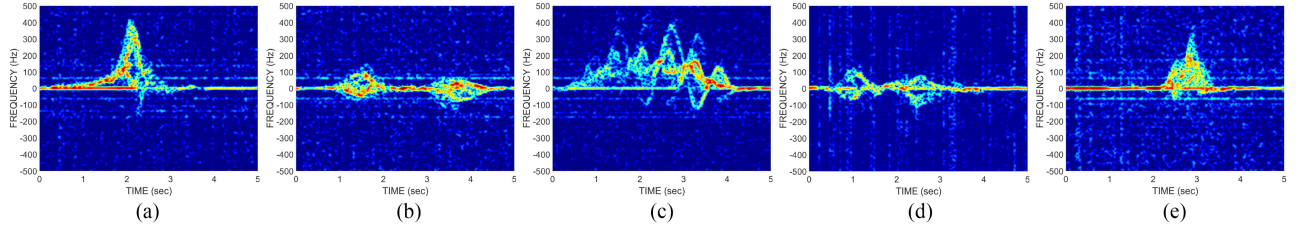

Fig. 2. TF representations of different motions. (a) Falling. (b) Bending. (c) Walking. (d) Kneeling. (e) Sitting.

five different motions are provided in Fig. 2(a) through (e). Note that, in the walking motion [see Fig. 2(c)], the torso exhibits a sinusoidal pattern with the strongest return. The periodic motion of the limbs causes frequency oscillations modulated around the main Doppler shift. The motion of the legs causes the highest frequency oscillation, followed by that of the arms, which appear to be more distinct. Sitting spectrogram, provided in Fig. 2(e), appears to be more dense along the frequency axis because of the motion of the upper body as a whole. Interestingly, in bending, positive and negative frequency components are present within the same time interval [see Fig. 2(b)]. As the upper body translates closer to the radar, the lower body (human back) moves away from the radar. Kneeling motion has the same kinematics as bending with one significance difference; motion of the knee. Kneeling is performed as in tying the shoelace. Therefore, subjects put their one knee on the ground while tying the shoe. This effect can be clearly seen in Fig. 1(d). Fall spectrogram, shown in Fig. 2(a), depicts a sudden drop in frequency in the shape of a positive "hump." In falling, the human subject puts the knee first on the ground to slow down the fall and this effect can be clearly seen around 1.8 s in the TF domain.

Compared to 1-D single-variable domain, the 2-D joint-variable representations is more effective in human motion classification. Each 2-D motion data representation provides distinct and valuable information that might not be present in other 2-D domains. Acquiring all the information, which is needed for better classification, often requires a parallel processing of 2-D domains, which might result in increased computational complexity and loss of information present in the interrelation between the domains. Therefore, we utilize a 3-D joint-variable signal representation to exploit the underlying dependency and correlations among the three radar signal variables, namely, range, slow time, and Doppler frequency. RD representation includes the effects of both target velocity and range [21]–[23].

Typically, a single RD frame can be obtained by applying the Fourier transform for each range bin over a nonoverlapping coherent processing interval (CPI) of slow-time on the range map. In this paper, the CPI is determined as 32. We create a tensor structure, called RDC, by stacking consecutive RD frames. Visualization of the RDC is usually done by a video sequence of RD frames, however, in [24], RDC was visualized by creating a surface that has the same intensity value within the slices of data cube. This can be accomplished by isosurface method, which is a 3-D extension of an analog isoline. Volumetric representations of the RDC for falling, bending, walking, kneeling, and sitting are presented in Fig. 3(a) through (e), respectively. Note that, RDC gathers fast-time, slow-time, and Doppler frequency information in a single domain, aiming at increased detection and reduced false alarms.

The first step of the our RDC extraction procedure includes a preprocessing approach called extended CLEAN (eCLEAN) and aims at suppressing unwanted distortions or noise effects while enhancing the natural structural integrity of the data. It is noted that the eCLEAN algorithm directly operates on the individual RD frames. The original CLEAN algorithm strives to find the highest peaks in an image, which corresponds to a real target location. At each step of the algorithm, the maximum peak is extracted, and a portion of the point spread function centered at that peak is subtracted until some threshold is met [24]. The number of points, which are needed to be removed, are automatically determined before extraction of peaks using a simple and efficient histogram-based method. The output of the algorithm is depicted in Fig. 4(c) when a noisy falling RDC data are given in Fig. 4(a).

III. MULTILINEAR SUBSPACE LEARNING

In dimensionality reduction, the simplest approach contains a preprocessing step where the original raw data
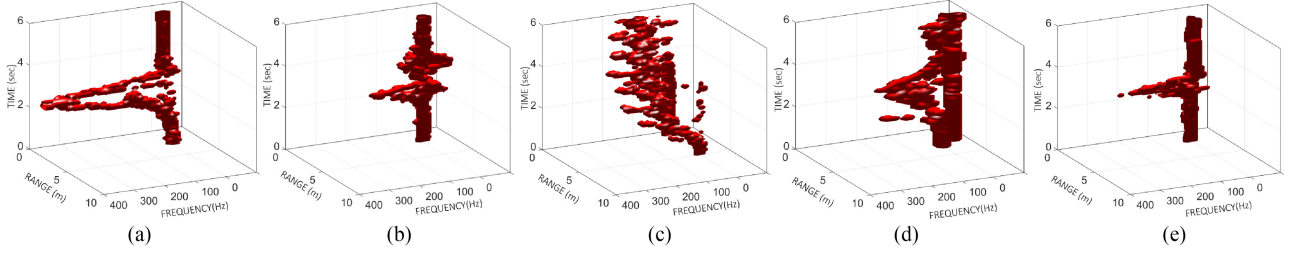
Fig. 3.   Multidimensional RDC representation of different motions. (a) Falling. (b) Bending. (c) Walking. (d) Kneeling. (e) Sitting.
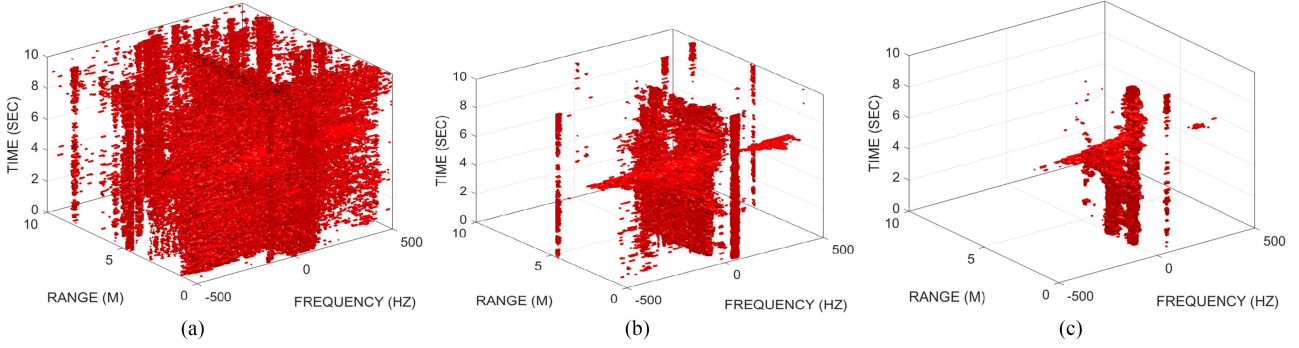

Fig. 4.   Visualization of a noisy falling RDC data. (a) Without any filtering. (b) After basic thresholding. (c) After proposed eCLEAN.

(spectrograms or range maps) are reshaped into vectors. This operation is called vectorization or flattening, which entails a high processing cost in terms of increased computational time and memory access. For example, a typical RDC of size $(64 \times 64 \times 64)$ would translate into a vector with a size of $(262, 144 \times 1 \times 1)$ upon flattening. Moreover, it is well known that the reshaping operation breaks the natural structure and correlation in the original data [25].

In order to overcome the limitations and problems with flattening, some approaches have been recently introduced where the images are represented as matrices (second order tensors) [20]. The image covariance matrix can then be constructed. Spatial correlation of the image pixels within a localized neighborhood is used to create a less restrictive PCA. However, these approaches cannot be applied to higher other tensors (3 or higher). Since our RDC structure has three modes (slow time, range, and frequency), we employ the MPCA proposed in [26]. In addition, we apply multilinear discriminant analysis (MLDA) and compare its performance with that of MPCA [27].

A.   Multilinear Algebra Basics and Notations

This section briefly introduces basic multilinear notations, which are commonly used in multilinear subspace learning algorithms. Vectors are denoted by lowercase symbols, such as p; the normal uppercase symbols represent matrices, e.g., A. An Nth-order tensor is denoted as $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \cdots \times I_N}$. The implemented algorithms often require reshaping of the data which, in the case of tensors, is called unfolding, or matricization. Unfolding $\mathcal{X}$ along the n-mode is denoted as $\mathcal{X}^{(n)} \in \mathbb{R}^{I_n \times (I_1 \times \cdots \times I_{n-1} \times I_{n+1} \cdots \times I_N)}$. Finally, the

n-mode product of a tensor $\mathcal{X}$ by a matrix A is defined by $\mathcal{Y} = \mathcal{X} \times_n A$.

B.   Multilinear Principal Component Analysis

The MPCA is an unsupervised subspace learning algorithm for higher order tensors targeting the same objective function as in 1D-PCA: variation maximization while reducing the dimensions along multiple modes. Assume a set of training tensor samples $(m = 1, 2, \ldots, M)$ given as $\mathcal{X}_m \in \mathbb{R}^{I_1 \times I_2 \cdots \times I_N}$. The objective is to find a tensor-to-tensor projection (TTP) matrix subspace $\tilde{U}^{(n)} \in \mathbb{R}^{I_N \times P_n}$ that projects the original tensors into low-dimensional tensors $\mathcal{Y}_m \in \mathbb{R}^{P_1 \times P_2 \cdots \times P_N}$ (with $P_n \leq I_n$). To project an RDC sample, we utilize three projection matrices, $U^{(1)} \in \mathbb{R}^{I_1 \times P_1}$, $U^{(2)} \in \mathbb{R}^{I_2 \times P_2}$, and $U^{(3)} \in \mathbb{R}^{I_3 \times P_3}$ as

$$\mathcal{Y}_m = \mathcal{X}_m \times_{(1)} U^{(1)^T} \times_{(2)} U^{(2)^T} \times_{(3)} U^{(3)^T}. \quad (5)$$

Note that $\mathcal{Y}_m$ captures most of the variation observed in the original RDC sample. RDC-based dimensionality reduction procedure is illustrated in Fig. 5. It depicts the TTP of an RDC (shown in individual RD slices) with a size of $(64 \times 64 \times 64)$ to a smaller size of tensor with a size of $(P_1 \times P_2 \times P_3)$. Mathematically, the RDC-based objective function can be expressed as the determination of the three projection matrices that maximize the total tensor scatter

$$\{\tilde{U}^{(n)}\} = \arg \max_{U^{(n)}} \sum_{m=1}^{M} \left\| \mathcal{Y}_m - \tilde{y} \right\|_F^2 \quad (6)$$

where the mean of the projected tensor features is defined as $\tilde{\mathcal{Y}} = \frac{1}{M} \sum_{m=1}^{M} \mathcal{Y}_m$. After few iterations, the objective function converges to an eigenvalue decomposition problem and can be solved by using Lagrangian optimization. The
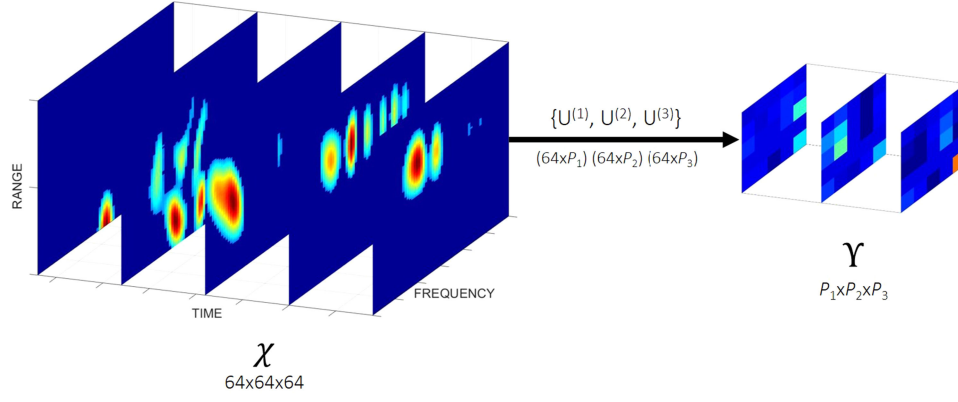
Fig. 5. Illustration of the RDC-based dimensionality reduction procedure through TTPs.

maxima of the objective function are obtained if $\{\tilde{U}^{(n)}\}$ consists of the P_n eigenvectors of the covariance matrices (along each mode) corresponding to the largest P_n eigenvalues [25]. Since the computation of the approximation is multilinear, iterative optimization of (8) can be solved through a sequence of linear subproblems using alternating least squares (ALS), whereby the least squares of the cost function is optimized for one mode at a time, while keeping the other mode subspace matrices fixed. Finally, the feature tensor is obtained by projecting the original tensor using optimized subspace, $\tilde{U}^{(n)}$, as

$$\tilde{\mathcal{Y}}_m = \mathcal{X}_m \times_{(1)} \tilde{U}^{(1)^T} \times_{(2)} \tilde{U}^{(2)^T} \times_{(3)} \tilde{U}^{(3)^T}. \quad (7)$$

Then, feature vector of the mth training tensor can be constructed as $f_m = \text{vec}(\mathcal{Y}_m)$, $\in \mathbb{R}^{1 \times (P_1 \times P_2 \times P_3)}$, and $\text{vec}(\cdot)$ is the matrix vectorization operator. This framework is followed for each of the samples in the database and stored in a matrix representation. Note that the dimensionality of the projected tensor subspace, P_n, is assumed to be known or predetermined. The effect of the P_n is shown in Section IV in terms of classification accuracy and computational complexity.

C. Multilinear Discriminant Analysis

MLDA is a supervised algorithm aiming at maximizing the between-class scatter while minimizing the within-class scatter in the projected subspace. In this paper, we use the DATER algorithm to reduce the dimensionality of the RDC in a supervised fashion [27]. Similar to MPCA, DATER aims to solve the optimization problem through TTPs $\{U^{(n)} \in \mathbb{R}^{I_N \times P_n}\}$ by projecting the original RDC to a low-dimensional tensor while maximizing a discriminative objective criterion. Since MLDA is a supervised method, class label for the mth RDC sample is defined as c_m, and there are C classes in total. In this specific application, we have 5 classes and total of 827 training samples. The objective function of DATER can be expressed as

$$\{\tilde{U}^n\} = \underset{\{U^n\}}{\arg\max} \frac{\sum_{c=1}^{C} M_c \left\| \tilde{\mathcal{Y}}_c - \tilde{\mathcal{Y}} \right\|_F^2}{\sum_{m=1}^{M} \left\| \mathcal{Y}_m - \tilde{\mathcal{Y}}_{c_m} \right\|_F^2} \quad (8)$$

where $\tilde{\mathcal{Y}}$ and $\tilde{\mathcal{Y}}_c$ are defined as the overall and the class means, respectively.

The above optimization problem can be treated as an eigenvalue problem similar to MPCA. The objective function is maximized only if $\{\tilde{U}^n\}$ consists of the P_n eigenvectors of the between-class and within-class scatter matrices associated with the P_n largest eigenvalues. Feature extraction can be performed by following the same steps as in (7).

IV. OPTIMIZATION CONSIDERATIONS FOR MULTI-DIMENSIONAL METHODS

A. Iterative Projections

TTP-based projections usually have N sets of parameters for N projections matrices, one in each mode. These N mode optimization problems cannot be solved simultaneously, except for the case with vectorization. Therefore, the solution of one mode generally depends on the remaining mode projection matrices. To overcome this problem, an iterative algorithm, inspired by ALS, is employed for multilinear subspace learning, referred to as alternating partial projections (APPs) [26]. This algorithm does not cover the whole optimization space, hence, it can be categorized as a suboptimal solution. The APP basically splits the underlying optimization problem into N conditional problems that can be solved sequentially, one by one. In each mode n, it solves the defined conditional subproblem through n-mode partial multilinear projections utilizing the projection matrices obtained from the other modes.

B. Initialization

Because of the iterative nature of the APP solutions, the projection matrices have to be initialized at the beginning of the algorithm. Then, with each iteration, initialization subspaces are updated. Random initialization is one of the most popular approaches where each element of the n-mode projection matrices is constructed from a zero-mean uniform or Gaussian distribution. Another method for initialization is called full projection truncation [25], which requires the utilization of some of the training data.
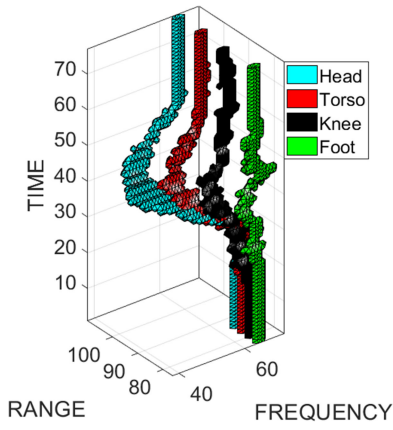
EROL AND AMIN: RADAR DATA CUBE PROCESSING FOR HUMAN ACTIVITY RECOGNITION USING MULTISUBSPACE LEARNING 3621

Fig. 6. Simulated falling RDC from motion capture data for individual joint trajectories.

Fig. 7. Convergence of diversified and random initialization for MPCA.

Since, we have a small database (827 samples), compared with those in image recognition (approximately 1.5 million samples), we utilize a diversified simulation database proposed for transfer learning of spectrograms. This database is generated from 55 MOCAP measurements of five different test subjects using a Kinect RGB-D camera. The Kinect sensor is used as a markerless system for capturing the time-varying coordinate information of human joints needed for radar simulations. The radar return from the human body can be computed as the sum of the reflected signals. For a human target, the RCS is represented by a sphere for the head, and ellipsoid for the torso and limbs. According to these changes, simulation can be computed by following the steps in the radar signal mode by only adding the individual K received signals.

In this paper, four different activities were recorded, namely, walking, running, falling, and sitting. Simulation parameters are matched with the real radar system as: center frequency of 25 GHz, sampling frequency of 1 kHz, and bandwidth of 2 GHz. Three transformations are applied to the underlying skeletal model tracked by the Kinect sensor:

1) Scaling in size, to capture the effect of different heights.
2) Scaling in time, to capture different motion speeds.
3) Random variation of the parameters in a Fourier series model of the trajectory of different points on the body to emulate the effect of different walking styles.

A more detailed description of the diversification methodology and its validation can be found in [28]. Using this approach, original 55 MOCAP measurements were used to generate simulations of 32 000 RDCs. An example for simulated falling RDC is illustrated in Fig. 6. Consistent with the arguments made earlier, the highest Doppler return for falling is obtained from the upper body (head), followed by torso, and lower body. The representation is also similar to a real radar RDC shown in Fig. 3(a).

In image databases, initialization plays an important role in both computational cost, speed of convergence, and performance, albeit this role is both task and domain dependent. In the underlying problem, the diversified
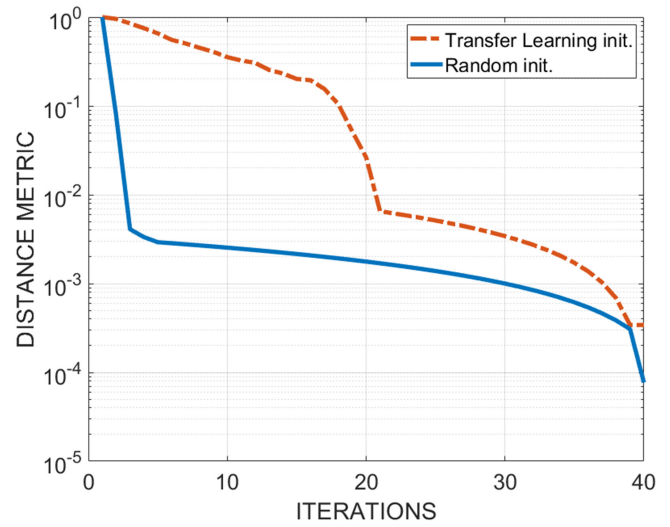
initialization achieves a much faster convergence compared to random initialization, as depicted in Fig. 7, while gaining an increase in the overall classification performance. The convergence termination criteria is met within 3 iterations with the proposed initialization method, whereas for random initialization, it nearly takes 20 iterations to achieve the same value. Clearly, these results prove the merit of using the diversified database to initialize the subspace learning methods.

In this paper, the optimized subspace matrices obtained from diversified database are used in two different ways [29]: 1) *Bottleneck analysis:* Diversified projection matrices are directly utilized to extract the features from the real radar data without any fine tuning (No training performed on real radar data) and 2) *Fine-tuning:* Diversified projection matrices are fine-tuned with the real radar data.

C. Termination Criteria and Convergence

In subspace learning, there are two commonly used termination criteria. In this paper, we examine the tensor distance between the original samples and the reconstructed approximations. The approximation of the mth RDC sample can be expressed as

$$\tilde{\mathcal{X}}_m = \tilde{\mathcal{Y}}_m \times_{(1)} \tilde{U}^{(1)} \times_{(2)} \tilde{U}^{(2)} \times_{(3)} \tilde{U}^{(3)}. \qquad (9)$$

Then, the termination distance criteria can be defined as

$$d = \frac{1}{M} \sum_{m=1}^{M} \left\| \mathcal{X}_m - \tilde{\mathcal{X}}_m \right\|_F^2 . \qquad (10)$$

This distance is computed in each iteration and when it drops below some user-defined hyper-parameter threshold, the algorithm breaks and outputs the optimized projection matrices. However, for computational considerations and in order to avoid any infinite loops, we also set a maximum number of iterations in case the algorithm would not convergence. Note that this iteration is only for controlling the convergence of the algorithm and different than the APP iterations.
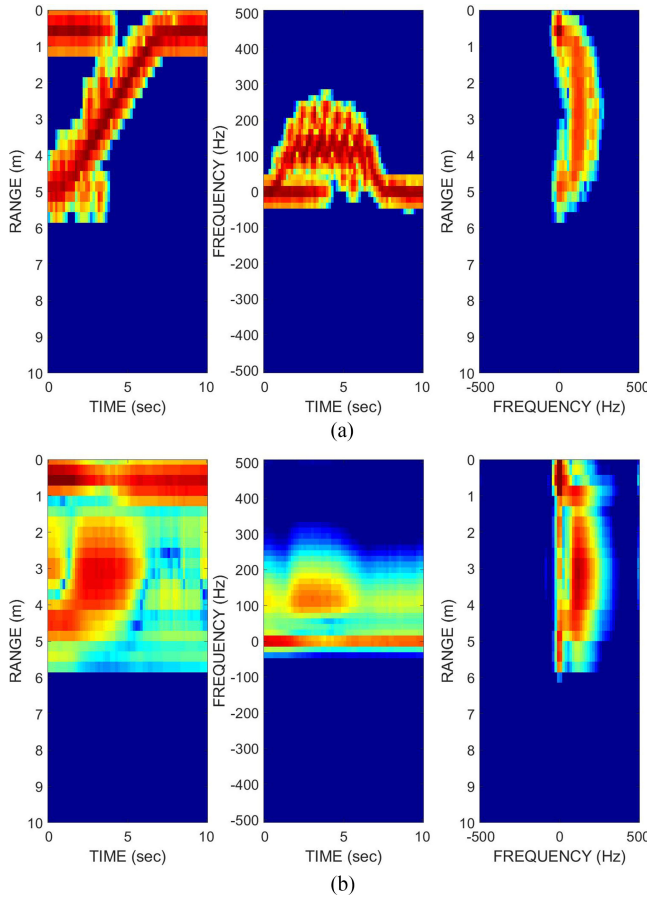
Fig. 8. Visualization of agglomeration operator along different modes for a walking sample. (a) Original. (b) Reconstructed.

Agglomeration along different modes in RDC will result in individual domains, such as TF (note that TF domain is obtained without performing any TF techniques), range map, and time-integrated RD map as depicted in Fig. 8(a). To evaluate and validate our termination criteria, we also analyze the reconstructed RDC (with only five components) and perform the agglomeration on the reconstructed RDC, as shown in Fig. 8(b). It is clear from the figure that the main discriminative properties of these domains are still present after the reconstruction only with five principal components in each mode.

D. Number of Projections

Before solving the multilinear subspace projections, the number of projections (P_1, P_2, P_3) needs to be determined for different modes. For the TTP-based method, the search space could become very large, depending on the size of the tensor ($\prod_{n=1}^{3} I_n$). For the RDC activity recognition task, there are 262 144 possible subspace dimensions. There are several ways to determine the number of projections used in the projection matrices.

1) Q Metric: This metric is a suboptimal, simplified dimension determination procedure that requires no iteration or extensive testing [25]. We start with defining the

ratio as

$$Q^{(n)} = \frac{\sum_{i_n=1}^{P_n} \lambda_{i_n}^{(n)}}{\sum_{i_n=1}^{I_n} \lambda_{i_n}^{(n)}} \tag{11}$$

where $\lambda_{i_n}^{(n)}$ is the i_nth full-projection n-mode eigenvalue. In the Q-based method, the first P_n eigenvectors are kept in n-mode so that it ensures the user-defined threshold parameter. Q metric is an extension of the traditional dimension selection in PCA. In essence, this method discards the least significant eigenvectors in each mode and does not require any classification.

The $Q^{(1)}$, $Q^{(2)}$, and $Q^{(3)}$ ratios are provided in Fig. 9(a) for different number of components. For mode-2, six components are enough to capture 90% of the variation, whereas for mode-1 and mode-1, it nearly takes 15 and 25 components to exceed 90%, respectively. Another important question is related to the prediction time, since the number of components defines the number of features.

2) Brute Force Selection: In brute force selection, we directly compute the classification performance for different number of projections. However, this method does not allow coverage of the entire feature space because the same number of projections are used for each mode. The classification accuracies of MPCA are provided in Fig. 9(b) for different numbers of projections. Moreover, the run-time complexity (time elapsed per prediction) increases in this process owing to the higher dimensionality search. This specific phenomena is indicated in juxtapositions in the same plot [see Fig. 9(b)] where the prediction time follows nearly a linear trajectory for increased number of projections.

3) Genetic Algorithm (GA)-Based Multiobjective Parameter Selection: Q metric does not guarantee any classification improvement, and brute force selection does not span the whole projection space. Therefore, we take a more heuristic and supervised approach where we attempt to find the optimized number of projections in each mode, while maximizing the classification performance together with minimizing the prediction time per sample. This forms a multiobjective optimization (MOO) problem and can be formulated by considering a decision maker that seeks to optimize K objective functions. Without loss of generality and overfitting, all objective functions need to be the minimization type. Given an n-dimensional (in our case $n = 3$, (P_1, P_2, P_3)) decision variable vector x $= x_1, \ldots, x_n$ in the solution space. The main idea is to find a vector $\tilde{x}$ that minimizes the given set of objective functions. The general solution space generally restricted by a series of constraint, such as the bounds on the decision variables (size of the RDC). To describe the terminology of MOO, several explanations have to be made at this point. A feasible solution is said to be Pareto optimal if it is not dominated by any other solution. A Pareto optimal solution cannot be improved with respect to any objective without worsening at least one other objective. The set of all feasible nondominated solutions in a solution space is referred as the Pareto optimal set, and for a given Pareto optimal set, the
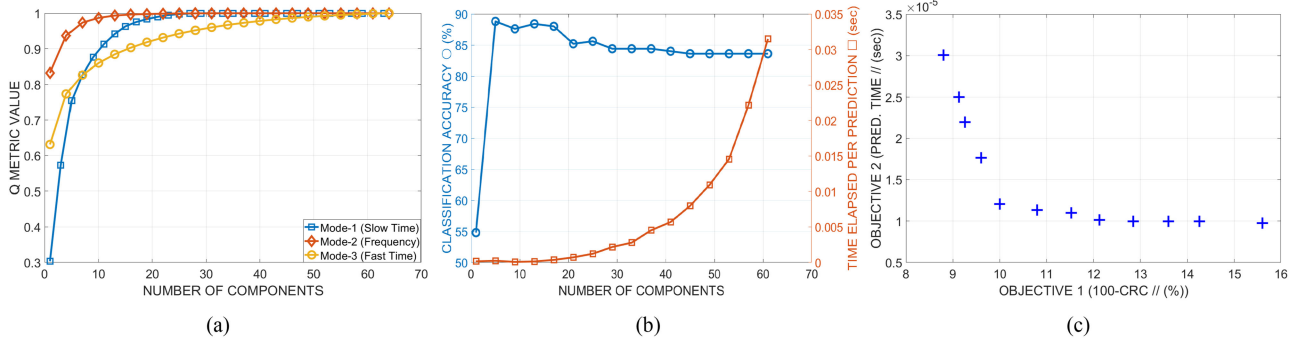
Fig. 9. Determination of the number of projections. (a) Q metric along different modes. (b) Brute force classification with time elapsed per prediction. (c) GA-based multiobjective parameter selection—Pareto plot—.

corresponding objective function values in the objective space are called Pareto front. A more detailed explanation about MOO can be found in [30].

GA can be categorized as a population-based optimization approach, and because of their properties, they are well suited for MOO problems. GA is inspired by the evolutionist theory where in nature, weak and unfit species within their environment are faced with extinction by natural selection. The strong ones have greater opportunity to pass their genes to future generations via reproductions [31]. Fitness functions of the GA are determined as the classification accuracy and the computation time attained by the minimum distance classifier (MDC). We employed the NSGA-II, which is one of the most popular MOO algorithms with three special characteristics: fast nondominated sorting approach, fast crowded distance estimation procedure, and simple crowded comparison operator. Resulting Pareto plot can be seen in Fig. 9(c), where the two objective functions are plotted: classification error and prediction time per sample.

V. BOOSTING THE MPCA

Boosting is an ensemble meta-algorithm for reducing bias and variance in supervised learning. It is stemmed by the question of whether a set of weak learners can be boosted into an accurate strong learner. In this paper, the boosting terminology used in a different sense, where two supervised methods are attached to the features extracted by MPCA. The idea is to extend the feature extraction procedure by selecting more discriminative features from f_m (7). Therefore, we added two supervised method, LDA and SNN, and fuse the achieved soft outputs in decision level. The results were also computed where LDA and SNN were used separately without any decision fusion at the end, however, the achieved results were not satisfactory.

Since the core matrices extracted from the RDCs are already vectorized, assume that S_B and S_W are to be the between-class scatter and within class-scatter based on the f_m, respectively. Then, the corresponding LDA projection matrix can be expressed as U_{LDA} where it consists of the generalized eigenvectors associated with the largest generalized eigenvalues of the generalized eigenvalue problem.

Finally, the feature vector can be extracted as $z_m = U_{LDA}^T f_m$ and given as an input to the classifier.

In this paper, an SNN is also implemented in the boosting scheme. Weight elimination property of the neural network is included by the back propagation algorithm to place a penalty on the MPCA features with redundant information in the fitted neural network model. Weight and biases of neurons are learned such that the cost function J is minimized. Given a set of training MPCA features and corresponding class labels, the cost function can be expressed as

$$J(W, b) = \left[\frac{1}{m} \sum_{i=1}^{m} \left(\frac{1}{2} \left\| h_{W,b}(f_m) - c_m \right\|^2 \right) \right] \quad (12)$$

where W and b are the weights and biases of the network. $h_{W,b}$ is defined as the nonlinear hypothesis that provides the classification. The cost function represents the error between the input data and the neural net output. Back propagation exploits these error values to compute the gradient of the cost function with respect to the weights and biases in the neural network. Afterwards, the gradient is fed to the scaled conjugate gradient descent method, which in turn uses it to update the weights and biases, in an attempt to minimize the cost function. Finally, the output of the fitted neural network is given as an input to the Softmax regression classifier, which is a generalization of logistic regression when there are multiple classes.

In the final step of the scheme, the soft classification labels obtained from LDA and SNN are fused using the L-place operators minimum, maximum, average, and product as the aggregation rule F as

$$c_m^F = F\left(d_{(m,i)}^1, d_{(m,i)}^2 \right) \quad (13)$$

where $d_{(m,i)}^1, d_{(m,i)}^2$ are the soft outputs (each class) of the LDA and SNN, respectively. An illustration of the boosting scheme is shown in Fig. 10.

VI. EXPERIMENTAL RESULTS

Two data sets were collected from two different indoor environments. The first location is the Radar Imaging Lab (RIL), which is a semicontrolled laboratory
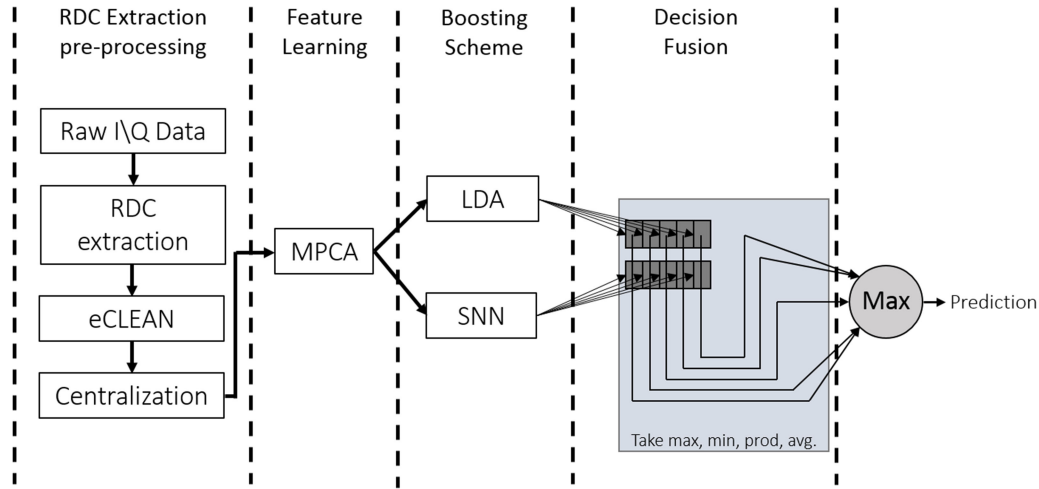
Fig. 10. RDC-based MPCA boosting scheme with LDA, SNN, and decision fusion.

TABLE I
Baseline Classification Performance Comparison Between 1D-PCA, Predefined Physical Features, 2D-PCA,
12-Layer CNN, and MPCA in Terms of Accuracy

	1D-PCA-MDC	Physical Features-SVM	2D-PCA-MDC	CNN	MPCA-MDC
Test accuracy	65.32%	73.65%	83.10%	84.54%	91.4 %

Note that, only MPCA-MDC operates on the RDC.

environment, located at Villanova University, CEER Building. Back and rear walls of the RIL covered with pyramid radiant-absorbent material to eliminate unwanted reflections and avoid any risk of measurement errors and ambiguities. In the RIL experiments, the radar was placed on a table raised 3.2 ft above the ground and pointing directly the back wall.

The second data set was acquired in the Center for Advanced Communications conference room located at Villanova University, Tolentine Building. This location is selected to mimic an uncontrolled environment similar to a senior residence apartment where an office furniture, such as television, plants, tables, bookshelves, chairs, was used. The radar was again placed on a table with a height of 3.2 ft. A Kinect was also located next to the radar in both configurations to record the ground truth optical videos of the motions.

Experiments were performed by 14 human subjects. A total of 827 signals were collected from the test subjects with posed heights ranging from 5.1 to 6.3 ft, weights ranging from 119 to 220 lbs, and included 12 male and 2 female subjects. Five different motions were considered in the experiments: falling (191 samples), sitting (213 samples), bending (203 samples), kneeling (108 samples), and walking (112 samples). Motions were performed for five different directions, namely, $0°$, $30°$, $45°$, $60°$, and $90°$. Each subject performed the experiments with three different speeds: slow (enacting as elderly people), medium, and fast.

We start the simulations by a comparison with 1D-PCA, predefined physical features, 2D-PCA, a 12-layer CNN, and RDC-based MPCA as depicted in Table I. Note that the 1D-PCA, predefined features, 2D-PCA, and CNN are applied on the spectrograms. The average classification accuracies for 1D-PCA, pre-defined features, 2D-PCA, CNN, and MPCA are determined to be 65.32%, 73.65%, 83.10%, 84.54%, and 91.4%, respectively. The MPCA produces the lowest number of missed detections and highest fall detection rate. In essence, MPCA performance is drastically higher than that achieved with existing commonly used algorithms, demonstrating the importance of considering RDC in the application at hand. Then, we move to compare four different RDC-based methods: MPCA with MDC, MLDA with MDC, boosted-bottleneck performance of the simulation projection matrices on real data (BMPCA-Bottleneck), and boosted-fine-tuned-MPCA (BMPCA-Fine-tuned). The classification accuracies are provided in Table II for different dimensions of the RDC. As we increase the dimension, the performances of the RDC MPCA-MDC, BMPCA-Bottleneck, and BMPCA-Fine-tuned also improve. The MLDA-MDC provides its best performance as 87.22% when the dimensions are set to ($16 \times 16 \times 16$). BMPCA-Bottleneck features provide a good classification performance of 93.6%. Note that this result is achieved only using the simulation projection subspaces, proving the importance and merits of the initialization step. Finally, the nearly perfect classification accuracy belongs to the BMPCA-Fine-tuned at a rate of 97.2%. The confusion matrix performance for the test set given by BMPCA-Fine-tuned is shown in Table III. The primary source of misclassification is observed to be between the classes of sitting and bending, with about 3.6% confusion. This is not surprising as these two classes are most similar, and because our database contains different speeds of motions.

TABLE II

Classification Performance (on Test Data) Comparison for Different Dimensions of the RDC Between
MPCA-MDC, MLDA-MDC, BMPCA-Bottleneck, and BMPCA-Fine-Tuned in Terms of Accuracy

RDC Dimension	MPCA-MDC	MLDA-MDC	BMPCA (Bottleneck)	BMPCA (Fine Tuned)
16×16×16	90.56%	87.22%	89.14%	94.1%
32×32×32	90.94%	82.58%	92.56%	95.14%
64×64×64	91.4%	65.4%	93.6%	97.2%

TABLE III

Confusion Matrix of BMPCA-Fine-Tuned Features on the Test Data
(TA: 97.20%)

Activity-Class	Falling	Sitting	Bending	Kneeling	Walking
Falling	98.7	0	0	0	1.3
Sitting	2.3	94.1	3.6	0	0
Bending	1.3	2.4	95.1	1.2	0
Kneeling	0	0	0	100	0
Walking	2	0	0	0	98

VII. CONCLUSION

In this paper, we proposed an RDC-based multilinear subspace method for human activity recognition. Utilization of RDC offers an effective way to combine motion information from individual domains to capture cross correlations and interdependency. The proposed subspace method benefits from a single representation utilizing the entwined relationship between fast-time, slow-time, and Doppler frequency and their corresponding joint-variable domains. The proposed RDC-based approach starts with a preprocessing step called eCLEAN and attempts to suppress unwanted distortions and noise effects in the data. By employing MDC, we demonstrated that the multidimensional PCA method outperforms those based on predefined features, 1-D and 2-D PCA features, and a 12-layer CNN. We showed that improved performance by RDC-based MPCA under diverse data repository can be achieved by using different optimization techniques. We employed a transfer learning initialization and GAs-based optimization method to find the optimum number of projections. Furthermore, we introduced a boosting scheme with two supervised methods to enhance the classification performance attained by the RDC-based MPCA. The proposed approach provided a high classification accuracy of 97.2% for a 5-class activity recognition problem.

REFERENCES

[1]   J. A. Stevens, J. E. Mahoney, and H. Ehrenreich
      Circumstances and outcomes of falls among high risk
      community-dwelling older adults
      Injury Epidemiol., vol. 1, Mar. 2014, Art. no. 5.

[2]   M. G. Amin, Y. D. Zhang, F. Ahmad, and K. C. D. Ho
      Radar signal processing for elderly fall detection: The future
      for in-home monitoring
      IEEE Signal Process. Mag., vol. 33, no. 2, pp. 71–80, Mar.
      2016.

[3]   M. G. Amin
      *Radar for Indoor Monitoring: Detection, Classification,
      and Assessment*, 1st ed. Boca Raton, FL, USA: CRC Press,
      Sep. 2017. [Online]. Available: https://www.crcpress.com/
      Radar-for-Indoor-Monitoring-Detection-Classification-and-
      Assessment/Amin/p/book/9781138746091

[4]   B. Y. Su, K. C. Ho, M. J. Rantz, and M. Skubic
      Doppler radar fall activity detection using the wavelet trans-
      form
      IEEE Trans. Biomed. Eng., vol. 62, no. 3, pp. 865–875, Mar.
      2015.

[5]   V. C. Chen, F. Li, S. Ho, and H. Wechsler
      Micro-doppler effect in radar: Phenomenon, model, and simu-
      lation study
      IEEE Trans. Aerosp. Electron. Syst., vol. 42, no. 1, pp. 2–21,
      Jan. 2006.

[6]   L. Liu, M. Popescu, M. Skubic, M. Rantz, T. Yardibi, and P.
      Cuddihy
      Automatic fall detection based on Doppler radar motion sig-
      nature
      In *Proc. 5th Int. Conf. Pervasive Comput. Technol. Healthcare
      Workshops*, May 2011, pp. 222–225.

[7]   B. G. Mobasseri and M. G. Amin
      A time-frequency classifier for human gait recognition
      In *Proc. SPIE*, May 2009, vol. 7306, Art. no. 730628. [Online].
      Available: https://www.spiedigitallibrary.org/conference-
      proceedings-of-spie/7306/730628/A-time-frequency-classifier
      -for-human-gait-recognition/10.1117/12.819060.short

[8]   J. Zabalza, C. Clemente, G. Di Caterina, J. Ren, J. J. Soraghan,
      and S. Marshall
      Robust PCA micro-Doppler classification using SVM on em-
      bedded systems
      IEEE Trans. Aerosp. Electron. Syst., vol. 50, no. 3, pp. 2304–
      2310, Jul. 2014.

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

[10] P. Molchanov, J. Astola, K. Egiazarian, and A. Totsky
      Ground moving target classification by using DCT coefficients
      extracted from micro-Doppler radar signatures and artificial
      neuron network
      In *Proc. Microw., Radar Remote Sens. Symp.*, Aug. 2011,
      pp. 173–176.

[11] M. Otero
      Application of a continuous wave radar for human gait recog-
      nition
      In *Proc. SPIE*, 2005, vol. 5809, pp. 538–548.

[12] C. Clemente, L. Pallotta, A. D. Maio, J. J. Soraghan, and A.
      Farina
      A novel algorithm for radar classification based on Doppler
      characteristics exploiting orthogonal Pseudo-Zernike polyno-
      mials
      IEEE Trans. Aerosp. Electron. Syst., vol. 51, no. 1, pp. 417–
      430, Jan. 2015.

[13] R. Ricci and A. Balleri
      Recognition of humans based on radar micro-Doppler shape
      spectrum features
      IET Radar, Sonar Navigation, vol. 9, no. 9, pp. 1216–1223,
      2015.

[14] Y. Kim and T. Moon
      Human detection and activity classification based on micro-
      Doppler signatures using deep convolutional neural networks
      IEEE Geosci. Remote Sens. Lett., vol. 13, no. 1, pp. 8–12, Jan.
      2016.

[15] B. Jokanovic and M. Amin
Fall detection using deep learning in range-Doppler radars
IEEE Trans. Aerosp. Electron. Syst., vol. 54, no. 1, pp. 180–189, Feb. 2018.

[16] B. Jokanovic, M. Amin, F. Ahmad, and B. Boashash
Radar fall detection using principal component analysis
In *Proc. SPIE*, May 2016, vol. 9829, Art. no. 982919. [Online]. Available: https://www.spiedigitallibrary.org/conference-proceedings-of-spie/9829/982919/Radar-fall-detection-using-principal-component-analysis/10.1117/12.2225106.short

[17] J. Ye, R. Janardan, and Q. Li
GPCA: An efficient dimension reduction scheme for image compression and retrieval
In *Proc. 10th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2004, pp. 354–363. [Online]. Available: http://doi.acm.org/10.1145/1014052.1014092

[18] F. H. C. Tivive, S. L. Phung, and A. Bouzerdoum
Classification of micro-Doppler signatures of human motions using log-Gabor filters
IET Radar, Sonar Navigation, vol. 9, no. 9, pp. 1188–1195, 2015.

[19] A. K. Seifert, L. Schfer, M. G. Amin, and A. M. Zoubir
Subspace classification of human gait using radar micro-Doppler signatures
In *Proc. 26th Eur. Signal Process. Conf.*, Sep. 2018, pp. 311–315.

[20] A. Cichocki *et al.*
Tensor decompositions for signal processing applications: From two-way to multiway component analysis
IEEE Signal Process. Mag., vol. 32, no. 2, pp. 145–163, Mar. 2015.

[21] F. R. Dickey, M. Labitt, and F. M. Staudaher
Development of airborne moving target radar for long range surveillance
IEEE Trans. Aerosp. Electron. Syst., vol. 27, no. 6, pp. 959–972, Nov. 1991.

[22] S. Fukushima, H. Yamada, H. Kobayashi, and Y. Yamaguchi
Human motion detection and extraction by using FM-CW range-Doppler radar
In *Proc. Int. Symp. Antennas Propag. Conf.*, Dec. 2014, pp. 173–174. [Online]. Available: http://ieeexplore.ieee.org/document/7026586/

[23] D. Tahmoush and J. Silvious
Time-integrated range-Doppler maps for visualizing and classifying radar data
In *Proc. IEEE RadarCon*, May 2011, pp. 372–374.

[24] Y. He, P. Molchanov, T. Sakamoto, P. Aubry, F. L. Chevalier, and A. Yarovoy
Range-Doppler surface: A tool to analyse human target in ultra-wideband radar
IET Radar, Sonar Navigation, vol. 9, no. 9, pp. 1240–1250, 2015.

[25] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos
Multilinear Subspace Learning: Dimensionality Reduction of Multidimensional Data. London, U.K.: Chapman and Hall, 2012.

[26] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos
MPCA: Multilinear principal component analysis of tensor objects
IEEE Trans. Neural Netw., vol. 19, no. 1, pp. 18–39, Jan. 2008.

[27] S. Yan, D. Xu, Q. Yang, L. Zhang, X. Tang, and H. Zhang
Multilinear discriminant analysis for face recognition
IEEE Trans. Image Process., vol. 16, no. 1, pp. 212–220, Jan. 2007.

[28] M. S. Seyfioglu, B. Erol, S. Z. Gurbuz, and M. G. Amin
Diversified radar micro-Doppler simulations as training data for deep residual neural networks
In *Proc. IEEE Radar Conf.*, Apr. 2018, pp. 0612–0617.

[29] S. J. Pan and Q. Yang
A survey on transfer learning
IEEE Trans. Knowl. Data Eng., vol. 22, no. 10, pp. 1345–1359, Oct. 2010.

[30] K. Miettinen
Nonlinear Multiobjective Optimization. New York, NY, USA: Springer, 1999.

[31] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan
A fast and elitist multiobjective genetic algorithm: NSGA-II
IEEE Trans. Evol. Comput., vol. 6, no. 2, pp. 182–197, Apr. 2002.

Baris Erol (S'17) received the B.S. degree in electrical and electronics engineering and the M.S. degree in electrical engineering from TOBB University of Technology and Economics, Ankara, Turkey, in 2014 and 2015, respectively. He received the Ph.D. degree in electrical and computer engineering from the Villanova University, Villanova, PA, USA, in 2018.

He joined the Center for Advanced Communications, Villanova University in 2015 where he was a Research Assistant. He is currently with the Siemens Corporate Technology, Princeton, NJ, USA. His research interests include radar signal processing, machine learning, and data science.

Moeness G. Amin (F'01) received the B.Sc. degree from the Faculty of Engineering, Cairo University, Cairo, Egypt, in 1976, the M.Sc. degree in electrical engineering from King Fahd University of Petroleum and Minerals Dhahran, Saudi Arabi, in 1980, and the Ph.D. degree in electrical engineering from the University of Colorado, Boulder, CO, USA, in 1984.

Since 1985, he has been with the Faculty of the Department of Electrical and Computer Engineering, Villanova University, Villanova, PA, USA, where he became the Director of the Center for Advanced Communications, College of Engineering, in 2002. He has more than 750 journal and conference publications in signal processing theory and applications, covering the areas of wireless communications, radar, sonar, satellite navigations, ultrasound, healthcare, and RFID. He has coauthored 21 book chapters and is the Editor of 3 books: *Through the Wall Radar Imaging* (CRC Press, 2011), *Compressive Sensing for Urban Radar* (CRC Press, 2014), and *Radar for Indoor Monitoring* (CRC Press, 2017).

Dr. Amin is a Fellow of the Institute of Electrical and Electronics Engineers; a Fellow of the International Society of Optical Engineering; a Fellow of the Institute of Engineering and Technology; and a Fellow of the European Association for Signal Processing. He is a recipient of the 2017 Fulbright Distinguished Chair in Advanced Science and Technology; a recipient of the 2016 Alexander von Humboldt Research Award; a recipient of the 2016 IET Achievement Medal; a recipient of the 2014 IEEE Signal Processing Society Technical Achievement Award; a recipient of the 2010 NATO Scientific Achievement Award; a recipient of the 2009 Individual Technical Achievement Award from the European Association for Signal Processing; a recipient of the 2015 IEEE Aerospace and Electronic Systems Society Warren D White Award for Excellence in Radar Engineering; and a recipient of the 2010 Chief of Naval Research Challenge Award. He is the Recipient of the IEEE Third Millennium Medal. He was a Distinguished Lecturer of the IEEE Signal Processing Society from 2003 to 2004, and is the past Chair of the Electrical Cluster of the Franklin Institute Committee on Science and the Arts.