Abstract—Radar for indoor monitoring is an emerging area of research and development, covering and supporting different health and wellbeing applications of smart homes, assisted living, and medical diagnosis. Different human motion articulations present themselves more visibly in certain joint-variables data domains, most notably, time-frequency (TF) and range vs slow time. In this paper, we present a human motion data-driven classifier that utilizes both domains through a feature fusion approach. With data in each domain considered as an image, the features are extracted from lower dimension projections. These projections recognize the correlations across each image dimension, and are pursued using the generalized principal component analysis (GPCA). It is shown, through the confusion matrices, that feature fusion provides improved classification performance of human daily activities over the case where only the features of either domain are considered.

Index Terms—human activity recognition, micro-Doppler, radar, principal component analysis

I. INTRODUCTION

It has become evident from recent results and findings that the electromagnetic sensing modality plays an important role in monitoring normal and abnormal human activities indoor [1]–[5]. The safety and affordability of radars, combined with its ability to function irrespective of light and weather conditions, has made it one of the prime candidate technologies to be implemented in smart homes, assisted living facilities, and clinics. Research in this area is focusing on achieving high sensitivity and specificity consistent with the demands and expectations of both the providers and the recipients of the technology. Even a percentage or two improvement in classification performance can make a difference in how the technology is perceived and ultimately adopted.

So far, techniques of human motion classification using radar can be categorized into those based on handcrafted features and others that implement data driven feature learning [6]–[13]. The latter include principal component analysis (PCA) [14], autoencoders [15], convolution neural networks (CNNs) [16]–[18], convolutional autoencoders [17], and transfer learning [19]. Deep neural networks (DNNs) have shown great potential and may soon lead the way as a preferred motion classification choice. Issues with training data availability, processing speed and capabilities, and competitive performance with non-neural network (NN) approaches, however, need to be first and foremost resolved and thoroughly examined before such a choice is finally made.

PCA is an unsupervised data-driven dimensionality reduction (feature learning) method proposed to solve problems related to high dimensionality [20]. It reduces the dimensionality of an input dataset 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. On the other hand, since the principal components are uncorrelated, PCA can also be categorized as an automatic feature learning method, which has proven effective, offering high fall classification rates [1], [2], [10], [21]. PCA has been applied to different data representation domains and used to determine suitability of each domain for motion discrimination [22]. An extension of the PCA, called 2 dimensional (2D)-PCA, assumes that the inputs are 2D images and does not require a pre-processing vectorization step [23]. This method operates along the row direction of the input images where an image covariance matrix is calculated. In this work, we use another extension of the one dimensional (1D)-PCA called generalized PCA (GPCA) [20], [24]. GPCA is different from the traditional 2D-PCA methods in term of its underlying multiple subspace approach where it uses two subspaces to examine the different modes of the image. More details of the algorithm is provided in Section III.

Since the human body in motion exhibits different velocities that change over time, it is necessary to represent the data in such a way that the time-varying frequency characteristics of these radar returns are properly revealed and captured. In addition to the velocity information, target range and target motion translation extent can present useful parameters for activity recognition. Each human motion professes different levels of distinctions in different domains. We, therefore, combine information from range vs slow time domain (range maps) and spectrograms through fusing the outputs of their respective GPCA. We apply feature level fusion, where features of these individual domains are concatenated into a single feature vector, which is then given as an input to the k-nearest neighbors (kNN).

The paper is organized as follows. In Section II, the radar signal model and different joint variable domains are presented. In Section III, the GPCA method is introduced. In Section IV, the feature level domain fusion scheme is discussed. In Section V, experimental results are provided for performance demonstration. Finally, conclusions are drawn in Section VI.
II. RADAR SIGNAL MODEL AND SIGNAL REPRESENTATION

For a frequency modulated continuous wave (FMCW) radar, the backscattering signal from a target located at a distance \( R \) can be expressed as

\[
s_{\text{rx}}(t) = A_{\text{rx}} \cos \left( 2\pi f_0 (t - \tau) + \frac{r}{2} (t - \tau)^2 + \phi_{\text{rx}} \right),
\]

where \( \tau \) is the round trip time delay, \( f_0 \) is the carrier frequency, \( r \) is the chirp rate, \( \phi_{\text{rx}} \) is the phase of the received signal, and \( A_{\text{rx}} \) is the amplitude of the received signal computed from the range radar equation as

\[
A_{\text{rx}} = \frac{G \sqrt{P \sigma}}{(4\pi)^{1.5} R^2 \sqrt{L_s} \sqrt{L_a}}.
\]

Here, \( G \) is the antenna gain, \( P \) is the transmitter power, \( \sigma \) 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 processed by the I/Q demodulator, providing the in-phase and quadrature-phase components of the complex baseband signal, expressed as

\[
s(t) = I(t) + jQ(t) = A \exp(\psi(t)),
\]

where \( \psi \) is the signal phase. The sampled I/Q signal can be transformed into a 2D 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 (DFT) along the matrix columns. An example of the range maps for different human motion activities is depicted in Figure 1-(a) through (e).

Raw slow time data, \( x(n) \) where \( n \) is the slow time, can be provided by agglomeration of the range map along its range axis (columns). The resulting data associated with human in motion (individual being tracked) is non-stationary with time-varying frequencies which are associated with velocity, acceleration, and higher order terms of different human body parts. A joint-variable representation, in lieu of a single-variable representation whether it belongs to time or frequency, is a natural tool for revealing the local signal behavior and depicting its time-varying Doppler signatures [25]. Spectrograms are the simplest and most commonly used time-frequency (TF) distribution [26]. Spectrogram is the energetic form of the short-time Fourier transform (STFT). The latter is obtained by splitting the time domain signal into overlapping or disjoint consecutive segments, and then taking the DFT of each segment. The spectrogram is mathematically defined as

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

where \( h(m) \) is a window function. The spectrograms for different human motion activities are depicted in Figure 2-(a) through (e).

III. GENERALIZED PCA

In this section, we discuss the unsupervised GPCA algorithm. GPCA is a 2D extension of the 1D-PCA with the same objective function: maximizing the captured variation in the projected subspaces [27]. First, assume that the database contains \( M \) images \( \{S_1, S_2, ..., S_M\} \) which are available for training. Each image is denoted as \( S_m \in \mathbb{R}^{I_1 \times I_2} \) where \( I_1 \) and \( I_2 \) define the size of the image. In line with 1D-PCA, our objective is to find a matrix subspaces \( \tilde{U}^{(1)} \in \mathbb{R}^{I_1 \times P_1} \) and \( \tilde{U}^{(2)} \in \mathbb{R}^{I_2 \times P_2} \) that project the original tensor into a low dimensional matrix subspace \( Y_m \in \mathbb{R}^{P_1 \times P_2} \) (with \( P_1 \leq I_1 \) and \( P_2 \leq I_2 \)) defined as

\[
Y_m = S_m \times_1 \tilde{U}^{(1)T} \times_2 \tilde{U}^{(2)T},
\]

such that low dimensional subspace captures most of the variation observed in the original range map and spectrogram.
images. The objective function of the GPCA can be written as

\[
(\hat{U}^{(1)}, \hat{U}^{(2)}) = \arg \max_{U^{(1)}, U^{(2)}} \sum_{m=1}^{M} \|Y_m - \hat{Y}\|_F^2, \quad (6)
\]

where \(Y = \frac{1}{M} \sum_{m=1}^{M} Y_m\). The core matrix for each \(m\) samples can be obtained by projecting the original images using optimized subspaces, \(\hat{U}^{(1)}, \hat{U}^{(2)}\), as

\[
\hat{Y}_m = S_m \times_1 \hat{U}^{(1)T} \times_2 \hat{U}^{(2)T}. \quad (7)
\]

Finally, the feature vector of a training sample, \(m\), can be constructed as \(C_m = \text{vec}(\hat{Y}_m)\), \(\in \mathbb{R}^{1 \times D}\), where \(D = P_1 \times P_2\) and \(\text{vec}(\cdot)\) is the matrix column-wise vectorization operator. Note that, the dimensionality of the feature matrix, \(P_1\) and \(P_2\), is assumed to be known or predetermined. In this work, we employed the \(Q\)-based method, an extension of the traditional dimension selection in PCA. This metric depicts the total variation (energy) captured by a specified number of components. The \(Q^{(1)}\) and \(Q^{(2)}\) ratios are provided in Figure 3-(a) for spectrograms. For mode-1 and mode-2, 15 components are sufficient to capture 90% of the variation. The \(Q^{(1)}\) and \(Q^{(2)}\) ratios are provided in Figure 3-(b) for range maps. For mode-1 and mode-2, 8 components are enough to capture 90% of the variation. Therefore, for both range maps and spectrograms, the number of projections used in mode 1 and 2 are determined as 10 and 15, respectively.

IV. Domain Fusion

Fusion of domains is typically carried out at three different levels/architectures of abstraction closely interrelated with the flow of the classification process: data level, feature level, and decision level fusion. The selection among these architectures is application specific, and several considerations need to be made prior to implementing the fusion methods, such as characteristics of the domains and availability of computational resources. In this paper, we apply feature level fusion where the GPCA features from range maps and spectrograms are combined. In this scheme, GPCA operates on individual domains (range map and spectrogram) from which two feature vectors are extracted. Then, the respective feature sets are concatenated into a single feature vector which represents an input to the kNN classifier. The classifier output becomes, in essence, a joint or fused declaration of motion identity based on the combined feature vector from the two domains. Figure 4 depicts the feature level fusion scheme employed in this work.

V. Experimental Results

The radar system used in the experiments, named SDRKIT 2500B, is developed by Ancortek, Inc [28]. The center frequency is 25 GHz, whereas the bandwidth is 2 GHz which provides 0.075 m range resolution.

The FMCW radar signals were collected from two different indoor environments. The first location is the Radar Imaging Lab (RIL) located in CEER Building, where the radar was placed on a table raised 3.2 ft above the ground and pointing directly the back wall. The second dataset was acquired at the Center for Advanced Communications (CAC) conference room located in Tolentine Building. Both building are in Villanova University campus. 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 etc, was used. A Kinect was also located next to the radar in both configurations to record the ground truth optical videos of the experiments.

The combined dataset contained five human motions: falling, sitting, bending, kneeling and walking. Each motion was performed for 4 different directions, namely, \(0^\circ, 30^\circ, 45^\circ, 60^\circ\). Experiments were performed by 14 human subjects who posed...
Fig. 4: Proposed parallel processing and feature level fusion scheme.

TABLE I: Confusion matrix of only spectrogram-based GPCA features on the test data (TA: 82.24%)

<table>
<thead>
<tr>
<th>Activity-Class</th>
<th>Walking</th>
<th>Sitting</th>
<th>Bending</th>
<th>Kneeling</th>
<th>Falling</th>
</tr>
</thead>
<tbody>
<tr>
<td>Walking</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Sitting</td>
<td>0</td>
<td>70.4</td>
<td>18.5</td>
<td>7.4</td>
<td>3.7</td>
</tr>
<tr>
<td>Bending</td>
<td>0</td>
<td>11.1</td>
<td>81.5</td>
<td>7.4</td>
<td>0</td>
</tr>
<tr>
<td>Kneeling</td>
<td>0</td>
<td>11.1</td>
<td>11.1</td>
<td>74.1</td>
<td>3.7</td>
</tr>
<tr>
<td>Falling</td>
<td>0</td>
<td>7.4</td>
<td>7.4</td>
<td>0</td>
<td>85.2</td>
</tr>
</tbody>
</table>

TABLE II: Confusion matrix of only range map-based GPCA features on the test data (TA: 82.60%)

<table>
<thead>
<tr>
<th>Activity-Class</th>
<th>Walking</th>
<th>Sitting</th>
<th>Bending</th>
<th>Kneeling</th>
<th>Falling</th>
</tr>
</thead>
<tbody>
<tr>
<td>Walking</td>
<td>83.3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>16.7</td>
</tr>
<tr>
<td>Sitting</td>
<td>0</td>
<td>77.8</td>
<td>18.1</td>
<td>0</td>
<td>4.1</td>
</tr>
<tr>
<td>Bending</td>
<td>0</td>
<td>37.1</td>
<td>63.0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Kneeling</td>
<td>0</td>
<td>0</td>
<td>100.0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Falling</td>
<td>0</td>
<td>3.7</td>
<td>3.7</td>
<td>3.7</td>
<td>88.9</td>
</tr>
</tbody>
</table>

TABLE III: Confusion matrix of proposed fusion scheme on the test data (TA: 93.94%)

<table>
<thead>
<tr>
<th>Activity-Class</th>
<th>Walking</th>
<th>Sitting</th>
<th>Bending</th>
<th>Kneeling</th>
<th>Falling</th>
</tr>
</thead>
<tbody>
<tr>
<td>Walking</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Sitting</td>
<td>0</td>
<td>92.6</td>
<td>7.4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Bending</td>
<td>0</td>
<td>11.1</td>
<td>84.5</td>
<td>4.4</td>
<td>0</td>
</tr>
<tr>
<td>Kneeling</td>
<td>0</td>
<td>0</td>
<td>100.0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Falling</td>
<td>0</td>
<td>0</td>
<td>7.4</td>
<td>0</td>
<td>92.6</td>
</tr>
</tbody>
</table>

heights ranging from 5.1 to 6.3 ft, weights ranging from 119 to 220 lbs, and included 12 male and 2 female subjects.

It is frequently observed that the Doppler signatures for different events (falling and sitting) may become very similar in the TF domain, depending on the speed of the activity. Therefore, each subject performed the experiments with three different speeds: slow, medium, and fast, resulting in a database that covers a wide variety of motions with sufficient intra and inter class variance. From a total of 825 samples, 80% were used for training and remaining 20% were used for testing.

The test confusion matrices are provided for three different cases: spectrogram only (Table I), range map only (Table II), and proposed fusion scheme (Table III). It is observed that the proposed scheme provides the best test accuracy around 94% which is followed by 82% for both of the individual domain features. The primary sources of confusion for spectrogram and range map based GPCA features are sitting and bending respectively. The proposed fusion scheme identifies these classes at rate of 92.6% and 84.5%. Moreover, a high detection performance for fall (92.6%) is achieved by employing the fusion scheme. It should be emphasized that the data used in the experiment includes various motion orientations which render classification rather challenging.

VI. CONCLUSION

Stemming from the fact that different human motions are manifested with subtlety in different data domain representations, we applied in this paper a fusion of the information (features) extracted from the time-frequency (TF) and range vs slow time (range-map) domains. We employed data driven feature learning through the use of the generalized
principal component analysis (GPCA). This approach exploits
the data correlations over each variable of each of the two
representation domains and, as such, lowers dimensionality
to the principal components along range, slow time, and
frequency. These component features are fused though simple
concatenation scheme and applied to a k-nearest neighbors
(kNN) classifier. It was shown that the fusion step markedly
improves classification performance compared to the case
where the GPCA of each domain is considered separately.
The few percentage improvement in classification rates goes
a long way in cementing the offering of the radar monitoring
technology and preparing it for adoption in various industry
sectors.

REFERENCES
Signal Processing for Elderly Fall Detection: The future for in-home
80, Mar. 2016.
Available: https://www.crcpress.com/Radar-for-In-Door-Monitoring-
Detection-Localization-and-Assessment/Amin/p/book/9781138746091
outcomes of falls among high risk community-dwelling older adults,”
Injury Epidemiology, vol. 1, no. 5, Mar. 2014.
Diversity for Optimized Human Micro-Doppler Classification Using
Multistatic Radar,” IEEE Transactions on Aerospace and Electronic
Doppler-based in-home aided and unaided walking recognition with
multiple radar and sonar systems,” IET Radar, Sonar &amp; Navigation,
library.theiet.org/content/journals/10.1049/iet-rsn.2016.0055
Activity Detection Using the Wavelet Transform,” IEEE Transactions
[7] Y. Kim and H. Ling, “Human Activity Classification Based on Multi-
Doppler Signatures Using a Support Vector Machine,” IEEE Transac-
tions on Geoscience and Remote Sensing, vol. 47, no. 5, pp. 1328–1337,
May 2009.
target classification by using DCT coefficients extracted from micro-
Doppler radar signatures and artificial neuron network,” in 2011 MI-
CROWAVES, RADAR AND REMOTE SENSING SYMPOSIUM, Aug.
Gait Recognition,” The MITRE Corporation, Sep. 2013. [Online]. Avail-
bable: https://www.mitre.org/publications/technical-papers/application-of-
a-continuous-wave-radar-for-human-gait-recognition
[10] F. H. C. Tivive, S. L. Phung, and A. Bouzerdoum, “Classification of
micro-Doppler signatures of human motions using log-Gabor filters,”
detection based on Doppler timefrequency signatures for assisted living,”
[12] L. Liu, M. Popescu, M. Skubic, M. Rantz, T. Yardibi, and P. Cuddihy,
“Automatic fall detection based on Doppler radar motion signature,” in
2011 5th International Conference on Pervasive Computing Technolo-
gies for Healthcare (PervasiveHealth) and Workshops, May 2011, pp.
222–225.
“Pseudo-Zernike moments based radar micro-Doppler classification,” in
recognition/10.1117/12.819060.short
using deep learning,” in 2016 IEEE Radar Conference (RadarConf),
May 2016, pp. 1–6.
Transactions on Knowledge and Data Engineering, vol. 22, no. 10,
Methods for Micro-Doppler Classification With Low Training Sample
Based on Micro-Doppler Signatures Using Deep Convolutional Neural
Networks,” IEEE Geoscience and Remote Sensing Letters, vol. 13, no. 1,
pp. 8–12, Jan. 2016.
Based Classification of Human Aquatic Activities via Transfer
Learning of Convolutional Neural Networks,” Sensors (Basel,
linear Principal Component Analysis of Tensor Objects,” IEEE Transac-
classification of human gait using radar micro-doppler signatures,” in
[22] B. Jokanovic and M. Amin, “Suitability of Data Representation Domains
in Expressing Human Motion Radar Signals,” IEEE Geoscience and
a new approach to appearance-based face representation and recogni-
tion,” IEEE Transactions on Pattern Analysis and Machine Intelligence,
https://link.springer.com/article/10.1007/s10994-005-3561-6
[26] B. Boashash, Time-Frequency Signal Analysis and Processing: A Com-
Reduction Scheme for Image Compression and Retrieval,” in Proceed-