Golipoor, Sahar; Sigg, Stephan

Environment and Person-independent Gesture Recognition with Non-static RFID Tags Leveraging Adaptive Signal Segmentation

Published in:
2024 IEEE 29th International Conference on Emerging Technologies and Factory Automation, ETFA 2024

DOI:
10.1109/ETFA61755.2024.10710733

Published: 01/01/2024

Document Version
Peer-reviewed accepted author manuscript, also known as Final accepted manuscript or Post-print

Environment and Person-independent Gesture Recognition with Non-static RFID Tags Leveraging Adaptive Signal Segmentation

Sahar Golipoor, Stephan Sigg
Department of Information and Communications Engineering, Aalto University, Finland
e-mails: {sahar.golipoor, stephan.sigg}@aalto.fi

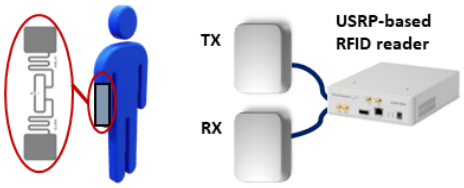
***Abstract*—Gesture recognition for human machine interaction enhances the efficiency, safety, and usability of industrial and factory automation systems. We investigate hand-gesture recognition using battery-less body-worn reflective tags. Particularly, we propose two methods for hand gesture recognition using radio frequency identification (RFID). From backscattered signals we utilize in-phase and quadrature (IQ) constellation, as well as the phase. We convert extracted IQ samples into images and interprete them for gestures using a pre-trained VGG16. As a second approach we alternatively conduct pre-processing on the phase of the backscattered signals and propose Zero Crossing-Modified Derivative (ZCMD) for signal segmentation. Through signal resampling and wavelet denoising we mitigate undesired fluctuations introduced during this process, while retaining crucial signal characteristics. Subsequently, we integrate time-domain and frequency-domain features of the signals and train a random forest classifier based on these features to identify different gestures. Utilizing battery-free body-worn RFID tags, we are able to outperform a state-of-the art method and recognize four gestures with an accuracy of 81% with the VGG16-based model. Employing phase, we achieve an accuracy of 94%.**

***Index Terms*—RFID, gesture recognition, human-sensing, signal processing, signal segmentation**

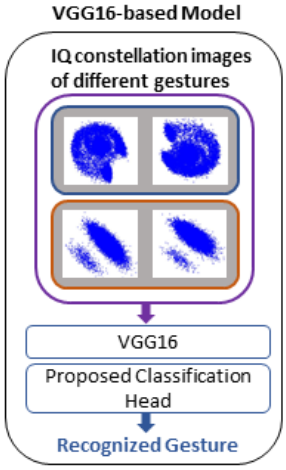
I. Introduction

Radio-based human sensing describes the analysis of electromagnetic signals reflected back from individuals or objects, for localization [1], gesture [2], vital sign [3] or emotion recognition [4]–[6]. Radio sensing has a high potential in automation due to the nearly ubiquitous availability of RF interfaces in industrial IoT and other devices [7]. The high energy consumption of these approaches is a challenge though [8], [9]. Further challenges in RF-based human sensing are, for instance, that clothes and the human body absorb a remarkable portion of the signal and only a smaller part is reflected and can therefore be utilized for RF-based human sensing. Furthermore, existing work typically defines a region of interest for RF-based human sensing and the recognition performance degrades for subjects outside of the region of interest or when the subject is rotated with respect to the receiving antenna. Finally, RF-based sensing is often perceived to have privacy issues since subjects who do not intend to take part in the sensing have indeed no way to opt out.

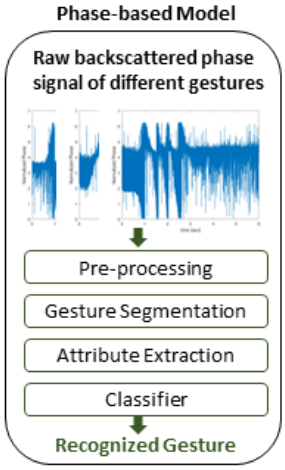
We address these challenges by requiring that the subject from which we sense gestures, pose, location and activity, is instrumented with reflective material (in our case, backscattering Radio frequency identification (RFID) tags). For instance, RFID tags can be integrated into professional work clothes or into visitor or name-tag badges at industrial facilities but they are otherwise not normally found in clothing in general. RFID tags are conventionally known for identification applications, but also their potential for human sensing has been demonstrated [10]. In contrast to many other RF-sensing technologies, RFID and backscattering tags feature a low cost (e.g. price, energy, complexity). Attached to the surface of a human body, RFID and backscatter technology may a) amplify the energy reflected back from a human target, b) move together with the human subject, so that the sensing field follows the moving subject, and c) when employed as body-worn sensing technology, they allow an opt-in human sensing system, where the subject decides whether or not to



(a) Image and phase corresponding to 4 different gestures are built using UHF RFID tag and a USRP equipped with two circularly polarized antennas.



(b) Image data fed to a VGG16 pre-trained model; the classification head contains convolutional and dense layers.



(c) Phase signals are pre-processed and segmented and attributes are extracted and fed to the classifier.

Fig. 1: System Models and Methodology. We propose and compare image and phase-based gesture recognition models.

wear clothes that enable smart interaction with an environment or service.

Signals reflected from fabric which is not equipped with RFID tags have a significantly lower energy and are therefore more difficult to detect compared to signals reflected from RFID. Furthermore, we have demonstrated in [11] that it is possible to distinguish different body parts by exploiting variable phase profiles for body-worn RFID tag groups. Such marker-based setup allows, for instance, high recognition rates for complex movements of individual body parts. Note that RFID may also distribute a unique Electronic Product Code (EPC), which may give human subjects the possibility to disclose an identity, e.g. for smart environment-situated services.

It should be pointed out that proper learning approaches and pre-processing can significantly enhance the performance of recognition systems and their robustness. In this regard, we study transfer learning for gesture recognition. The conversion of channel state information (CSI) [12] and IQ samples [13] to images has shown promising results thanks to the existence of powerful pre-trained image classifier models provided for vision based tasks, such as VGG16 [14], and ResNet-50 [15]. On top of that, since good representation of a signal profile corresponding to different gestures is vital, we apply signal processing steps proposed in ReActor [16], i.e. , an RFID-based gesture recognition system that used Varri signal segmentation, and propose our own signal segmentation schemes to design a highly accurate and robust gesture recognition system. In summary, this paper has the following contributions.

- We study human gesture recognition extracting IQ samples according to two approaches for four different gestures. Specifically, we convert IQ samples to images before we feed them to a VGG16 pre-trained model.
- We extract and pre-process phase of backscattered signals and present two heuristic signal segmentation methods that isolate gesture signals from noise. The approaches are effective in new environments with new participants, and simultaneously cope with speed variation in gestures.
- We resample signals and suppress unwanted fluctuation introduced during the resampling with wavelet denoising while preserving important features of signals. Finally, we merge time-domain and frequency-domain features of smooth signals and build a classifier based on the these features to recognize 4 key gestures.

We demonstrate in case studies with 14 subjects and in three environments that our approaches outperform the state-of-the-art.

II. Related work

RFID-aided human activity recognition is a well researched domain [17]. The studies mainly were conducted in scenarios in which tag arrays were embedded into environments. Using ResNet, fall detection and daily activity recognition was examined in [18]. Fusing phase and received signal strength (RSS) of the RFID backscattered signal, a spatiotemporal graph convolutional neural network based-model was proposed to recognize human actions [19]. For fine-grained RF-based gesture recognition, such as the detection of hand and finger movements in front of an RFID array, different classifiers were evaluated [20]. The authors in [16] proposed ReActor and ReActor+ for real-time hand gesture recognition, handling the impact of varying speeds. Multi-touch fine-grained gestures were studied in [21] leveraging K-Nearest Neighbors to track finger trace and Convolutional Neural Networks (CNN) to recognize gestures. Sign language recognition was considered using a CNN [22], and adversarial learning [23]. Authors in [24] proposed a permutation-based dataset generation and a network with designed loss function and distance metric to distinguish different gestures. Contrary to the aforementioned works, we consider body-attached RFID tags.

RFID tags are light and can be attached or woven on or into outfits. The authors in [25] embedded RFID tags into gloves and studied hand gesture recognition by interpreting movement patterns with the help of dynamic time warping. A position-independent sign language recognition was examined in [26] by normalizing the hand's horizontal rotation angle and radial distance.

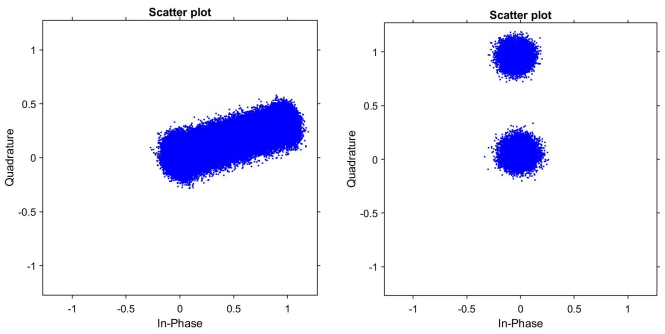
Using RFID tags on the human body for coarse-grained gesture recognition has challenges. First, the tag's orientation changes constantly because of large movement, making it difficult for RFID readers to consistently read the tag. Therefore, the quality of reflected signals is poor due to polarization mismatch [27]. Second, the reflection of signals is affected by the human body's anatomy and tissues [28]. Different body compositions and shapes can cause variations in how RF signals are reflected, potentially leading to various patterns for the same gestures and makes recognition difficult. Authors in [29] studied relatively coarse-grained gesture recognition by integrating multi-modal CNN and Long Short-Term Memory networks to represent features in RFID signals. We use different methodologies in contrast to this work, proposing two models, image-based gesture recognition leveraging transfer learning and phase-based gesture recognition using a random forest classifier. We also consider a different set of gestures.

III. System model and Problem statement

We implement the USRP-based UHF RFID reader developed in [30]. The reader software utilizes GNU Radio and consists of 6 blocks (USRP source, matched filter, gate, tag decoder, Generation-2 UHF RFID logic and USRP sink). For the study, we attach a UHF RFID tag on the hand of a human subject. Specifically, the IQ samples reflected via the RFID tag and also their extracted phase are then interpreted for different gestures.

For the IQ sample extraction, we investigate two approaches as detailed below (cf. Fig. 3).

DC-Matched filtering-based: We extract IQ samples after DC offset removal and matched filtering. The reader software is developed such that the DC offset component is estimated and removed from each sample. The impulse response of the matched filter is a square pulse with half a symbol period. Fig. 2a depicts IQ samples captured by this approach.



(a) Received samples after matched filtering and DC offset removal (accumulated)

(b) Received samples after synchronization; two separated states are visible.

Fig. 2: IQ constellation while the RFID tag is fixed

Synchronization-based: We extract IQ samples after synchronization. The tag decoder estimates the symbol rate by maximizing the energy of the matched filter output. The frame synchronization time is needed for the above process [30, eq. (14)]. It is estimated by maximizing the correlation between the received signal with known preamble [30, eq. (7)]. As depicted in Fig. 2b, IQ samples for this approach are separated into two states (RFID absorb state, RFID reflect state).

IV. Set-up and Data Collection

The RFID system operates in the 910 MHz range. The distance between transmit and receive antennas is 70 cm to suppress leakage (cf. Fig. 4). We use the Zebra Z-Select 2000T tag attached to the hands of human subjects. The IQ constellation diagrams of the tag's EPC responses for four different gestures, Lift (L), Lateral Raise (LR), Pull (PL), and Push (PS) are extracted (cf. Fig. 5). Fig. 1b depicts the scatter plot for two gestures of two pairs of IQ samples. The IQ sample plots for a particular gesture are similar. We deploy these IQ samples and create 3 datasets for the recognition of different gestures as follows.

- Dataset I: Converting IQ constellations after DC-Matched filtering to images
- Dataset II: Converting IQ constellations after Synchronization to images
- Dataset III: Phase extraction from IQ constellations after Synchronization

We utilized an USRP N200 and two circularly polarized antennas. Collecting data from 14 people (7 male and 7 female) with different height (155 cm-185 cm) and physique, we extracted IQ sample images and phases for every gesture and every person. The total number of data points for every gesture is 140. First, data was collected from 9 participants in one environment and used to create two models, VGG16-based and phase-based, to recognize gestures. Next, in two new environments, we collected data from 5 new participants; i.e., unknown to the trained model. The two presented models are evaluated on the unseen data set. The gestures were performed in a distance of 1.7 m from the reader.

TABLE I: Data Augmentation Parameters

Parameter	Value / state
rotation range	90
width shift range	0.1
height shift range	0.1
shear range	0.1
zoom range	0.1
rescaling factor	$1.0/255.0$
horizontal flip	True

V. VGG16-based gesture recognition model

In the following, we detail how we interpret IQ images for gesture recognition.

A. VGG16 Network

We fed our datapoints, IQ images, to a VGG16 pre-trained model which is particularly well suited for smaller datasets because it has learned rich and generalized features from its extensive training [31]. The VGG16 architecture consists of 16 layers, including 13 convolutional layers and 3 fully connected layers. The earlier layers of VGG16 learn low-level features such as edges, whereas the deeper layers learn high-level features such as shapes [32]. Applying a fine-tuning approach, we removed the fully connected layers of the pre-trained model and re-train the last three VGG16 convolutional layers. As illustrated in Fig. 1b, we then added a new classification head consisting of two convolutional layers each, followed by MaxPooling layers, a flatten layer, a fully connected layer with 256 units applying the rectified linear unit (ReLU) activation function to allow the model to learn complex patterns and relationships in the data, a dropout layer, and finally a fully connected layer with 4 units, which matches the number of classes in our classification task. The activation function used for the final layer is softmax, which computes the probability distribution over the classes and outputs the probability of the input belonging to each class.

B. Data Augmentation

Since our training data is limited, we created additional synthetic data samples (data augmentation) using the parameters specified in TABLE I.

C. Results and Discussion

As illustrated in Fig. 6b, the gestures LR and PS were recognized successfully with 100 % recognition accuracy. This is due to the relative uniqueness of their scatter plots. Several scatter plots extracted by the synchronization-based Approach corresponding to the 4 gestures are shown in Fig. 7. As it can be seen, PL and L gestures generate similar images. This is the reason why they are more frequently misclassified. As it can be seen from the second row of Fig. 7, PL and L are similar to the PS image in the first row. LR, however, since it has its own unique image, is easier to be distinguished by the recognition model.

Regarding the DC-Matched filtering-based Approach, LR has the highest accuracy due to its dissimilarity with other

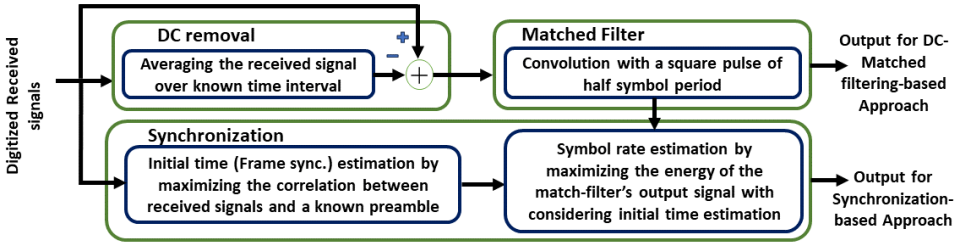


Fig. 3: Processing chain for DC-Matched filtering-based Approach and Synchronization-based Approach

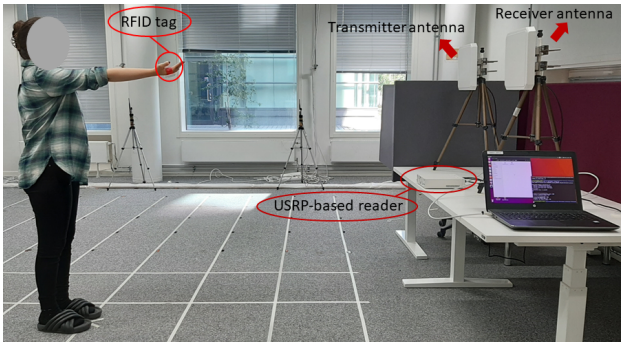
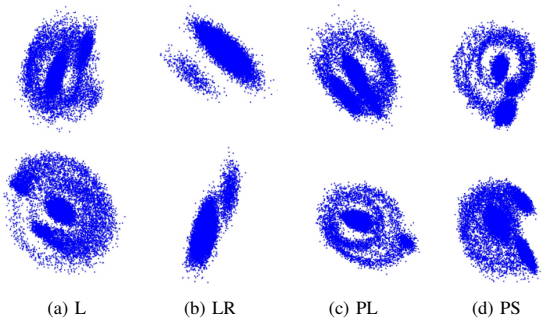
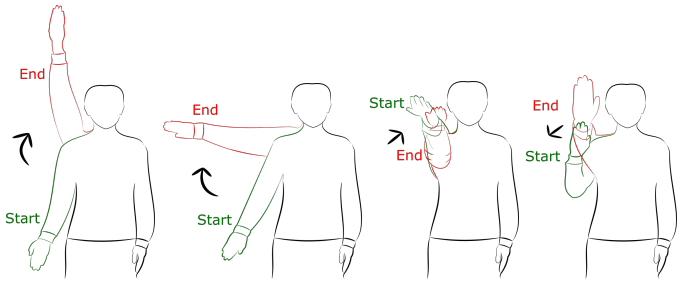


Fig. 4: Experimental Setup. RFID attached on the back of the hand moving in the distance of 1.7 m away from reader and antennas.



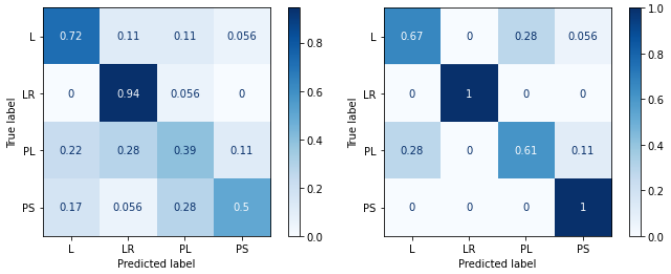
(a) L (b) LR (c) PL (d) PS

Fig. 7: Scatter plot samples of different gestures for the synchronization-based Approach



(a) L (b) LR (c) PL (d) PS

Fig. 5: The considered gestures.



(a) DC-Matched filtering-based Approach with test accuracy=63% (b) Synchronization-based Approach with test accuracy=81%

Fig. 6: Confusion matrices while the dataset was split 60%, 20%, and 20% for training, validation, and test sets, respectively.

gestures. On the other hand, other gestures are highly similar. This is because the scatter plots extracted by DC-Matched filtering have a shape as illustrated in Fig. 2a while the tag is fixed. This causes the images to not be distinct for different gestures. Thus, gesture recognition becomes then more challenging when using DC-Matched filtering.

VI. Phase-based Gesture Recognition Model

An extracted phase signal is noisy and erratic and should be pre-processed before feeding it to gesture recognition algorithms. We perform signal smoothing and noise reduction as well as signal normalization. It should be mentioned that we set a constant inventory round for reader and RFID communication and participants perform the gestures in various speed. Gestures might be fulfilled before the inventory round is completed. Thus, we apply signal segmentation to capture the gesture part of the signal. The length of segmented signals is diverse and we resample them to a certain length and perform wavelet denoising to achieve smooth signals while preserving important features of signals. Finally, we extract time-domain and frequency-domain features of the resulted signal and recognize gestures applying a random forest classifier.

A. Signal Filtering and Smoothing

Since high-frequency noise exists in the phase of the signal, we exploit Savitzky-Golay (S-G) [33] and moving average filtering, which preserve the integrity of the underlying signal.

B. Signal Normalization

We normalize the phase of the received signal to enhance gesture-relevant changes and alleviate the impact of back-

ground signals by mapping them to a range of $[-1, 1]$. The normalization of the m-th reading phase φ_m is as follows

$$\tilde{\varphi}_m = \begin{cases} \frac{\varphi_m - \bar{\varphi}}{\varphi_{\max}} & \varphi_m \geq 0 \\ -\frac{\varphi_m - \bar{\varphi}}{\varphi_{\min}} & \varphi_m < 0 \end{cases}, \tag{1}$$

where $\bar{\varphi}$, $\varphi_{\max}$, and $\varphi_{\min}$ denote the mean, maximum, and minimum of all phase readings of the tag (i.e., $\boldsymbol{\varphi} = [\varphi_1, \varphi_2, \ldots, \varphi_M]$), respectively. The total number of phase readings is M.

C. Gesture Segmentation

Participants perform gestures with different speed and the actual start and end of a gesture within the signal is initially unknown and must be segmented adaptatively. This step is crucial for proper training of the recognition algorithm. Two methods are considered as explained.

1) Gesture segmentation based on amplitude and frequency measurement: We apply the Varri method [34], [35] as it was used in ReActor [16] to extract the active part of the signal containing the gesture information. This method calculates the amplitude measurement $\mathcal{A}$ and frequency measurement $\mathcal{F}$ of the signal over sliding windows of length L. Accordingly, these measurements within the n-th window are computed as:

$$\mathcal{A}_n = \sum_{\ell=0}^{L} |\tilde{\varphi}_{n,\ell}| \tag{2}$$

$$\mathcal{F}_n = \sum_{\ell=0}^{L} |\tilde{\varphi}_{n,\ell} - \tilde{\varphi}_{n,\ell-1}| \tag{3}$$

where $\tilde{\varphi}_{n,\ell}$ indicates the ℓ-th data point (i.e., normalized phase) in the n-th sliding window. Next, the measurement difference between two successive sliding windows is calculated as

$$\mathcal{G}(n) = \mathcal{C}_{\mathcal{A}} |\mathcal{A}_{n+1} - \mathcal{A}_n| + \mathcal{C}_{\mathcal{F}} |\mathcal{F}_{n+1} - \mathcal{F}_n|, \tag{4}$$

where $\mathcal{C}_{\mathcal{A}}$ and $\mathcal{C}_{\mathcal{F}}$ are application specific coefficients. We set $\mathcal{C}_{\mathcal{A}} = 7$ and $\mathcal{C}_{\mathcal{F}} = 1$. Finally, the local maxima of $\mathcal{G}$ is calculated, which determines the boundary of the gesture part of the signal.

2) Gesture segmentation based on zero crossing and modified derivative: In this part, we introduce gesture segmentation schemes that are able to cope with gesture speed variation and also are environment independent. Based on our observations, we have two signal categories and accordingly, we present two separate segmentation schemes. For L, PL, and PS gestures (cf. Fig. 8a), zero crossing (ZC) happens along the actual gesture part of the signal (light green area) within large amplitude changes. However, for many gestures, we observe that the non-gesture part of signals (between red lines) also experience ZCs but with small fluctuations. To capture ZC and large changes, we calculate a metric over sliding windows, which is defined as

$$\varpi_n \triangleq \zeta_n(\Phi_n^{\max} - \Phi_n^{\min}) \quad n = 1, \ldots, N_g, \tag{5}$$

where ζ_n and N_g denote the number of ZC in the nth sliding window and the number of sliding windows, respectively. The maximum and minimum values of the data point in the nth sliding window are $\Phi_n^{\max}$ and $\Phi_n^{\min}$, respectively. Next, we compare the calculated metric vector $\boldsymbol{\varpi} = [\varpi_1, \varpi_2, \ldots, \varpi_{N_g}]$ with the predefined threshold. The data points among the first and the last sliding windows whose values ϖ exceed the threshold are considered as gesture part. To precisely capture the data point corresponding to the gesture part, the overlap between two successive sliding windows is set to one sample.

LR gestures, however, do not follow the above ZC scheme as illustrated and explained in Fig. 8b. Thus we monitor the signal length after ZC segmentation. Based on our observations, the length of LR gestures are either too short or too long with considerable differences compared to other gestures. Accordingly, separate segmentation is employed based on the length. If the length of the signal after ZC is not in the expected range, the algorithm will continue with the normalized signal. As can be seen from Fig. 8b, there is a sudden and significant change in the phase (light green area). To segment this part of the signal, a sliding window is applied and the modified derivative (MD) measurement of the n-th window is calculated as

$$\mathcal{D}_n = \frac{\Phi_n^{\max} - \Phi_n^{\min}}{I_n^{\max} - I_n^{\min}} \quad n = 1, \ldots, N_g \tag{6}$$

where $\Phi_n^{\max}$, and $\Phi_n^{\min}$ indicate maximum and minimum value within the n-th sliding window, respectively. $I_n^{\max}$ and $I_n^{\min}$ represent the corresponding index of maximum and minimum value, respectively. Next, we find the index of the maximum value of the MD, i,e., $i_{\max} = \text{argmax}(\boldsymbol{\mathcal{D}})$ where $\boldsymbol{\mathcal{D}} \triangleq [\mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_{N_g}]$. Accordingly, the gesture part of the signal can be

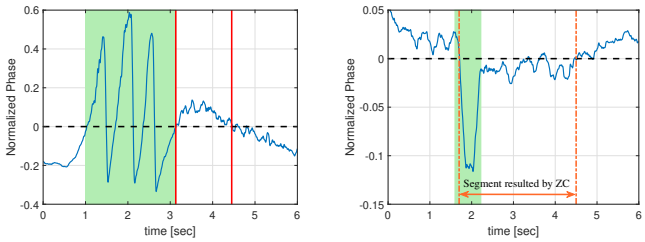
$$\boldsymbol{S}_g = \begin{cases} \tilde{\boldsymbol{\varphi}}\left[i_{\max} - \lfloor 0.8|i_{\max} - 1| \rfloor : i_{\max} + \text{Th}\right] & a \\ \tilde{\boldsymbol{\varphi}}\left[i_{\max} - \text{Th} : i_{\max} + \lfloor 0.8|i_{\max} - N| \rfloor\right] & b \\ \tilde{\boldsymbol{\varphi}}\left[i_{\max} - \text{Th} : i_{\max} + \text{Th}\right] & \text{otherwise,} \end{cases} \tag{7}$$

where $a \triangleq |i_{\max} - 1| < \text{Th}$ and $b \triangleq |i_{\max} - N| < \text{Th}$ are conditions where the large MD value occurs near to the beginning or end of the signal, respectively. $\lfloor . \rfloor$ is the floor function and $\tilde{\boldsymbol{\varphi}}[\, a_0 : b_0]$ captures the component of vector $\tilde{\boldsymbol{\varphi}}$ from index a_0 to index b_0. We set the value of Th based on the data.

D. Resampling and Wavelet Smoothing

Segmented signals vary in length (e.g. duration of gesture, speed, etc.). Consequently, the attribute feature vector for each segmented signal might have different lengths since we extract the wavelet coefficients as part of the feature vector[1]. Hence, we apply resampling to harmonize the length of all segmented signals. Resampling employs an anti-aliasing filter [36], which is sensitive to large transients of a signal. Therefore, we observed some resonance effects on the resampled signals. To alleviate it, we apply wavelet de-noising on the resampled signals to receive a smooth signal.

[1]The number of Wavelet coefficients depends on the signal length.



(a) An exemplary PS signal instance (b) An exemplary LR signal instance

Fig. 8: Normalized signals before segmentation: In (a) that is an exemplary push signal, both gesture part (light green area) and non-gesture part (between red lines) have zero crossings. Significant changes in the gesture part (light green area) are captured by Eq (5). In (b) that is an exemplary lateral raise signal, zero crossing happens between orange dashed lines. Eq (5) that is the combination of zero crossings and capturing significant changes can not isolate the desired segment (shown in light green). To capture the pattern, the modified derivative (Eq (6)) is instrumental.

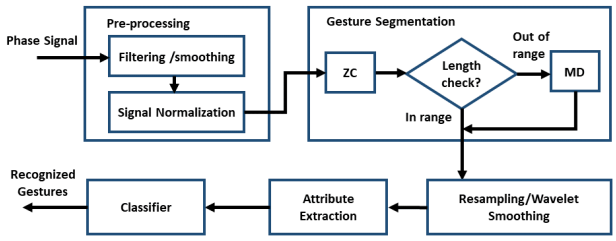


Fig. 9: The workflow of phase-based gesture recognition that contains zero crossing and modified derivative segmentation schemes

E. Attribute Extraction

We extract 13 statistical features, namely the mode, the median, the first quartile, the third quartile, the mean, the max, the min, the range, the variance, the standard deviation, the third-order central moment, the kurtosis and the skewness. Moreover, wavelets provide a multi-resolution analysis of a signal, meaning that they can capture information at various levels of detail. This is useful for identifying both coarse and fine-grained patterns within the data. In this regard, we use single level discrete wavelet transform with Daubechies wavelet and extract both low-frequency and high-frequency coefficients. Wavelet-based features are often more robust to variations such as translation, scaling, and rotation. This makes them particularly useful in recognition tasks where such variations are common. Statistical features and wavelet coefficients are then concatenated and used to train a random forest classifier. The workflow of phase-based gesture recognition is shown in Fig. 9.

F. Results and Discussion

The normalized confusion matrices of classifying different gestures are presented in Fig. 10. We use data from 9 participants to train the model. For the phase-based model, the dataset is divided into 90% and 10% for training and testing, respectively. Fig. 10a shows the confusion matrix on test data points applying the ZCMD method. The overall test accuracy achieved is 94.44%. Note that, applying ReActor [16], including Varri signal segmentation, achieves only

TABLE II: Accuracy (Acc.), Precision (Pre.), and Recall (Rec.) of Methods

Test data categories*	Category 1			Category 2		
Metric / Method	Acc.	Pre.	Rec.	Acc.	Pre.	Rec.
ZCMD	94.44	95	94	87	86.9	87
ReActor [16]	77.77	80.2	79.2	62.5	58.5	62.5
Synchronization-based	81	81.1	81.9	56.66	59.1	56.5

* Category 1 is for trained model on test data pints. Category 2 is for trained model on the data points from unseen participants in the new environments.

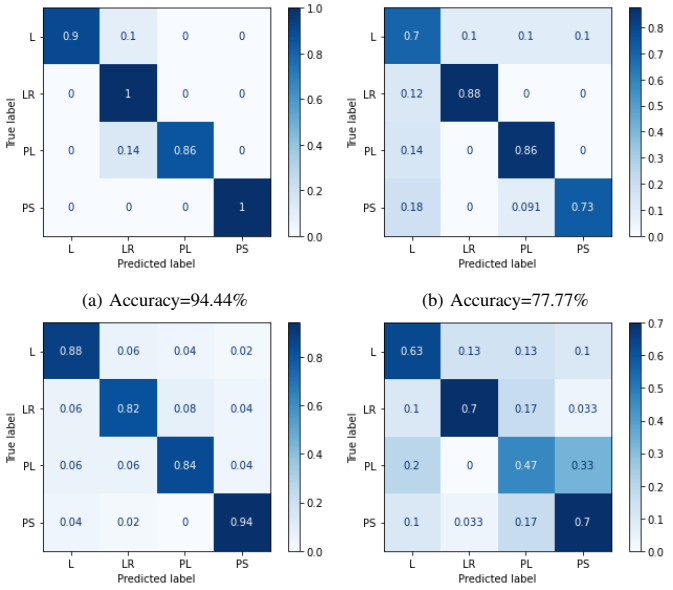
77.77% accuracy (cf. Fig. 10b). This is because Varri includes parameters $\mathcal{C}_{\mathcal{A}}$ and $\mathcal{C}_{\mathcal{F}}$ that are basically application-dependent coefficients. Moreover, it is well suited in biomedical signals such as electroencephalogram where the multi-path effects have already been mitigated. In ReActor [16], where Varri was applied, the tags were fixed and gestures were performed very close to tags and antenna, so that the signals were less effected by multi-path than in our study. LR and PS are classified with 100% accuracy in ZCMD.

The trained models based on these two segmentation methods are then evaluated in terms of robustness. We assess the models on collected data from 5 new participants (unseen participants via model) in new environments. The corresponding confusion matrices are demonstrated in Fig. 10c and Fig. 10d. ZCMD performs effective in the new condition with an overall accuracy of 87% and the recognition accuracy of all gestures is more than 80%. Even though LR test data points are classified with 100%, the accuracy of this gesture drops by almost 20% on the new test data. This is because signal propagation and noise in new environments impact on the thresholding of MD segmentation.

An overall comparison for the 3 methods is shown in TABLE. II in terms of accuracy, precision, and recall for two test data categories. Although the VGG16-based model can reach 81% test accuracy, it is highly dependent on environments and participants. Its accuracy drops to 56.66%. While the test accuracy of the ReActor-based model [16], i.e. 77.77%, is lower than that of the VGG16-based model, it performs better in new conditions and reaches 62.5%. The ZCMD-based model, however, outperforms both models with an overall test accuracy of 94.44% and also remains functional and effective in new conditions and its accuracy is 87%.

VII. Discussion and Limitations

There are certain constraints when it comes to non-static RFID-based gesture recognition. The tag's orientation shifts continuously due to movement, complicating the RFID readers' ability to consistently read the tag. Consequently, the quality of reflected signals suffers due to polarization mismatch. Additionally, variations in body composition and shape can alter RF signal reflections, potentially creating different patterns for the same gestures and making recognition challenging. Data augmentation can help to address this issue. Besides, since angular measurements are not provided in a single receiver antenna scenario, gestures might not seem



(a) Accuracy=94.44%

(b) Accuracy=77.77%

(c) Accuracy=87%

(d) Accuracy=62.5%

Fig. 10: Confusion matrices: (a) Confusion matrix with ZCMD segmentation on test data points; (b) Confusion matrix with Varri segmentation applied in ReActor [16] on test data points (c) Confusion matrix with ZCMD segmentation on test data points, i.e. , collected from new participants in new environments (d) Confusion matrix with Varri segmentation applied in ReActor [16] on test data points, i.e. , collected from new participants in new environments.

distinct easily. Sweep (hand motion from right to hand) and lift, for example, might generate similar backscatterd signals from the receiver's perspective. In this regard, exploring new features through feature engineering may help.

VIII. Conclusion

We have investigated the use of body-mounted RFID tags for human gesture recognition. Using a USRP-based reader and a single UHF RFID tag, we have successfully distinguished four gestures through two distinct methodologies. Firstly, we applied a VGG16 pre-trained model on IQ sample images, and secondly, we trained a random forest classifier using pre-processed and segmented backscattered phase signals. In the second methodology, we introduced the ZCMD technique as an efficient segmentation to deal with varied speed of participants' gestures. Accuracy of 81% and 94% were attained for VGG16-based and phase-based models, respectively. Thanks to our adaptive gesture segmentation, the phase-based model could operate robustly in new environments with new participants with an accuracy drop of only 7%.

Acknowledgment

We acknowledge funding by the European Union through the Horizon Europe EIC Pathfinder project SUSTAIN (project no. 101071179). Views and opinions expressed are those of the authors and do not necessarily reflect those of the European Union.

We would like to thank Maryam Ghorbansabagh for her assistance with data collection for this study.

References

[1] R. Ghazalian *et al.*, "Joint user localization and location calibration of a hybrid reconfigurable intelligent surface," *IEEE Transactions on Vehicular Technology*, 2023.

[2] S. Palipana *et al.*, "Pantomime: Mid-air gesture recognition with sparse millimeter-wave radar point clouds," *Proceedings of the ACM on IMWUT*, vol. 5, no. 1, pp. 1–27, 2021.

[3] X. Jiang *et al.*, "Automatic RF leakage cancellation for improved remote vital sign detection using a low-IF dual-PLL radar system," *IEEE Transactions on Microwave Theory and Techniques*, 2023.

[4] A. Alabsi *et al.*, "Emotion recognition based on wireless, physiological and audiovisual signals: A comprehensive survey," in *International conference on smart computing and cyber security: strategic foresight, security challenges and innovation*. Springer, 2021, pp. 121–138.

[5] B. Yan *et al.*, "mmgesture: Semi-supervised gesture recognition system using mmwave radar," *Expert Systems with Applications*, vol. 213, p. 119042, 2023.

[6] C. Xu *et al.*, "Improving dynamic gesture recognition in untrimmed videos by an online lightweight framework and a new gesture dataset ZJUGesture," *Neurocomputing*, vol. 523, pp. 58–68, 2023.

[7] T. Fei, S. Mukhopadhyay, J. P. J. Da Costa, C. RoyChaudhuri, L. Lan, and N. Demitri, "Spatial environment perception and sensing in automated systems: A review," *IEEE Sensors Journal*, 2024.

[8] W. Qi *et al.*, "A resource-efficient cross-domain sensing method for device-free gesture recognition with federated transfer learning," *IEEE Transactions on Green Communications and Networking*, vol. 7, no. 1, pp. 393–400, 2023.

[9] Y. Yao *et al.*, "Interference-negligible privacy-preserved shield for RF sensing," *IEEE Transactions on Mobile Computing*, 2023.

[10] H. Landaluce *et al.*, "A review of IoT sensing applications and challenges using RFID and wireless sensor networks," *Sensors*, vol. 20, no. 9, p. 2495, 2020.

[11] S. Golipoor and S. Sigg, "Accurate RF-sensing of complex gestures using RFID with variable phase-profiles," in *IEEE 32nd International Symposium on Industrial Electronics (ISIE)*, 2023, pp. 1–4.

[12] Q. Bu *et al.*, "Deep transfer learning for gesture recognition with WiFi signals," *Personal and Ubiquitous Computing*, pp. 1–12, 2020.

[13] G. Jaswal *et al.*, "Range-doppler hand gesture recognition using deep residual-3DCNN with transformer network," in *Pattern Recognition. ICPR International Workshops and Challenges: Virtual Event, January 10–15, 2021, Proceedings, Part VI*. Springer, 2021, pp. 759–772.

[14] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[15] K. He *et al.*, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[16] S. Zhang *et al.*, "Real-time and accurate gesture recognition with commercial rfid devices," *IEEE Transactions on Mobile Computing*, 2022.

[17] J. Xiao and othes, "A survey on wireless device-free human sensing: Application scenarios, current solutions, and open issues," *ACM Computing Surveys*, vol. 55, no. 5, pp. 1–35, 2022.

[18] C. Zhao, L. Wang *et al.*, "Wear-free indoor fall detection based on RFID and deep residual networks," *International Journal of Communication Systems*, p. e5499, 2023.

[19] C. Zhao *et al.*, "RFID-based human action recognition through spatiotemporal graph convolutional neural network," *IEEE Internet of Things Journal*, 2023.

[20] M. Merenda *et al.*, "Edge machine learning techniques applied to RFID for device-free hand gesture recognition," *IEEE Journal of Radio Frequency Identification*, vol. 6, pp. 564–572, 2022.

[21] C. Wang *et al.*, "Multi-touch in the air: Device-free finger tracking and gesture recognition via cots rfid," in *IEEE INFOCOM 2018-IEEE conference on computer communications*, 2018, pp. 1691–1699.

[22] H. Xu *et al.*, "Rf-csign: A Chinese sign language recognition system based on large kernel convolution and normalization-based attention," *IEEE Access*, vol. 11, pp. 133 767–133 780, 2023.

[23] C. a. Dian, "Towards domain-independent complex and fine-grained gesture recognition with RFID," *Proceedings of the ACM on Human-Computer Interaction*, vol. 4, no. ISS, pp. 1–22, 2020.

[24] Z. Ma *et al.*, "RF-siamese: approaching accurate rfid gesture recognition with one sample," *IEEE Transactions on Mobile Computing*, 2022.

[25] K. Cheng *et al.*, "In-air gesture interaction: Real time hand posture recognition using passive RFID tags," *IEEE access*, vol. 7, pp. 94 460–94 472, 2019.
[26] H. Zhang *et al.*, "RF-sign: Position-independent sign language recognition using passive RFID tags," *IEEE Internet of Things Journal*, 2023.
[27] R. H. Clarke *et al.*, "Radio frequency identification (RFID) performance: the effect of tag orientation and package contents," *Packaging Technology and Science: An International Journal*, vol. 19, no. 1, pp. 45–54, 2006.
[28] D. Vasisht *et al.*, "In-body backscatter communication and localization," in *Proceedings of the 2018 Conference of the ACM Special Interest Group on Data Communication*, 2018, pp. 132–146.
[29] Y. Yu *et al.*, "RFID based real-time recognition of ongoing gesture with adversarial learning," in *Proceedings of the 17th Conference on Embedded Networked Sensor Systems*, 2019, pp. 298–310.
[30] N. Kargas *et al.*, "Fully-coherent reader with commodity SDR for Gen2 FM0 and computational RFID," *IEEE Wireless Communications Letters*, vol. 4, no. 6, pp. 617–620, 2015.
[31] S. S. Yadav and S. M. Jadhav, "Deep convolutional neural network based medical image classification for disease diagnosis," *Journal of Big data*, vol. 6, no. 1, pp. 1–18, 2019.
[32] S. Tammina, "Transfer learning using vgg-16 with deep convolutional neural network for classifying images," *International Journal of Scientific and Research Publications (IJSRP)*, vol. 9, no. 10, pp. 143–150, 2019.
[33] R. W. Schafer, "What is a savitzky-golay filter?[lecture notes]," *IEEE Signal processing magazine*, vol. 28, no. 4, pp. 111–117, 2011.
[34] A. Varri, "Digital processing of the eeg in epilepsy," in *Licentiate Thesis, Tampere University of Technology, Tampere, Finland*, 1988.
[35] H. Azami *et al.*, "An improved signal segmentation using moving average and Savitzky-Golay filter," 2012.
[36] A. V. Oppenheim, *Discrete-time signal processing*. Pearson Education India, 1999.