Phase-based Classification for Arm Gesture and Gross-Motor Activities using Histogram of Oriented Gradients

Guendel, Ronny Gerhard; Fioranelli, Francesco; Yarovoy, Alexander

DOI
10.1109/JSEN.2020.3044675

Publication date
2020

Document Version
Accepted author manuscript

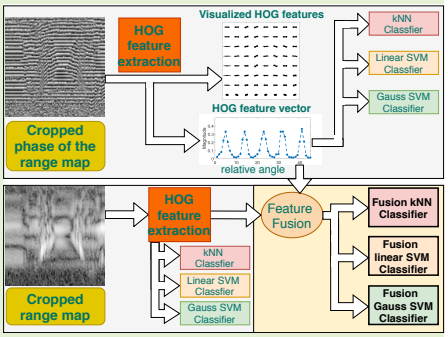
Published in
IEEE Sensors Journal

Citation (APA)
Guendel, R. G., Fioranelli, F., & Yarovoy, A. (2020). Phase-based Classification for Arm Gesture and Gross-Motor Activities using Histogram of Oriented Gradients. *IEEE Sensors Journal*, *21*(6), 7918 - 7927. Article 9293354. https://doi.org/10.1109/JSEN.2020.3044675

Phase-based Classification for Arm Gesture and Gross-Motor Activities using Histogram of Oriented Gradients

Ronny Gerhard Guendel, *Student Member, IEEE*, Francesco Fioranelli, *Senior Member, IEEE*, and Alexander Yarovoy, *Fellow, IEEE*

***Abstract*— Micro-Doppler spectrograms are a conventional data representation domain for movement recognition such as Human Activity Recognition (HAR) or gesture detection. However, they present the problem of time-frequency resolution trade-offs of Short-Time Fourier Transform (STFT), which may have limitations due to unambiguous Doppler frequency, and the STFT computation may be onerous in constrained embedded environments. We propose in this paper an alternative classification approach based on the radar phase information directly extracted from high-resolution Range Map (RM). This novel approach does not use the aforementioned micro-Doppler processing, and yet achieves equivalent or even superior classification results. This shows a potential advantage for low-latency, real-time applications, or computationally constrained scenarios. The proposed method exploits the Histogram of Oriented Gradients (HOG) algorithm as an effective feature extraction algorithm, specifically its capability to capture the unique shape and patterns present in the wrapped phase domains, such as their contour intensity and distributions. Validation results consistently above 92% demonstrate the effectiveness of this method on two independent datasets of arm gestures and gross-motor activities. These were classified with three algorithms, namely the Nearest Neighbor (NN), the linear Support Vector Machine (SVM), and the Gaussian SVM classifiers using the proposed phase information. Feature fusion of different data domains, e.g. the modulus of the RM fused with the RM phase information, is also investigated and shows classification improvement specifically for the robustness of activity performances, such as the aspect angle and the speed of performance.**



***Index Terms*— Micro-Doppler radar, assisted living, range map, phase, classification, Histogram of Oriented Gradients (HOG), feature fusion, human activity recognition (HAR)**

I. Introduction

The amount of research on human activity recognition (HAR) with radar sensors has tremendously increased over the past decades, with significant progress made in almost every area related to activities of daily living (ADL) [1], [2]. Closely associated are the areas of gesture and arm motion recognition which attracted interest for their potential for remote control of smart devices [3]–[6]. This field has seen the development of many different classification approaches, including those inspired by deep learning techniques, such as Recurrent Neural Networks (RNN) with their bidirectional implementations known as Bidirectional Long Short-Term Memory (BI-LSTM) [7], frameworks to generate synthetic radar signatures via Generative Adversarial Networks (GANs) [8], and effective cross-frequency training for multiple radar sensors used for HAR [9]. These techniques for HAR and gesture classification also include multimodal frameworks where different sensing modalities can be combined together with radar. For example, in recent studies, magnetic induction systems and more in general wearables are also used for HAR in conjunction with radar applications [10], [11].

Nonetheless, the majority of research work in radar for HAR and gestures has focused mostly on the modulus (magnitude) of the micro-Doppler (μD) spectrogram, and in part on the Range Map (RM), the range-Doppler (RD), or the Range-Doppler-Surface (RDS) as radar data domains to start the classification process [12]–[16].

In this paper, we propose a different and innovative approach based on the usage of the phase information directly extracted from complex high-resolution RM matrices. To the best of our knowledge, this data domain has been very marginally explored for radar-based HAR and gesture classification, whereas other researchers have for example applied Phase Unwrapping Techniques (PUT) on the phase of the μD spectrograms [17]. In this work we compare conventional radar data domains, such as the μD spectrograms and RM, with different formats of the proposed phase-based domain information, namely the phase of the μD spectrogram and the phase of the RM. In both cases, their original form (which includes all recorded range bins) and a ”cropped” form (which only considers a spatial window, e.g. of approximately 1m around the detected target for the RM) are considered.

Different features to be extracted from the aforementioned

data domains are investigated. We tested that conventional well-performing feature extracted by the Two-Dimensional (2-D) Principal Component Analysis (PCA) were not able to capture the relevant information for HAR and gesture classification from the unique shapes in the phase of the RM matrices. On the contrary, features derived from the Histogram of Oriented Gradients (HOG) technique proved to be suitable when applied to phase matrices, as they are capable to capture the salient patterns in terms of strength and orientation of the typical "line structures" in such plots, while still retaining a relatively simple mathematical formulation compared to less easily explainable convolutional neural networks.

The paper is organized as follows. In Section II, the radar signal model along with the HOG feature extraction method and the data domains are presented. In Section III, the experimental setup is outlined together with the detailed results on the gross-motor activities and the arm gesture datasets. Finally, concluding remarks are provided in Section IV.

II. Signal Representation and Feature Selection

In this section the different radar data domains are presented, specifically the phase domain information to be used in combination with the HOG features for HAR and arm gesture classification.

A. Radar data representation

$$S_{\mathrm{mn}} = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1\mathrm{n}} \\ s_{21} & \ddots & & s_{2\mathrm{n}} \\ \vdots & & \ddots & \vdots \\ s_{\mathrm{m}1} & s_{\mathrm{m}2} & \cdots & s_{\mathrm{mn}} \end{bmatrix}$$

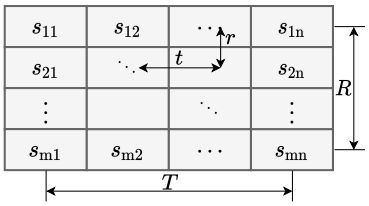


Fig. 1: Matrix representation for the received radar signal of the Humatics P410 radar.

The Humatics (former PulsON) P410 radar provides the Range Map (RM), S_{mn}, as shown in Fig. 1, where, s_{mn}, represents the real samples (in-phase components) for the individual range and slow time bins with index m and n, respectively. Each range bin in the m-direction has a time separation, τ, of 61.024ps. The resulting range bin resolution, r, is computed as follows by, $r = \frac{\tau \cdot c}{2} = 9.153\mathrm{mm}$. This leads to an area of coverage, R, of 4.39m with 480 received range bins, $\mathrm{m} = 1, \cdots, \mathrm{M}$. The slow time samples, n, are indicated by, $\mathrm{n} = 1, \cdots, \mathrm{N}$, for the total time T [18]. Cleaning steps, such as MTI filtering, mean subtraction, and adaptive thresholding are applied to generate the RM, indicated as *RM-O* in Fig. 2.

B. Range map and phase angle representation

As this radar provides only the in-phase components as presented in the matrix S_{mn} in Fig. 1, a common practice is using the Hilbert transform for reconstructing the complex signal along the range (R). The process is repeated across each column vector $[s_{1n}, s_{2n}, \cdots, s_{mn}]^T$ for $\mathrm{n} = 1, \cdots, \mathrm{N}$

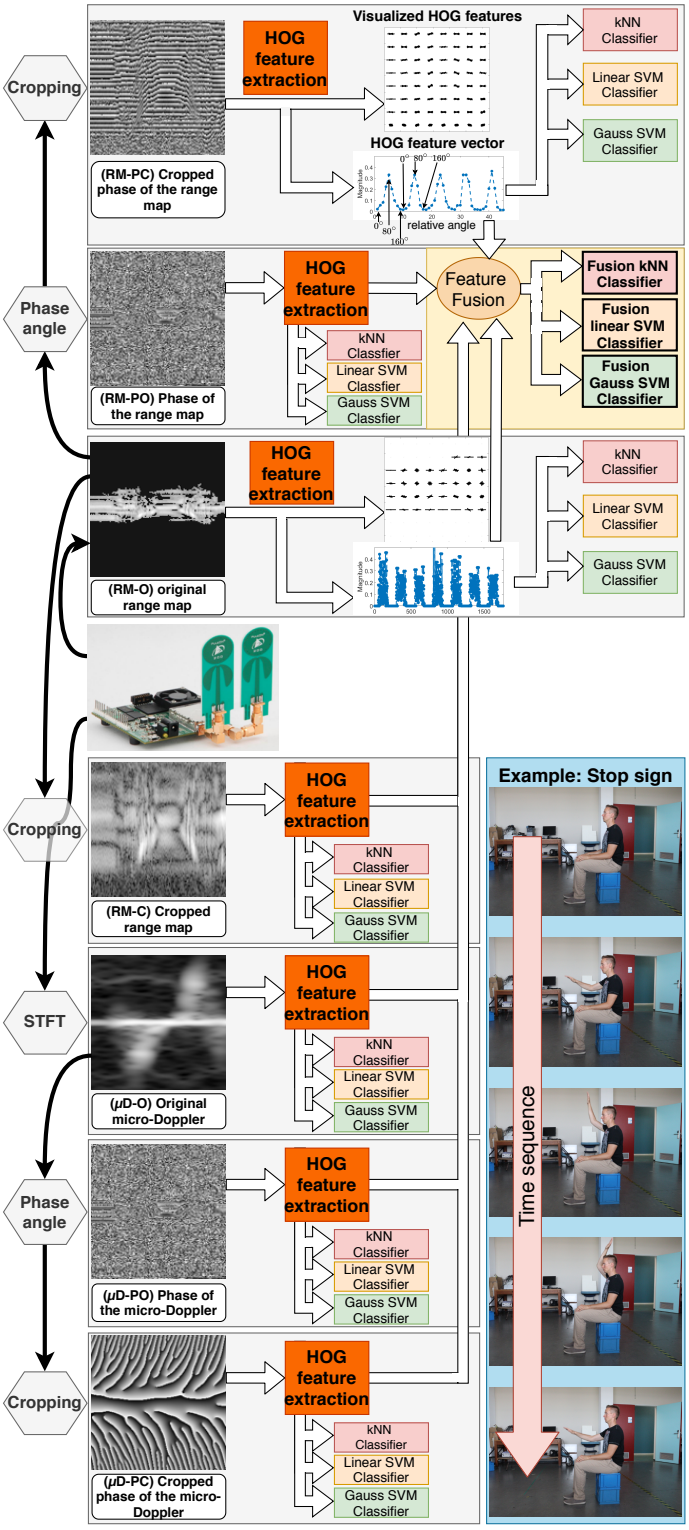


Fig. 2: Schematic representation of the feature extraction by the Histogram of Oriented Gradients (HOG) algorithm for the individual data domains, with feature fusion shown in the yellow box.

[19]. The Hilbert transform creates a complex-valued causal function from the purely real-valued range profile with the property of a $\frac{\pi}{2}$ (90°) phase shift, such that $\hat{S}_{\mathrm{mn}} = Hil(S_{\mathrm{mn}}) =$

$Re(\hat{S}_{\text{mn}}) + i \cdot Im(\hat{S}_{\text{mn}})$ with $i = \sqrt{-1}$. Now, $\hat{S}_{mn}$ represents the complex signal matrix of the RM with in-phase samples, $Re(\hat{S}_{\text{mn}})$, and quadrature samples, $Im(\hat{S}_{\text{mn}})$ [20], [21].

From the Euler representation of complex numbers, the phase angle ϕ_{mn} of the signal $\hat{S}_{\text{mn}}$ can be computed as $\phi_{\text{mn}} = \angle\hat{S}_{\text{mn}}$, with ϕ_{mn} in the value domain of $\phi_{\text{mn}} \in \{\mathbf{R}| - \pi < \phi_{\text{mn}} \leq \pi\}$. In fact, the matrix ϕ_{mn} has the same range and slow time resolution as S_{mn} and $\hat{S}_{\text{mn}}$. The resulting RM phase plot can be seen in Fig. 2 indicated as, *RM-PO*, originating from the phase angle block in the flow chart.

We introduce the "cropped" RM, since studies have shown that the arm span to body height ratio is between 0.98 and 1.08, so that the range stretch of an arm towards the radar is approximately $^1/_2 \times$ bodyheight, which is roughly equal to $^1/_2 \times$ armspan. Accounting for this, the tallest test person in the dataset with 1.84cm height can stretch their arms at a maximum of about 92cm towards the radar [22]. To capture effectively the span of all motions, also including possible torso movements when performing fast gesture motions, we are capturing 20% beyond the expected maximum range which results in 1.10m. In other words, a "cropped" version of the RM matrices and their phases is considered using this spatial window centred on the target range. In this paper, the location of the subject to perform the cropping operation is provided by the Derivative Target Line (DTL) [23] which can determine the person's distance to the radar (the DTL could also be replaced by other suitable target trackers [24]). This cropped phase plot is shown in Fig. 2 as *RM-PC*, originating from the cropping block of the flow chart. The cropped window of 1.10m is also applied to the original RM, with an example shown in Fig. 2 and denoted as *RM-C*.

C. Spectrogram representation

From the Hilbert transformed signal, $\hat{S}_{\text{mn}}$, the Fast Fourier Transform (FFT) across each scan is computed. Then, the Short-Time Fourier Transform (STFT) is applied to the vector of the 4GHz frequency for the computation of the μD spectrogram [25], [26]. Using a f_{PRF} of 121.95Hz yields to an unambiguous Doppler frequency of ± 60.97Hz. Thus, the unambiguous velocity is ± 2.17m/s and is computed as $\pm v_{un} = c_0 \cdot f_{\text{PRF}}/(4 \cdot f_o)$, with, c_0, the speed of light and, f_0, the center frequency of 4.2GHz (operational frequency band: 3.1–5.3GHz) [14]. An example of the μD is shown in Fig. 2 as *μD-O*. The phase angle and the cropped phase angle of the μD are computed from the RM explained in Sec. II-B. The phase information of the micro-Doppler spectrograms are also computed. The original phase angle and the cropped phase angle of the μD can be seen in Fig. 2 as *μD-PO* and *μD-PC*, respectively. Specifically, the *μD-PC* is resized by a factor 0.25 in the cropping process, leading to a maximum Doppler frequency extent of ± 15.24Hz around the OHz Doppler bin.

To summarize, the data representation domains in Fig. 2 are:

RM-PC	Cropped phase of the range map
RM-PO	Phase of the range map
RM-O	Original range map
RM-C	Cropped range map
μD-O	Original micro-Doppler
μD-PO	Phase of the micro-Doppler
μD-PC	Cropped phase of the micro-Doppler

D. Histogram of Oriented Gradients (HOG)

The Histogram of Oriented Gradients (HOG) is a powerful tool for edge and contour detection and has been widely used in the computer vision and optical character recognition fields due to the ability to characterize strength and regularities of line patterns and contours in images [27]. This method first determines the gradients, g_x, and g_y, by the partial derivative as, $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$, so that the gradient vector is defined as

$$\nabla f(x,y) = \begin{bmatrix} g_x \\ g_y \end{bmatrix} = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix} = \begin{bmatrix} f(x+1,y) - f(x-1,y) \\ f(x,y+1) - f(x,y-1) \end{bmatrix} \tag{1}$$

for a matrix $f(x,y)$, where x and y represent the individual samples or pixels. From the $\nabla f(x,y)$ two important attributes are extracted, specifically

- The magnitude of the vector by computing the L_2-norm as, $g = \|\nabla f(x,y)\|_2 = \sqrt{{g_x}^2 + {g_y}^2}$
- The directional orientation as, $\theta = \arctan\left(\frac{g_y}{g_x}\right)$

It is noted that $f(x,y)$ defines only an area of the whole image matrix, where in our experiments the HOG sizes are $[16,16]$ or $[32,32]$ pixels. Examples of the visualized HOG features are shown in Fig. 3 on the bottom, where the top shows the related *RM-PC* matrices. Specifically, the contours in the *RM-PC* appear to be mainly horizontally orientated with an approximate variance of up to $\pm 45°$. Hence, vectors which are vertically orientated have almost zero length, which relates to the magnitude of g. The extracted features are represented by histograms, examples are shown in the flow chart of Fig. 2 for *RM-PC* and *RM-O* and denoted as "HOG feature vector". Then, the "HOG feature vectors" are the input for the classifiers [27]–[30].

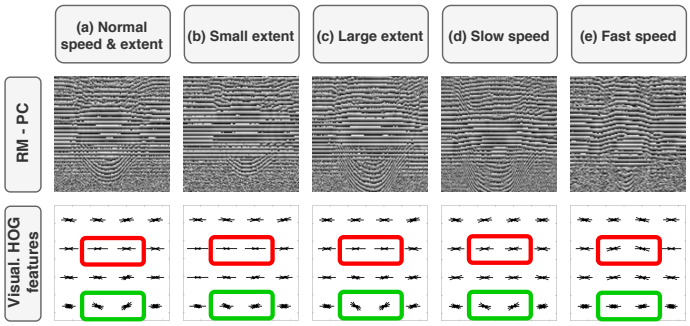


Fig. 3: The impact on the HOG features illustrated for the extent and speed variation for the *push and pull arm* activity.

Back to Fig. 3, we show the HOG features for the activity of *push and pull arms*. Here, we illustrate the differences in performing the motion with (b) small arm extent, (c) large arm extent, (d) slow speed, (e) fast speed, and compare those to the (a) normal performed motion. Specifically, two regions where deliberately selected, the red and green circled areas to characterize the features. Body movements occur typically in the middle regions, as those marked with the red

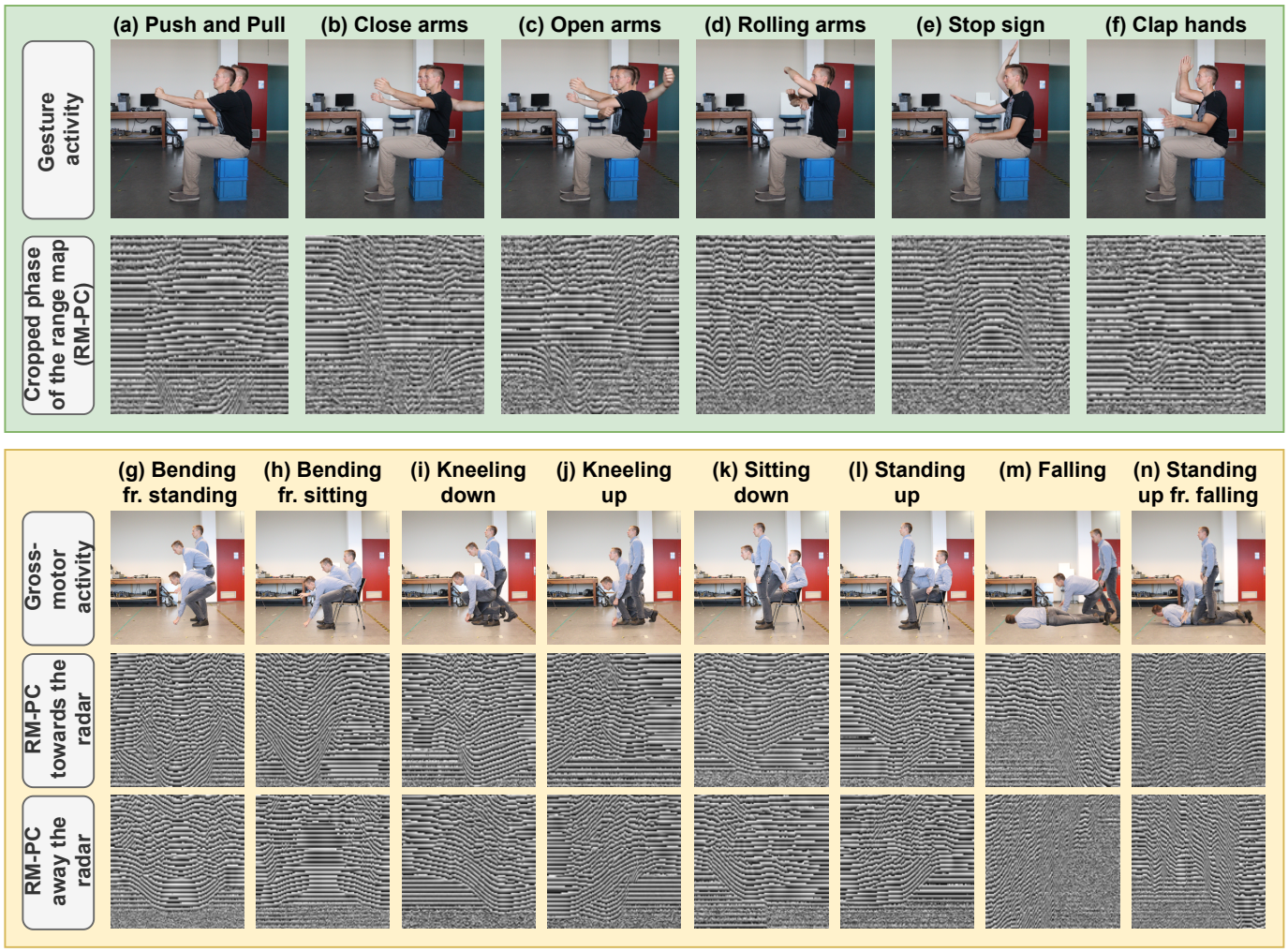


Fig. 4: Pictures of the activities and related cropped phase matrices: six gestures (top, green-shaded), and eight gross-motor activities performed bidirectionally (bottom, yellow-shaded).

rectangles. It can be seen that the small extent leads to little body movements, since the torso is mainly static on a fixed position, thus resulting in mostly horizontal HOG vectors. Furthermore, the green framed areas contain HOG vectors mostly originating from the arm movement towards the radar. As a result by performing the activities with large extents, the HOG feature vectors become more steeply orientated and diagonal in contrast to the small extent case. Another important difference can be seen by performing the motion with fast speed, where the inertia of the torso and the arms lead inevitably to a larger backward movement of the torso than for slower speed. As a result, the HOG features emphasize the backward movement of the torso in a very distinctive way.

E. Data representation and classification

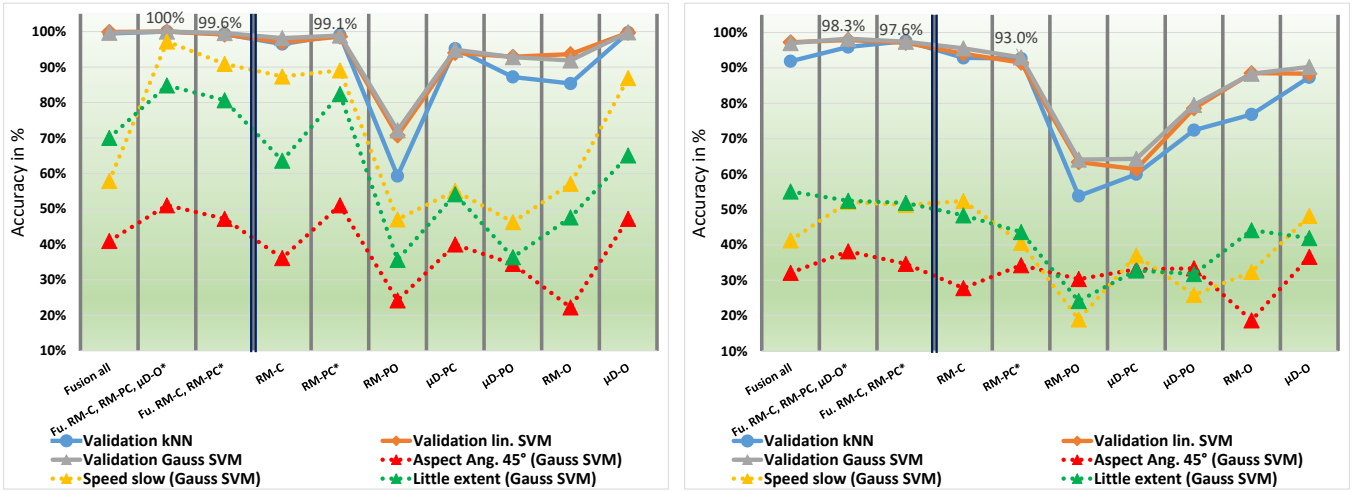
In the previous section, the HOG feature extraction was explicitly described. From here, the orientated gradients are collected over the 2D detection windows (specifically $[16, 16]$ or $[32, 32]$ pixels) to provide the histograms and the feature vector, while the orientation is discretized in histogram angular bins of $20°$ from $0°$ to $160°$. It is noted that vectorial gradients are proportionally split into the histogram bins if their orientation value is between the bins' nominal values. Furthermore, HOG only considers gradients from $0°$ to $180°$ since a contour in an image is nondirectional. Also, the histogram bin of $180°$ does not exist since it is equivalent to the $0°$ bin. This processed feature vector is then used for classification.

In this paper, we test our proposed method with few of the most common yet effective supervised learning classifiers, the Nearest Neighbor (NN), the linear Support Vector Machine (linear SVM), and the Gaussian SVM classifier. Specifically, the NN classifier was used with a number of neighbors of five with Euclidean distance computation. For the linear SVM and the Gaussian SVM classifier, we apply the multi-class setting one-versus-one.

The yellow-shaded square box in Fig. 2 shows that we also apply feature fusion for classification. For that, the individual feature vectors stemming from the individual radar data domains are joined into a concatenated feature vector, e.g., $\kappa_{\mathrm{Fu_{all}}}$, which is expressed as $\kappa_{\mathrm{Fu_{all}}}=[\kappa^{\mathrm{T}}_{\mathrm{RM\text{-}PC}}, \kappa^{\mathrm{T}}_{\mathrm{RM\text{-}PO}}, \ldots, \kappa^{\mathrm{T}}_{\mu\mathrm{D\text{-}PC}}]^{\mathrm{T}}$. We show that concatenating all possible feature vectors from different radar domains does not lead to the best classification results. In this regard, the best accuracy was achieved by using

TABLE I: Average performance of gesture experiments for two different f_{PRF} for the feature fusion case (3 left side columns) and individual domains (7 right side columns). Average test performance presented for AA of 45°, a slower speed, and a smaller extent.

Features / Classifier	Fusion all	Fu. RM-C, RM-PC, μD-O	Fu. RM-C, RM-PC	RM-C	RM-PC	RM-PO	μD-PC	μD-PO	RM-O	μD-O
Validation performance for f_{PRF}=122Hz, HOG=[16,16]										
kNN=5 (euclidean)	99.46	100.00	99.29	96.49	99.05	59.29	95.18	87.20	85.36	99.70
SVM linear	99.88	100.00	99.17	96.96	98.57	70.71	93.99	92.92	93.63	99.70
SVM Gauss	99.64	100.00	99.58	98.16	98.87	72.14	94.88	92.80	91.85	99.76
Test performance for SVM Gaussian classifier, f_{PRF}=122Hz, HOG=[16,16]										
(1) Aspect angle: 45°	40.96	51.04	47.21	36.15	51.04	24.21	39.98	34.50	22.23	47.21
(2) Speed: slow	57.87	97.22	90.90	87.35	89.04	47.07	55.09	46.30	57.10	86.88
(3) Extent: small	70.04	84.83	80.64	63.63	82.37	35.64	54.13	36.38	47.60	65.11
Validation performance for f_{PRF}=12.2Hz, HOG=[16,16]										
kNN=5 (euclidean)	91.89	95.92	97.60	92.84	92.67	53.80	59.96	72.43	76.85	87.36
SVM linear	97.26	97.88	97.09	94.02	91.50	63.37	61.41	78.58	88.54	88.37
SVM Gauss	97.04	98.27	97.43	95.53	92.95	64.09	64.32	79.53	88.37	90.27
Test performance for SVM Gaussian classifier, f_{PRF}=12.2Hz, HOG=[16,16]										
(1) Aspect angle: 45°	32.09	38.18	34.64	27.77	34.25	30.32	33.07	33.37	18.65	36.61
(2) Speed: slow	41.24	52.12	51.27	52.40	40.54	18.93	37.01	25.85	32.35	48.16
(3) Extent: small	55.08	52.46	51.86	48.34	43.62	24.12	32.76	31.76	44.12	41.91



(a) Performance with f_{PRF}=122Hz, HOG=[16,16]

(b) Performance with f_{PRF}=12.2Hz, HOG=[16,16]

Fig. 5: Validation and test performance of gesture experiments for the different f_{PRF} related to Tab. I. Red, yellow, green data labels for testing performance match in color those in Tab. I.

a subset of features, namely, the cropped phase of the RM (*RM-PC*), the cropped RM (*RM-C*), and the μD spectrogram (*μD-O*), so that the concatenated feature vector is formed as $\kappa_{\mathrm{Fu_{best}}}=[\kappa^{\mathrm{T}}_{\mathrm{RM\text{-}PC}}, \kappa^{\mathrm{T}}_{\mathrm{RM\text{-}C}}, \kappa^{\mathrm{T}}_{\mu\mathrm{D\text{-}O}}]^{\mathrm{T}}$.

On the other hand, very promising results were achieved by excluding the μD spectrogram and focusing deliberately on the RM which is directly provided by the radar, so that an additional STFT calculation or even more complex time-frequency distributions can be omitted. Computing the μD spectrogram requires some computational resources and time. Considering

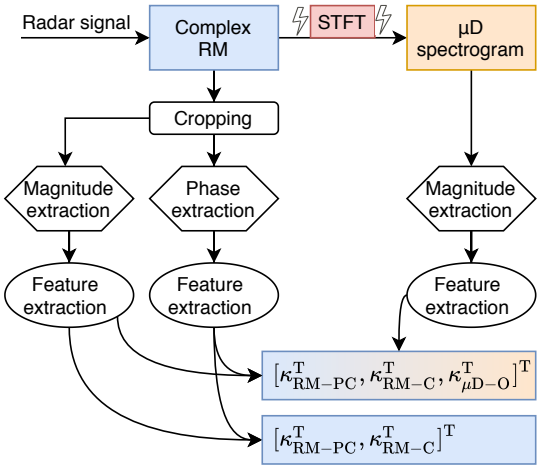


Fig. 6: The revealing scheme to extract the features from the range only as, $\kappa_{\mathrm{Fu_{range}}}$, and with containing the μD spectrogram feature as, $\kappa_{\mathrm{Fu_{best}}}$, which implies the STFT.

the STFT as the simplest approach to calculating spectrograms via time-frequency analysis, a measure of its complexity as the number of floating-point operations (FLOPS) can be computed as,

$$\text{with: } k = \frac{N-L}{n-L} \quad k \cdot log_2(n^n) \tag{2}$$

with n the STFT window function length, L the overlap length, and N the signal length. By using the STFT with the Discrete Fourier Transform (DFT) the FLOPS from Eq. 2 changes to kn^2 [31], [32]. It is noted, the simplified FLOP calculation does not consider additional smoothing window multiplication, i.e., the Hamming window. In this respect, avoiding the STFT leads to the concatenated feature vector consisting of $\kappa_{\mathrm{Fu_{range}}}$=$[\kappa^{\mathrm{T}}_{\mathrm{RM\text{-}PC}}, \kappa^{\mathrm{T}}_{\mathrm{RM\text{-}C}}]^{\mathrm{T}}$, and includes only the cropped RM and the cropped phase of the RM. The scheme is illustrated in Fig. 6.

III. Experimental results

Two data sets were collected in the radar laboratory at the Delft University of Technology (TU Delft), consisting of a comprehensive number of classes for the gesture and the gross-motor experiments that are presented in this section. Both sets were recorded with Humatics (former PulsON) P410 pulsed radar systems. Four participants were involved in the experimental data collection, with a height between 1.65m and 1.84m, and a weight between 65kg and 86kg. The amount of data samples for the training sets is 280, 282, and 300 samples per class for the gross-motor activities with a f_{PRF} of 12.2Hz, and the gesture activities with a f_{PRF} of 12.2Hz, and 122Hz, respectively.

The test set for gross-motor activities includes 120 samples by considering an aspect angle of $45°$. The amount of test data for the gesture activities for both f_{PRF} amounts to 94, 78, and 84 samples for the aspect angle of $45°$, the slow speed, and the small extent, respectively.

Each of the considered radar data domain (e.g. *RM-O*, *RM-PC*, or *μD-O*) was resized to a matrix size of 128x128, from which further processing extracts the HOG feature vectors (κ_{i}), with, i, the seven different domains as listed in Section II-C. The collection of the training and validation data was performed under controlled aspect angle, spatial extent, and speed of the movement. For further validation, we collected a separate test set for the gesture activities with (1) an aspect angle (AA) of about $45°$, (2) a slower speed, and (3) a smaller extent as shown in Tab. I. The training/validation data of the gross-motor activities was collected with the same conditions, whereas the test set contains only data with aspect angle of $45°$, since a slower speed or a smaller extent can be difficult for some gross-motor activities, e.g. falling. It is noted that the training/validation set does not include data samples which are reflecting the test set conditions (AA, slower speed, or smaller extent). In fact, the classifier is requested to classify data without being explicitly trained in those conditions.

A. Arm gesture results

The first data set consists of six gesture activities, namely, (a) push and pull arms, (b) close arms, (c) open arms, (d) rolling arms, (e) stop sign, and (f) clap hands. The individual gesture activities and the related *RM-PC* are shown in the green-shaded box in Fig. 4. The movements were performed facing the radar with a body center distance of 3.20m to the radar. For performance comparison, we collected two subsets with the f_{PRF} of 122Hz and 12.2Hz.

The validation and test performance of the gesture activities can be found in Tab. I. Specifically, the test data set includes three different cases, namely (1) higher Aspect Angle (AA), (2) slower speed, and (3) smaller spatial extent compared to the training/validation data. The results are visualized in Fig. 5 related to Tab. I. In the table, we show the test performance of the gesture activities only for the Gaussian SVM classifier which gives the best results and outperformed the kNN and the linear SVM classifier.

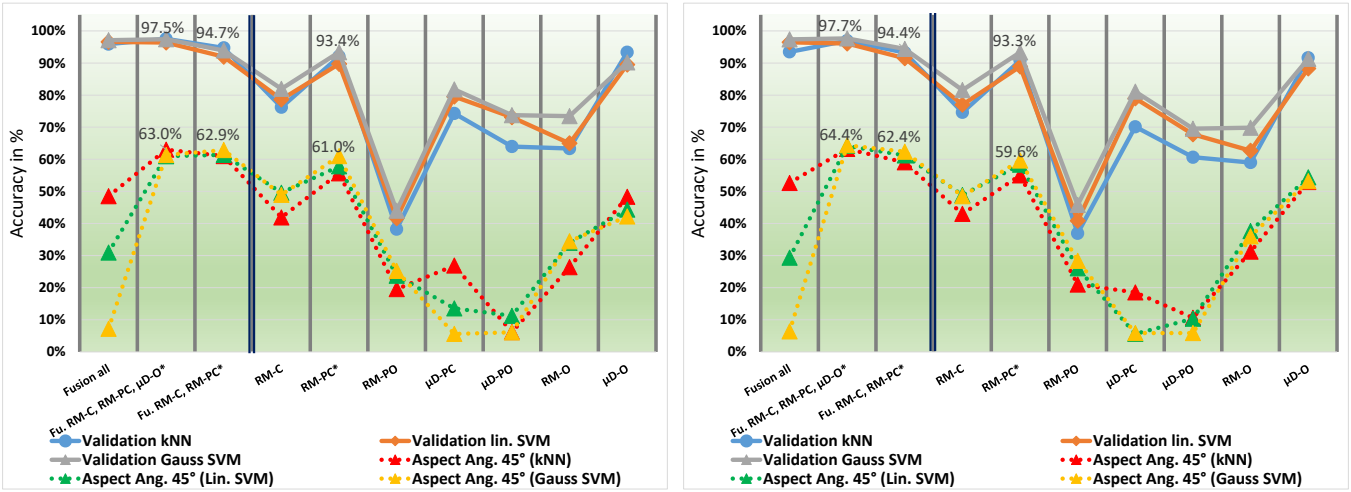
Remarkably, we illustrate that our proposed method of the phase of the RM (*RM-PC)* and HOG features is still able to classify gestures with sufficient accuracy when the unambiguous Doppler frequency of the μD spectrogram decreases by a factor 10x by using the lower f_{PRF}=12.2Hz instead of the f_{PRF}=122Hz. As expected, because of the resulting ambiguity in micro-Doppler, the classification accuracy decreases when using the *μD-O* domain by more than 9% with the best-performing classifier, Gaussian SVM ($99.76\% \rightarrow 90.27\%$). The RM and especially the *RM-PC* is almost unaffected by lowering the radar f_{PRF} by such significant amount. Regarding the HOG feature extraction, the highest classification results were obtained by using a HOG cell size of [16,16] samples. Other HOG cell sizes of [8,8] or [32,32] have also been tested, but provided lower classification results.

B. Gross-motor activities results

The second data set contains an even larger number of classes, while the activities were performed away from the radar in addition to facing the radar. The activities are (g)

TABLE II: Average performance of gross-motor experiments for two different HOG cell sizes for the feature fusion case (3 left side columns) and individual domains (7 right side columns). Average test performance presented for 45° AA for the three different classifiers kNN, linear SVM, and Gaussian SVM.

Features / Classifier	Fusion all	Fu. RM-C, RM-PC, µD-O	Fu. RM-C, RM-PC	RM-C	RM-PC	RM-PO	µD-PC	µD-PO	RM-O	µD-O
Validation performance for f_{PRF}=122Hz, HOG=[16,16]										
kNN=5 (euclidean)	95.89	97.53	94.72	76.21	92.04	38.24	74.30	63.98	63.35	93.37
SVM linear	96.61	96.38	91.95	78.67	89.70	41.66	79.54	73.11	64.97	89.50
SVM Gauss	97.12	97.42	93.91	81.97	93.39	44.00	81.79	73.74	73.45	90.20
Test performance for aspect angle (AA): 45° f_{PRF}=122Hz, HOG=[16,16]										
AA: 45° (kNN)	48.51	63.01	61.01	41.89	55.63	19.56	26.90	6.18	26.41	48.33
AA: 45° (lin. SVM)	30.95	60.96	61.41	49.58	57.80	23.74	13.56	11.25	34.02	44.33
AA: 45° (Gauss SVM)	7.20	61.36	62.87	48.96	61.01	25.26	5.60	6.09	34.46	42.20
Validation performance for f_{PRF}=122Hz, HOG=[32,32]										
kNN=5 (euclidean)	93.46	96.94	93.23	74.62	90.81	36.96	70.14	60.66	58.99	91.61
SVM linear	96.45	96.07	91.48	77.02	88.92	40.87	79.00	67.78	62.68	88.29
SVM Gauss	97.37	97.66	94.40	81.57	93.32	45.82	81.09	69.54	69.85	91.30
Test performance for aspect angle (AA): 45° f_{PRF}=122Hz, HOG=[32,32]										
AA: 45° (kNN)	52.65	63.23	59.05	43.00	55.00	20.99	18.54	10.63	31.26	52.91
AA: 45° (lin. SVM)	29.35	64.16	61.23	48.87	58.34	26.06	5.60	10.36	37.66	54.38
AA: 45° (Gauss SVM)	6.40	64.38	62.38	48.51	59.63	28.37	5.87	5.87	36.02	53.18



(a) Performance with f_{PRF}=122Hz, HOG=[16,16]

(b) Performance with f_{PRF}=122Hz, HOG=[32,32]

Fig. 7: Validation and test performance of gross-motor experiments for two different HOG sizes related to Tab. II. Red, yellow, green data labels for testing performance match in color those in Tab. II.

bending from standing, (h) bending from sitting, (i) kneeling down, (j) kneeling up, (k) sitting down, (l) standing up, (m) falling, and (n) standing up from falling. Accounting for the bidirectional orientation, 16 classes were collected and considered for classification. The activities can be seen in Fig. 4 in the yellow-shaded box, together with the relevant bidirectional *RM-PC* plots.

In this paper, we show the average classification results for the gross-motor activities in Tab. II and visualize the results in Fig. 7. In this case, the data set was collected with

only a f_{PRF} of 122Hz, but two HOG cell sizes of [16,16] and [32,32] are compared. These are shown in Tab. II by the gray-shaded and yellow-shaded boxes, respectively. The results show that doubling the HOG cell size does not lead to a drastic classification performance reduction and both results are rather comparable. As a side note, doubling the HOG size reduces the computational load since the resulting input feature vector (κ_i) is shortened by factor $1/2$.

For the training/validation accuracy, it can be seen that the Gaussian SVM classifier gives slightly higher accuracy for the proposed method based on *RM-PC* data domain. When feature fusion is considered, the advantage of using the Gaussian SVM classifier with the HOG cell size of [32,32] is less dominant compared to the NN or the linear SVM classifier. Additionally, for classification based on single data domain, the proposed method of *RM-PC* yields a performance improvement of $+1.71\%$ compared to the best *μD-O* classification (*μD-O* [kNN] 91.61% $\rightarrow$ *RM-PC* [Gauss SVM] 93.32%). Although small in absolute terms, this improvement can be considered significant accounting for the large number of classes (16).

The test performance in Tab. II shows the results for the AA of 45° when using the three different classifiers of kNN, linear SVM, and Gaussian SVM. This is different from Section III-A where the test gesture activities were performed at slower speed and with smaller spatial extent. Specifically, the Gaussian SVM classifier outperforms the other tested classifiers in almost every category, except for the *Fusion all* case, which will be discussed in the next section.

C. Comparative and noise analysis

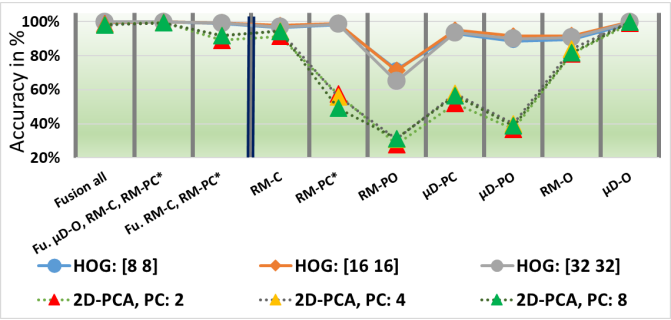
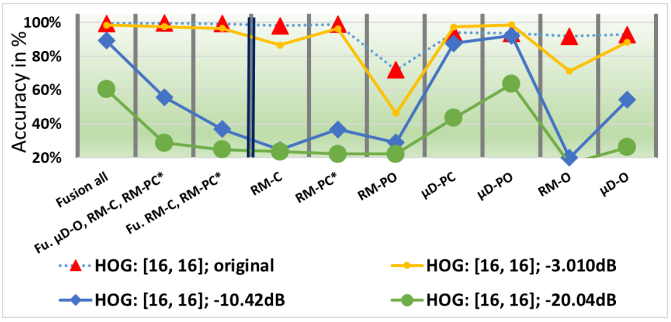
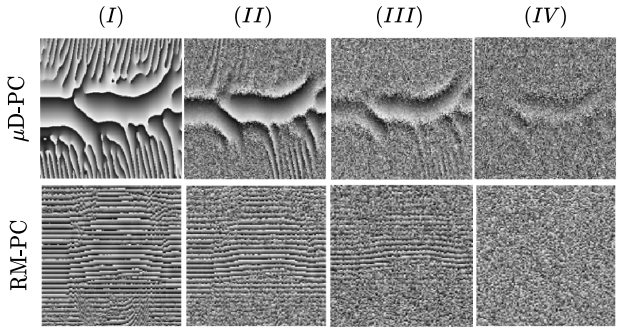


Fig. 8: Comparison between the HOG and the 2D-PCA feature extraction algorithm followed by the Gaussian SVM classifier for the HOG cell sizes 8, 16, and 32, as well as the principal component vectors (PC) of 2, 4, and 8.

In this section, a comparison between the investigated method of the HOG classification and the 2D-PCA classification is demonstrated, as well as a noise performannce analysis on the gesture data set. Regarding Tab. III and Fig. 8, it can be seen that the phase domains, such as RM-PC, RM-PO, μD-PC, μD-PO, as well as their fusion are better classified by the HOG algorithm. The typical μD-spectrogram (μD-O) classification performs almost equal with both methods. This leads to the conclusion that 2D-PCA cannot capture phase-related patterns as well as done by the HOG algorithm.



(a) Performance by adding noise for the HOG algorithm.



(b) Visualization by of the cropped phase-based RM (RM-PC) and μD-spectrogram (μD-RC), respectively, for push and pull arms.

Fig. 9: Validation performance of gesture experiments for different noise levels reductions of (*II*) -3.010dB, (*III*) -10.42dB, and (*IV*) -20.04dB, respectively, for the $f_{\text{PRF}} = 122Hz$. The related results can be found in Tab. III.

Fig. 9 shows the effect of decreasing the SNR by -3.010dB, -10.42dB, and -20.04dB, respectively. The change in the phase pattern can be seen in Fig. 9b for *push and pull arms*. The reduction of the SNR by -3.010dB to the original signal shows almost comparable classification results, specifically for the proposed method (RM-PC). A further SNR reduction leads to an expected classification drop for almost all feature domains, as provided in Tab. III.

D. Discussion on the results

The introduced method of classifying the *RM-PC* showed very promising results, which can be further improved by using feature fusion (early fusion) along with the *RM-C*, or *RM-C* together with *μD-O*. However, we do not suggest to use all available domains, which can be seen for the *Fusion all* case, as the classification accuracy can drastically decrease due to overfitting. Nevertheless, the use of our phase-based classification framework is suitable to avoid the calculation of the STFT completely, and omit any form of μD computation and related feature extraction, and rather focus on the range only. This can be a potentially interesting advantage to cut complexity in view of real-time applications and computationally constrained environments, as demonstrated in Eq. 2 together with Fig. 6.

The *RM-PC* plots in Fig. 4 (yellow-shaded box) for facing the radar can be compared with those recorded for movements

TABLE III: Average performance of gesture experiments by comparing the HOG algorithm (first sub-table) and 2D-PCA (second sub-table) for different cell sizes and by using different principal component (PC) vectors, respectively. The last sub-table shows the performance by adding different noise levels. The 3 left columns are the feature fusion accuracy values and the 7 right columns are the individual domains.

Features / Method	Fusion all	Fu. RM-C, RM-PC, μD-O	Fu. RM-C, RM-PC	RM-C	RM-PC	RM-PO	μD-PC	μD-PO	RM-O	μD-O
Validation performance using HOG, f_{PRF}=122Hz										
HOG: [8 8]	99.12	99.94	98.81	96.55	98.33	71.79	93.04	88.51	89.64	98.99
HOG: [16 16]	99.70	99.94	99.29	97.74	98.99	71.25	95.06	91.55	91.73	99.76
HOG: [32 32]	99.82	99.88	99.17	96.85	98.57	65.06	93.51	90.00	90.60	99.58
Validation performance using 2D-PCA, f_{PRF}=122Hz										
2D-PCA, PC: 2	98.69	99.41	89.05	91.43	57.50	27.92	51.91	36.85	81.07	99.05
2D-PCA, PC: 4	98.27	99.29	91.73	94.64	55.83	31.01	57.74	39.94	83.93	99.82
2D-PCA, PC: 8	98.10	99.29	91.85	94.35	49.11	31.31	56.61	38.81	81.73	99.88
Validation performance using HOG [16 16] and added noise, f_{PRF}=122Hz										
HOG, -3.01dB	98.41	97.55	96.50	86.69	96.28	46.41	97.34	98.58	71.35	88.39
HOG, -10.42dB	89.23	55.89	36.96	24.70	36.85	29.05	87.74	92.14	20.00	54.41
HOG, -20.04dB	60.48	28.81	24.94	23.69	22.26	22.38	43.51	63.63	16.67	26.25

away from the radar, and also with the gesture activities (green-shaded box in the same figure). Here, the changes in contour patterns of the phase plots are dominant and still visible for the human eye, although the differences are less clear and intuitive than those typically visible in spectrograms. However, the HOG algorithm can capture well and distinguish changes in contour patterns in the angular direction and intensity. Fig. 3 is also notable, where the HOG features for the same movement (*push and pull arms*) but performed at different speed and spatial extent are presented for comparison. Specifically, the red-framed HOG features represent the torso movement and are more dominant for a larger spatial extent or faster speed. Also, the green framed HOG features are distinguishable in strength and orientation, which represents mainly the arm movement. Here a difference can easily be seen between small and large extent.

IV. Conclusion

We propose a novel approach for classification of human gross-motor activities and arm gestures based on the phase information directly extracted from high resolution Range Maps (RM). This approach is an alternative compared to the more conventional use of the magnitude of the micro-Doppler (μD) spectrograms for classification. We investigated the wrapped phase of RM and μD spectrograms, whereas the phase-based RM provides superior results over the phase-based μD spectrograms. Nevertheless, the unique shape of those wrapped phases in terms of intensity and complexity of the line patterns requires a suitable feature extraction algorithm to capture the relevant information, differently from features typically used on a conventional μD spectrogram. For this, we exploited the Histogram of Oriented Gradients (HOG) algorithm to capture suitable features towards a phase-based classification by using three commonly known classifiers, namely the Nearest Neighbor (NN), the linear Support Vector Machine (SVM), and the Gaussian SVM.

We demonstrate this approach on two experimental datasets, namely one for gross-motor activities (e.g. sitting, standing, bending, kneeling, etc.), and the second for arm gestures (e.g. pushing and pulling arms, waving hands, or pointing, etc.). The latter dataset is recorded with two different Pulse Repetition Frequencies (f_{PRF}). We have shown that the proposed method can be applied to the arm gesture recognition measured with a 10-times lower f_{PRF} – which can be beneficial by using low cost hardware – without any noticeable decrease of performance while a conventional μD-based approach suffers with such data due to Doppler ambiguities. The method has shown to be robust with respect to the test scenario variables, e.g. the aspect angle to the radar line of sight, the velocity, and the extent of arm movements are also characterized.

Promising validation results for the proposed phase-based RM classification of consistently above 92% are demonstrated for both the arm gestures and gross-motor activities by using HOG features on the phase-based RM. These results based on phase domain classification can even be improved by fusing features from different radar data domains, such as the original RM and/or the μD spectrograms, which shows a more robust performance in different operational conditions

(e.g. different aspect angles, extent, or movement velocity). Superior performance was attained by fusing the proposed phase-based RM together with the magnitude of the RM. In regard to the priorly mentioned RM domain fusion, for the slow f_{PRF} of 12.2Hz of the gesture data set, a remarkable validation accuracy improvement of greater 10% compared to the conventional μD spectrogram classification was achieved. The proposed method may suite for radar systems providing the complex I and Q signal components directly without resorting to the Hilbert transform, as well as, for different operational frequencies and bandwidths.

REFERENCES

[1] S. A. Shah and F. Fioranelli, "RF sensing technologies for assisted daily living in healthcare: A comprehensive review," *IEEE Aerospace and Electronic Systems Magazine*, vol. 34, no. 11, pp. 26–44, Nov 2019.

[2] M. G. Amin and R. G. Guendel, "Radar human motion recognition using motion states and two-way classifications," in *Proceedings of the 2020 IEEE Radar Conference (RadarConf)*, 2020.

[3] G. Li, R. Zhang, M. Ritchie, and H. Griffiths, "Sparsity-based dynamic hand gesture recognition using micro-Doppler signatures," in *2017 IEEE Radar Conference (RadarConf)*, 2017, pp. 0928–0931.

[4] D. Wu, L. Pigou, P. Kindermans, N. D. Le, L. Shao, J. Dambre, and J. Odobez, "Deep dynamic neural networks for multimodal gesture segmentation and recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 8, pp. 1583–1597, 2016.

[5] Z. Zeng, M. Amin, and T. Shan, "Automatic arm motion recognition based on radar micro-Doppler signature envelopes," *IEEE Sensors Journal*, pp. 1–1, 2020.

[6] S. Wang, J. Song, J. Lien, I. Poupyrev, and O. Hilliges, "Interacting with soli: Exploring fine-grained dynamic gesture recognition in the radio-frequency spectrum," in *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, ser. UIST '16. New York, NY, USA: Association for Computing Machinery, 2016, p. 851–860.

[7] H. Li, A. Shrestha, H. Heidari, J. Le Kernec, and F. Fioranelli, "Bi-LSTM network for multimodal continuous human activity recognition and fall detection," *IEEE Sensors Journal*, vol. 20, no. 3, pp. 1191–1201, 2020.

[8] S. Z. Gurbuz and M. G. Amin, "Radar-based human-motion recognition with deep learning: Promising applications for indoor monitoring," *IEEE Signal Processing Magazine*, vol. 36, no. 4, pp. 16–28, 2019.

[9] S. Z. Gurbuz, M. M. Rahman, E. Kurtoglu, T. Macks, and F. Fioranelli, "Cross-frequency training with adversarial learning for radar micro-Doppler signature classification (Rising Researcher)," in *Radar Sensor Technology XXIV*, K. I. Ranney and A. M. Raynal, Eds., vol. 11408, International Society for Optics and Photonics. SPIE, 2020, pp. 58 – 68.

[10] N. Golestani and M. Moghaddam, "Human activity recognition using magnetic induction-based motion signals and deep recurrent neural networks," *Nature communications*, vol. 11, no. 1, pp. 1–11, 2020.

[11] H. Li, A. Shrestha, H. Heidari, J. L. Kernec, and F. Fioranelli, "A multisensory approach for remote health monitoring of older people," *IEEE Journal of Electromagnetics, RF and Microwaves in Medicine and Biology*, vol. 2, no. 2, pp. 102–108, 2018.

[12] X. Liang, H. Li, W. Wang, Y. Liu, R. Ghannam, F. Fioranelli, and H. Heidari, "Fusion of wearable and contactless sensors for intelligent gesture recognition," *Advanced Intelligent Systems*, vol. 1, no. 7, p. 1900088, 2019.

[13] M. G. Amin and R. G. Guendel, "Radar classification of consecutive and contiguous human gross-motor activities," *IET Radar, Sonar & Navigation*, p. 12, 2020.

[14] Y. He, P. Molchanov, T. Sakamoto, P. Aubry, F. Le 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.

[15] M. G. Amin, A. Ravisankar, and R. G. Guendel, "RF sensing for continuous monitoring of human activities for home consumer applications," in *Big Data: Learning, Analytics, and Applications*, Baltimore, Maryland, May 2019.

[16] H. Li, X. Liang, A. Shrestha, Y. Liu, H. Heidari, J. Le Kernec, and F. Fioranelli, "Hierarchical sensor fusion for micro-gestures recognition with pressure sensor array and radar," *IEEE Journal of Electromagnetics, RF and Microwaves in Medicine and Biology*, pp. 1–1, 2019.

[17] J. Kwon and N. Kwak, "Human detection by neural networks using a low-cost short-range Doppler radar sensor," in *2017 IEEE Radar Conference (RadarConf)*, 2017, pp. 0755–0760.

[18] S. A. Rane, A. Gaurav, S. Sarkar, J. C. Clement, and H. K. Sardana, "Clutter suppression techniques to detect behind the wall static human using UWB radar," in *2016 IEEE International Conference on Recent Trends in Electronics, Information Communication Technology (RTE-ICT)*, 2016, pp. 1325–1329.

[19] W. A. M. Al-Ashwal, "Measurement and modelling of bistatic sea clutter," Ph.D. dissertation, UCL (University College London), 2011.

[20] W. Fischer, *Basic Principles of Digital Modulation*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 209–236.

[21] B. R. Mahafza, *Radar Signal Analysis and Processing Using MATLAB*, 1st ed. Chapman and Hall CRC, 2008.

[22] P. H. Quanjer, A. Capderou, M. M. Mazicioglu, A. N. Aggarwal, S. D. Banik, S. Popovic, F. A. Tayie, M. Golshan, M. S. Ip, and M. Zelter, "All-age relationship between arm span and height in different ethnic groups," *European Respiratory Journal*, vol. 44, no. 4, pp. 905–912, 2014.

[23] R. G. Guendel, F. Fioranelli, and A. Yarovoy, "Derivative target line (dtl) for continuous human activity detection and recognition," in *Proceedings of the 2020 IEEE Radar Conference (RadarConf), Accepted*, Sept 2020.

[24] M. Mercuri, I. R. Lorato, Y. H. Liu, F. Wieringa, C. V. Hoof, and T. Torfs, "Vital-sign monitoring and spatial tracking of multiple people using a contactless radar-based sensor," *Nature Electronics*, vol. 2, no. 6, pp. 252–262, 2019. [Online]. Available: http://dx.doi.org/10.1038/s41928-019-0258-6

[25] V. Chen and H. Ling, *Time-frequency Transforms for Radar Imaging and Signal Analysis*, ser. Artech House Mobile Communicat. Artech House, 2002. [Online]. Available: https://books.google.nl/books?id=h-8knwEACAAJ

[26] W. Troy, M. Thompson, and Y. Li, "Isar imaging of rotating blades with an UWB radar," in *2015 Texas Symposium on Wireless and Microwave Circuits and Systems (WMCS)*, 2015, pp. 1–4.

[27] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 1, 2005, pp. 886–893 vol. 1.

[28] Y. Pan, X. Hou, and C. Liu, "A robust system to detect and localize texts in natural scene images," in *2008 The Eighth IAPR International Workshop on Document Analysis Systems*, 2008, pp. 35–42.

[29] L. Weng, "Object detection for dummies part 1: Gradient vector, HOG, and SS," *lilianweng.github.io/lil-log*, 2017, accessed July 20, 2020. [Online]. Available: http://lilianweng.github.io/lil-log/2017/10/29/object-recognition-for-dummies-part-1.html

[30] S. J. D. Prince, *Computer Vision: Models, Learning, and Inference*, 1st ed. USA: Cambridge University Press, 2012.

[31] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*, 3rd ed. USA: Prentice Hall Press, 2009.

[32] M. Alam, I. Islam, and M. R. Amin, "Performance comparison of STFT, WT, LMS and RLS adaptive algorithms in denoising of speech signal," *International Journal of Engineering and Technology*, vol. 3, no. 4, Jun 2011.

Ronny Gerhard Guendel received his Diploma-Engineering degree (Dipl.-Ing. (FH)) from the University of Applied Sciences Zwickau in Germany. He worked at the Fraunhofer Institute for Machine Tools and Forming Technology (IWU). In 2017/18 he joined Villanova University (USA) as a Fulbright scholar in Electrical Engineering. Over summer 2018, he worked on vehicular wireless communications at the Vodafone Chair Dresden guided by Prof. Gerhard Fettweis. In 18/19, the Center for Advanced Communications of Villanova University offered him a research assistantship under Dr. Moeness G. Amin, working on continuous activity classification. He graduated with a Master of Science degree in Signal Processing and Communications in 2019. Since January 2020 he is pursuing his Ph.D. at Delft University of Technology. He works on monitoring of continuous human activities in range and time beyond micro-Doppler by using RF-sensing technology.

Francesco Fioranelli (M'15–SM'19) received his Laurea (BEng, cum laude) and Laurea Specialistica (MEng, cum laude) degrees in telecommunication engineering from the Università Politecnica delle Marche, Ancona, Italy, in 2007 and 2010, respectively, and the Ph.D. degree from Durham University, U.K., in 2014. He is currently Assistant Professor at TU Delft in the Netherlands, and was an Assistant Professor at the University of Glasgow (2016-2019) and Research Associate at University College London (2014-2016).

His research interests include the development of radar systems and automatic classification for human signatures analysis in healthcare and security, drones and UAVs detection and classification, automotive radar, wind farm and sea clutter. He has authored over 85 publications between book chapters, journal and conference papers, edited the book on "Micro-Doppler Radar and Its Applications" published by IET-Scitech in 2020, and received three best paper awards.

Alexander Yarovoy (F'15) graduated from the Kharkov State University, Ukraine, in 1984 with the Diploma with honor in radiophysics and electronics. He received the Candidate Phys. & Math. Sci. and Doctor Phys. & Math. Sci. degrees in radiophysics in 1987 and 1994, respectively.

In 1987 he joined the Department of Radiophysics at the Kharkov State University as a Researcher and became a Full Professor there in 1997. From September 1994 through 1996 he was with Technical University of Ilmenau, Germany as a Visiting Researcher. Since 1999 he is with the Delft University of Technology, the Netherlands. Since 2009 he leads there a chair of Microwave Sensing, Systems and Signals. His main research interests are in high-resolution radar, microwave imaging and applied electromagnetics (in particular, UWB antennas). He has authored and co-authored more than 450 scientific or technical papers, six patents and fourteen book chapters. He is the recipient of the European Microwave Week Radar Award for the paper that best advances the state-of-the-art in radar technology in 2001 (together with L.P. Ligthart and P. van Genderen) and in 2012 (together with T. Savelyev). In 2010 together with D. Caratelli Prof. Yarovoy got the best paper award of the Applied Computational Electromagnetic Society (ACES).

Prof. Yarovoy served as the Chair and TPC chair of the 5th European Radar Conference (EuRA'08), Amsterdam, the Netherlands, as well as the Secretary of the 1st European Radar Conference (EuRAD'04), Amsterdam, the Netherlands. He served also as the co-chair and TPC chair of the Xth International Conference on GPR (GPR2004) in Delft, the Netherlands. He served as an Associated Editor of the International Journal of Microwave and Wireless Technologies from 2011 till 2018 and as a Guest Editor of five special issues of the IEEE Transactions and other journals. In the period 2008-2017 Prof. Yarovoy served as Director of the European Microwave Association (EuMA).