

# Micro-Doppler Gesture Recognition using Doppler, Time and Range Based Features

Matthew Ritchie  
Electronic and Electrical Engineering Department  
University College London  
London, WC1E 7JE  
m.ritchie@ucl.ac.uk

Aaron M Jones  
Sensors Directorate  
Air Force Research Laboratory  
Wright-Patterson AFB, Ohio 45433

**Abstract**—This paper presents micro-Doppler analysis and classification results from radar measurements of various hand gestures. A new database of 6 individuals completing 4 separate gestures with over 3,000 repetitions was recorded using a 24 GHz Ancortek radar system. The micro-Doppler signatures from these gestures were generated, features extracted and multiple different classifiers applied to this gesture data. A typical micro-Doppler classification process aims to use either a single range bin of data, average over a series of range bins or align all the target signal to a single bin. Different to previous techniques, the paper presents a method that uses multiple ranges bins to produce a spectrogram per range bin in order to represent the observed gesture over all four dimensions of time, Doppler, space and polarization. A comparison of the traditional and the newly proposed technique is shown and the improvements demonstrated are observed to be significant.

**Keywords**—Micro-Doppler, Classification, Machine-Learning, FMCW Radar

## I. INTRODUCTION

Micro-Doppler signals represent the component of a targets signature generated by vibration, rotation and so on modulated on top of the main bulk velocity and was first fully characterized by V. Chen [1]. These signatures contain a great deal of information and this has been exploited for broad range of different applications, including Jet Engine Modulation (JEM) analysis [2], analysis of ballistic target tumbling [3] and discrimination between bird and drone targets [4].

In the area of Radio Frequency (RF) sensing of people, [5] showed that micro-Doppler analysis of a person's gait could be used to discriminate individuals. Various human activities have been classified [6], along with discriminating humans or animals [7]. Multistatic radars have successfully been deployed for personnel recognition [8], [9]. In addition to human targets micro-Doppler has been explored for drone classification applications [10]. This recent significant increase in micro-Doppler classification research shows that it has many real world applications and provides vital information from a target unique to radar sensing.

The area of Human Machine Interface (HMI) control is well suited to short range low power RF sensors, and currently there is a strong trend in exploiting these sensors for diverse applications in this area. Companies such as Google [11] and Microsoft [12] are both actively researching in this area as well as a number of automotive companies. Rival technological solutions to RF sensors include optical imaging based sensors such as those that work alongside the Leap

Motion device [13] or a proposed 3D holoscopic camera solution discussed in [14]. Radar based sensors do have some advantages over optical based devices, these include resilience to variations in lighting conditions, reduced privacy concerns compared to a camera, the ability to sense through clothing (suitable for smart watch or phone interaction), and direct measurement of movement without the need for depth perception compensation.

This paper focuses on the application of a straightforward frequency modulated continuous wave (FMCW) radar sensor used to recognize four different gestures from a database of six individuals. Unlike previous classification results for micro-Doppler shown in [6]–[9] (and many other publications) the results presented here maintain the dimension of range in the classification process. It was hypothesized that maintaining features from multiple range bins would improve classification, as different gestures would be concentrated differently in the spatial domain. This would particularly be true when comparing gestures that have little movement in range, such as small hand actions, compared to movements that involve the individuals whole arm propagating over a number of range bins.

The remaining parts of this paper are as follows. Section II describes the experimental setup and database generated for this work along with the features extracted, Section III details the classifiers that were applied and shows the results from this processing by comparing single and multiple range bin results from various classifiers and finally Section IV concludes this paper and discusses potential future exploits for the research.

## II. EXPERIMENTAL SETUP AND DATABASE

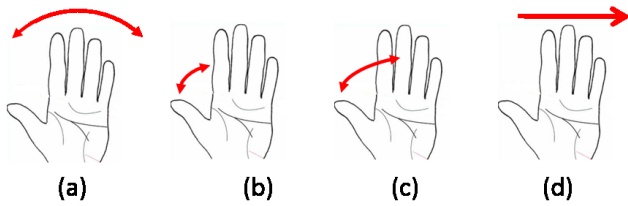
In this section, we describe the experimental equipment used to generate the data as well as the configuration of the geometry of the data capture and hardware settings. The amount of data generated for this database is then described along with how many individuals, gestures and repetitions that were used for classification. Finally, the signal processing chain used to analyze data is described, including key steps and features that are extracted from the data.

### A. FMCW Radar Hardware

The radar system used for the experiments was the Ancortek SDR-KIT 2400AD2 system. This is a 24 GHz FMCW radar that has up to 2 GHz instantaneous bandwidth and 1 ms chirp time. The sensor was setup with an output power of +13 dBm with horn antennas with 14 dBi gain. It was configured to use

a single transmit horn antenna in horizontal orientation (H-pol) and two receive horn antennas arranged in co and cross polarization (HH, HV, respectively). The radar was mounted on a desk and participants made hand gestures while sitting in front of the device at a range of 30 cm to 40 cm from the sensor.

The gestures recorded were four different actions; 1<sup>st</sup> waving (3-4 waves of the arm and hand), 2<sup>nd</sup> clicking (a single action between thumb and 2<sup>nd</sup> finger), 3<sup>rd</sup> pinching (between thumb and index finger) and 4<sup>th</sup> a swipe action (using the whole arm and hand). These gestures can be seen in a set of diagrams within Figure 1. The system was set to record 30 seconds of data and individuals completed as many repetitions as they could within this period. This process was repeated 10 times per gesture generating 40 data files for each individual.



**Figure 1** Hand gestures used (a) Wave (b) Pinch (c) Click (d) Swipe

The types of gesture can be subdivided into two groups with the waving and swipe actions generating a larger movement of arm and hand (perceived range migration) while the click and pinch action were more isolated in space (no range migration) with only the fingers moving. As we show, this distinction will help differentiate the gestures in range, which is a primary focus of the analysis of this paper.

These gestures were selected as they represent realistic HMI control actions. For example, the waving gesture is already utilized with the Xbox Kinetic sensor to initiate control and the pinch gesture is commonly used on smart phones to control zoom. Therefore, it is of interest if RF sensors are able to classify these commonly used actions as part of this newly proposed sensing modality. On generating the dataset, it was observed that individuals do complete each gesture differently and this does produce a diverse signature for the same action. This is clearly one of the challenges to the automatic recognition of these actions and a final implementation of the system may require the ability to learn an individual's actions to maximize classification success. This is partly explored via the blind classification analysis shown in Section III.

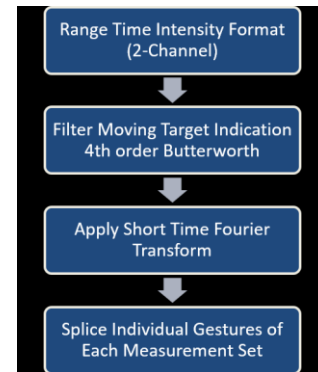
### B. Data Pre-Processing

The database generated includes 6 x 40 separate files from all the participants and all the repetitions. These participants are be labelled as A to F from now on. Each file was processed to produce a range-time matrix that covered the 30 sec recording. This was then manually segmented to extract the individual repeats within each recording. After this process the total database size had ground to 3164 measurements in total (6324 if considering both channels HH/HV of recorded data).

**Table 1** Table of number of captured gestures per person

Person	Gesture			
	<i>Pinch</i>	<i>Click</i>	<i>Swipe</i>	<i>Wave</i>
<i>A</i>	70	123	132	80
<i>B</i>	140	146	132	91
<i>C</i>	119	165	193	100
<i>D</i>	88	140	117	99
<i>E</i>	79	140	180	114
<i>F</i>	109	175	290	142

The processing flow for each recording can be seen in Figure 2. The FMCW signals are first formed into each channel of data and converted into Range Time Intensity (RTI) information. Then each file is then read in to a batch processing script which first stage is to filter the data using a Moving Target Intensity (MTI) process. The MTI filter applied was a 4<sup>th</sup> order Butterworth filter applied to the raw I/Q samples prior to conversion to range domain, further information on this processing is shown within [15][16]. This filter is used to enhance the moving components of the targets within the scene, as these methods are aimed a micro-Doppler classification and not on the static range profile information of the target present. In order to generate the spectrogram information a Short Time Fast Fourier Transform (STFT) is then applied with a window length of 0.2 sec and overlap of 95%. This has previously been shown to generate suitable micro-Doppler information for classification purposes [1], [5].

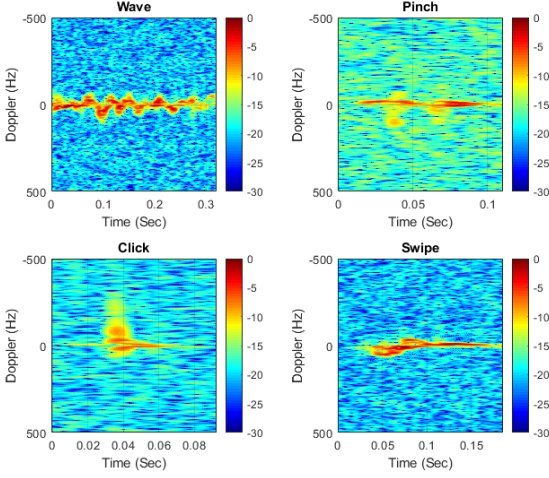


**Figure 2** Processing flow for each measurement

The data were then process using two different parallel techniques. The first was to take a coherent sum of the three adjacent range bins where the gesture action occurred in space, in order to produce a single summed range bin. This coherent summed signal was used to produce a single spectrogram. The results from this are labelled single range bin for the rest of this paper. The second method kept each of these range bins separate, thus maintaining the range dimension, and produced a spectrogram per range bin (three in total). These results are labelled multiple range bin for the rest of the paper. A key focus on this paper is on the comparison of these two methods in their ability to distinguish the various gestures.

Example plots of the spectrograms produced for four gestures is shown within Figure 3. These plots show that there are clear differences between the gestures in both Doppler and

time. The clink and pinch actions happen over a much shorter period, while the wave and swipe actions are more spread in time. It is important to note that these example plot do show a clear difference between classes, but this is not always the case across the whole dataset.



**Figure 3** Example spectrogram of the four gestures (Wave / Pinch / Click / Swipe). Data taken from Person A using a single range bin from the co-polar HH channel.

### C. Feature extraction

The data was then processed in order to extract the required features in order to successfully classify the different actions. This was completed on the single range bin data as well as three individual range bins that covered the area of maximum SNR of the action and the two adjacent bins. The features that were extracted have been obtained from analysis of the spectrograms directly. It has been shown that features can be extracted from different stages within the processing chain, from raw I/Q signals [17] to the RTI or MTI signatures [18].

Five features were extracted from each gesture example. Each of these features are defined within Table I below, where  $\mathcal{S}$  represents the amplitude values for the spectrogram matrix with  $n$  elements and  $\mu$  is the mean value for the spectrogram power (dB). For the entropy feature, the calculation uses the Shannon information definition of entropy and the  $p_i$  value is the probability of of seeing the  $i^{\text{th}}$  possible outcome  $\mathcal{S}$ . This was obtained from a histogram output of the  $\mathcal{S}$  matrix intensity values.

The eigenvalue based feature was extracted by applying Single Value Decomposition (SVD) analysis. The SVD analysis of an input matrix  $\mathbf{A}$ , (with rank of one) can produce the following output

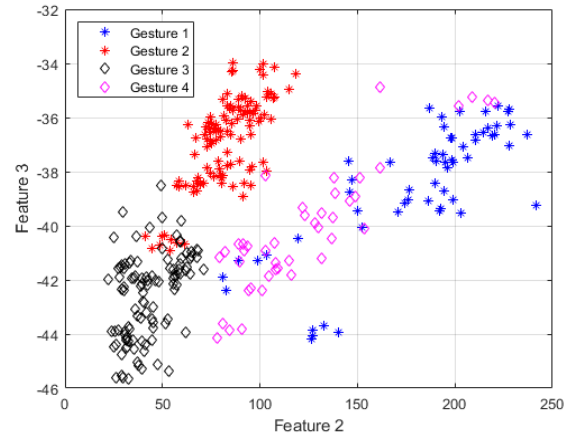
$$\mathbf{A} = \mathbf{u}\mathbf{\sigma}\mathbf{v}^T \quad (1)$$

where  $\mathbf{u} \in \mathbf{R}^m$ ,  $\mathbf{v} \in \mathbf{R}^n$ , and  $\sigma > 0$ . This can be expanded to when matrix  $\mathbf{A}$  has an arbitrary rank of  $r$ , where the matrix can be represented by a sum of rank one matrices. The  $\sigma$  matrix is commonly known as the singular values and is a diagonal matrix on non-negative real numbers. For the summation  $m$  is the length of the diagonal of the  $\sigma$  matrix.

An example feature space plot of the mean power of the spectrogram (feature 3) and the sum of the  $\sigma$  values (feature 2) from person A data in the co-polarised channel is shown in Figure 4. This plot shows that these two features have a good separation in the feature space and should lead to a reasonable classification performance.

TABLE I. EXTRACTED FEATURE DEFINITIONS

No	Feature	Equation
1	Spectrogram Intensity Summed	$\sum_{i=1}^n \mathcal{S}_i$
2	Spectrogram Variance	$\frac{1}{n} \sum_{i=1}^n ((20 \cdot \log_{10}(\mathcal{S}_i)) - \mu)^2$
3	Spectrogram Power (dB) Mean	$\frac{1}{n} \sum_{i=1}^n 20 \cdot \log_{10}(\mathcal{S}_i)$
4	Spectrogram SVD – Summed Singular Values	$\sum_{i=1}^m \text{diag}(\sigma_i)$
5	Entropy of Spectrogram Intensity	$-\sum_{i=1}^n p_i \log p_i$



**Figure 4** Example plot of Person A co-polarized features 2 and 3 plot

## III. CLASSIFICATION

### A. Methods

The classification processing applied to the obtained features was to use four separate classifiers. The classifiers selected for this paper were Linear and Quadrature Discriminate Analysis (DA) methods, a Random Forest (R-F) Fine-Tree mechanism and a K Nearest Neighbor (KNN) classifier with  $K=5$ . These were selected as testing with the quick look classification check using the Matlab *classificationLearner* utility showed them to be the most effective.

The DA methods are based around the assumption that the selected features are Gaussian distributed and statistically independent. For the linear case, the model has the same covariance matrix for each class, only the mean values change, while for the quadratic case, both the mean and covariance of each class can vary.

The R-F classifier used a supervised learning method of binary decisions deployed in a tree branched structure which enable the categorization of a set of input features. The tree is formed by considering all possible split combinations which are available from the input samples using an optimization criterion to fix these decisions thresholds. The optimization aims to minimize the Gini Diversity Index (GDI) which is defined as:

$$g(n) = 1 - \sum_k p^2(k) \quad (2)$$

where  $g$  is the GDI,  $k$  is the possible classes and  $p$  is the classification success rate that reached a given node in the RF decision tree.

Finally, the KNN classifier calculated the Euclidean distance between the test sample and adjacent training samples in the feature space. Based on the surrounding training samples that are the closest  $N$  (5 samples in this case) a decision is made on what class the new test sample belongs to.

Each classifier trained on a random subset of the features provided and then tested on the remaining data excluding this training set. The training subset size varied from 10% to 90% in steps of 10% in order to investigate the effects of training set on classification success rates. The classification process was repeated 100 times to produce an average result for each per training percentage size.

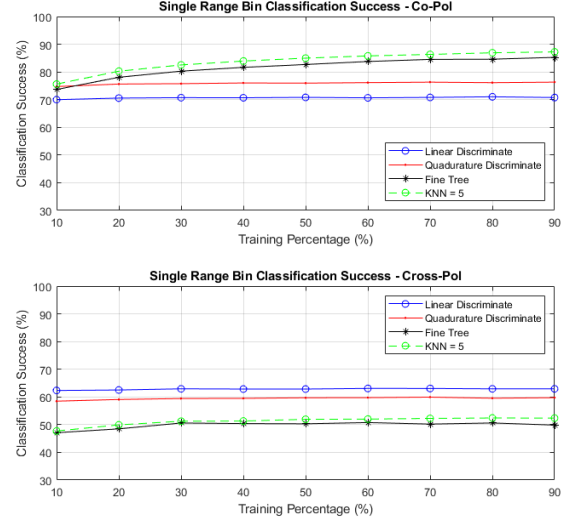
### B. Results

The results from processing either that co-pol or cross-pol data from each person using the four classifiers described above while taking only a single range bin of data (where the SNR of the gesture was maximum) are shown in Figure 5. It can be seen that, as a function of training data percentage, the linear and quadratic discriminant analysis classifiers were approximately flat; while the Random Forest R-F and KNN classifiers did improve as the percentage increased from 10 – 40% and then they leveled off. The best classification result of 87% from the single range bin data was obtained by the co-pol data with the KNN classifier at the highest training percentage of 90%. The cross-pol result maximum success was limited to 69% showing a significant disadvantage when using this information only. The result of the co-pol channel is a reasonably strong classifier success rate considering that this is a four class problem with a random chance of 25% and that there is a mixture of gestures recorded from different individuals which all have their own interpretation on how to perform the actions.

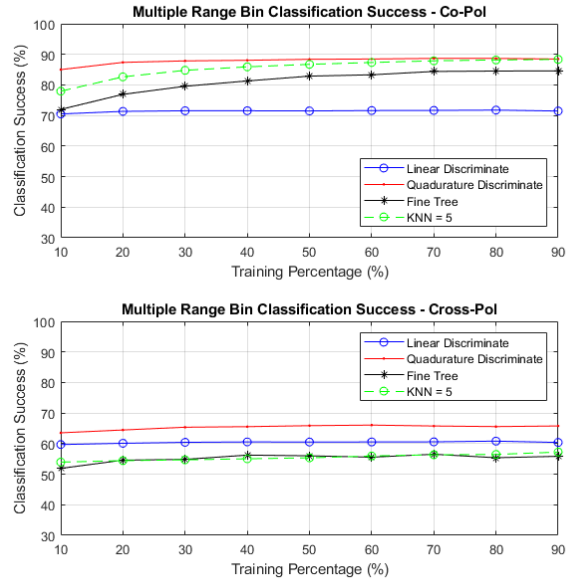
The multiple range bin data was then processed in the same way, for this analysis there was 3 times the features extracted from each range bins spectrogram, the results can be seen in Figure 6. These results show a peak classification success rate of 89%, which was achieved for the co-pol 90% training size result using the KNN classifier. This is very similar to that shown for the single range bin case, but it is worth noting that the the average classification success was much higher with a 10% different in success for the quadrature discriminate analysis method. This shows that for lower training set size there was a significant increase in classification success, but the difference is less noticeable at higher training set sizes.

The potential reason for not observing a significant increase in success rate is that the adjacent range bins features extracted contain a high degree of mutual information due to the range sidelobes of the processed signal. Therefore, limited additional information is obtained; the extent of mutual information could be quantified in the future to help evaluate this.

Within the outputs from the classifier it was noted that the most commonly misclassified outputs were generated when differentiating between clicking and pinching. These two actions happen over very short time frames and have a much less characteristic shape in the Doppler domain so it is understandable that they are the most challenging to classify.



**Figure 5 Classification success from single range bins using 4 classifiers for both co and cross-polarized data**



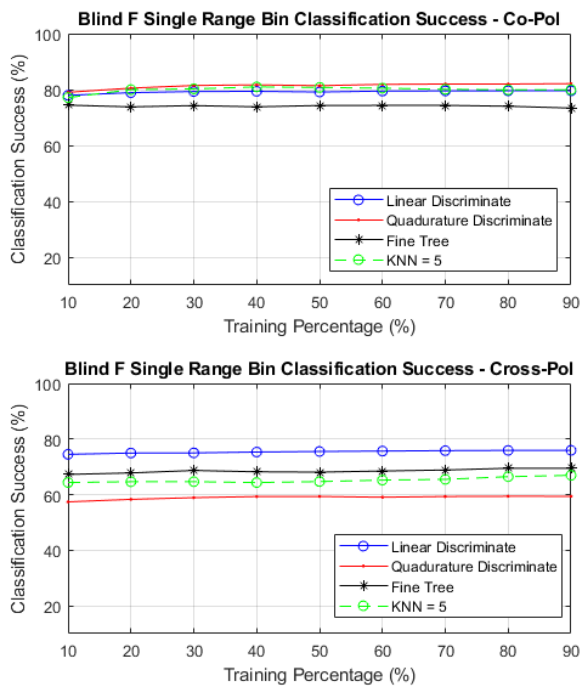
**Figure 6 Classification success from multiple range bins using 4 classifiers for both co and cross-polarized data**

The classifiers were also tested against blind input datasets, that have not been included in the training data. For this, data from 5 individuals was used to train a classifier and the data from the 6<sup>th</sup> individual was only used to test (not included in



training process). The results from this varied significantly, depending on which blind input dataset was used. Figure 7 and Figure 8 show the result generated for single and multiple range bin data respectively. In both of these cases the blind input to all the classifiers was from person F, which the training set was gathered from individuals A-E. Within the single range bin co-pol data the peak success rate was found to be 84%, while in the multiple range bin co-pol data a 96% peak was achieved. It is interesting to note that for the multiple range bin data we observe an increase in the blind classification results and the inverse is seen for the single range bin. The multiple range bin result is significantly higher than the general result shown in Figure 6. The lowest average results were obtained as 61% and 85% for single and multiple range bin analysis respectively; both for blind classification with person A. This demonstrates that person A may have a different signature in comparison to the others in the dataset and therefore without including their results in the training data you obtain a reduce success rate.

The aim of this analysis was to help understand how effective it is against an unknown individual's gesture. The outputs from this can help inform which feature sets and processing methods provide the best robust classification techniques against unknown users.



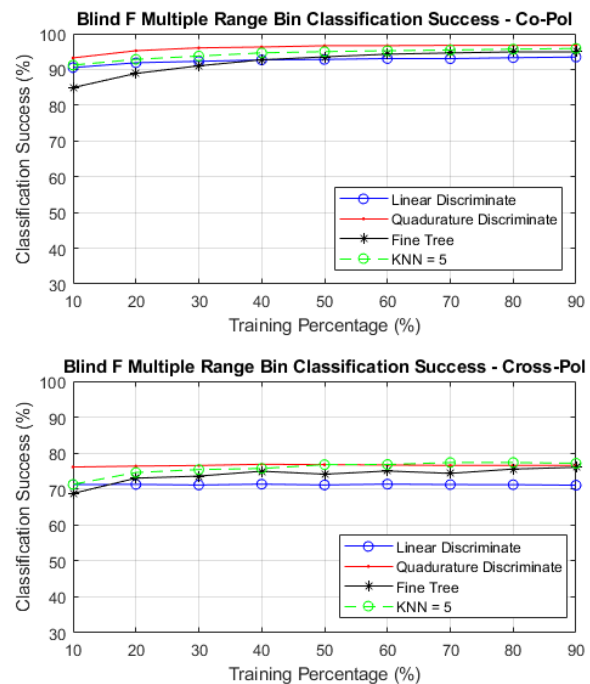
**Figure 7 Blind Classification success from single range bin using 4 classifiers for both co and cross-polarized data. The blind input individual was person F.**

#### IV. CONCLUSIONS

This paper has shown new radar data generated for this publication that was utilized to evaluate classification success rates for 4 different gestures by 6 individuals. The data was pre-processed in two different ways, one where adjacent range bins were coherent summed to produce a single range bin and spectrogram while the other method produced a spectrogram per range bin (3 in total). These two different input datasets were then used to extract features to classify

the actions. The multiple range bin/spectrogram technique was shown to achieve improved classification success rates and this difference was emphasized more at reduced training set sizes. Analysis was then performed on the resilience of the classifier to the situation when an individual was not included in the training set data. The multiple range bin processing technique showed an increased benefit in these situations and achieving up to 96% success rate in one case when person F was used in the blind test.

Future work will look to expand this into a real time classifier that aims to constantly evaluate spectrograms over the first N range bins in front of the radar and then take simple features from these spectrograms to continually update the predicted gestures that are occurring directly above the RF sensor.



**Figure 8 Blind Classification success from multiple range bins using 4 classifiers for both co and cross-polarized data. The blind input individual was person F.**

#### ACKNOWLEDGMENT

The authors would like to thank Colin Horne, Dr. Riccardo Palama', Alvaro Arenas Pingarron and Folker Hoffman for their support in data collection for this research.

#### REFERENCES

- [1] V. C. Chen and H. Wechsler, "Micro-doppler effect in radar: phenomenon, model, and simulation study," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 42, no. 1, pp. 2–21, Jan. 2006.
- [2] W. Y. Yang, J. H. Park, J. W. Bae, N. H. Myung, and C. H. Kim, "Automatic algorithm for estimating the jet engine blade number from the radar target signature of aircraft targets," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 30, no. 7, pp. 18–29, 2015.
- [3] A. R. Persico *et al.*, "On Model, Algorithms, and Experiment for Micro-Doppler-Based Recognition of Ballistic Targets," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 53, no. 3, pp. 1088–1108, 2017.
- [4] B. Torvik, K. E. Olsen, and H. Griffiths, "Classification of Birds

- and UAVs Based on Radar Polarimetry,” *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 9, pp. 1305–1309, 2016.
- [5] D. Tahmouh and J. Silvius, “Radar Micro-Doppler for Long Range Front-View Gait Recognition,” in *IEEE 3rd International Conference on Biometrics: Theory, Applications, and Systems*, 2009.
- [6] Y. Kim and H. Ling, “Human Activity Classification Based on Micro-Doppler Signatures Using a Support Vector Machine,” *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 47, no. 5, pp. 1328–1337, 2009.
- [7] Y. Kim, S. Ha, and J. Kwon, “Human Detection Using Doppler Radar Based on Physical Characteristics of Targets,” *Geoscience and Remote Sensing Letters, IEEE*, vol. 12, no. 2, pp. 289–293, 2015.
- [8] F. Fioranelli, M. Ritchie, and H. Griffiths, “Classification of Unarmed/Armed Personnel Using the NetRAD Multistatic Radar for Micro-Doppler and Singular Value Decomposition Features,” *Geoscience and Remote Sensing Letters, IEEE*, vol. 12, no. 9, pp. 1933–1937, 2015.
- [9] F. Fioranelli, M. Ritchie, and H. Griffiths, “Centroid features for classification of armed/unarmed multiple personnel using multistatic human micro-Doppler,” *IET Radar, Sonar Navig.*, vol. 10, no. 9, 2016.
- [10] B. K. Kim, H. S. Kang, and S. O. Park, “Drone classification using convolutional neural networks with merged doppler images,” *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 1, pp. 38–42, 2017.
- [11] J. Lien *et al.*, “Soli: Ubiquitous Gesture Sensing with Millimeter Wave Radar,” *ACM Trans. Graph.*, vol. 35, no. 4, pp. 1–19, 2016.
- [12] A. Alanis, T. Thai, G. Dejean, R. Gilad-backrach, and D. Lymberopoulos, “3D Gesture Recognition through RF Sensing,” 2014.
- [13] G. Marin, F. Dominio, and P. Zanuttigh, “Hand gesture recognition with jointly calibrated Leap Motion and depth sensor,” *Multimed. Tools Appl.*, vol. 75, no. 22, pp. 14991–15015, 2016.
- [14] Y. Liu, H. Meng, M. R. Swash, Y. F. A. Gaus, and R. Qin, “Holoscopic 3D micro-gesture database for wearable device interaction,” *Proc. - 13th IEEE Int. Conf. Autom. Face Gesture Recognition, FG 2018*, pp. 802–807, 2018.
- [15] M. Ash, M. Ritchie, and K. Chetty, “On The Application of Digital Moving Target Indication Techniques to Short-Range FMCW Radar Data,” *IEEE Sens. J.*, 2018.
- [16] M. Ritchie, A. Jones, J. Brown, and H. D. Griffiths, “Hand Gesture Classification Using 24 GHz FMCW Dual Polarised Radar,” in *IEEE International Radar Conference*, 2017.
- [17] C. Loukas and F. Fioranelli, “Activity Classification Using Raw Range and I & Q Radar Data With Long Short Term Memory Layers,” in *IEEE 16th Intl Conf on Dependable, Autonomic and Secure Computing, 16th Intl Conf on Pervasive Intelligence and Computing, 4th Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress(DASC/PiCom/DataCom/CyberSciTech)*, 2018, pp. 441–445.
- [18] D. B. Brown and J. S. Jones, “Classification of stationary movers with mti,” in *IEEE National Radar Conference*, 1996, pp. 112–117.