

Hand Gesture Classification Using 24 GHz FMCW Dual Polarised Radar

*M. Ritchie**, *A. Jones**, *J. Brown**, *H. D. Griffiths†*

**Radar Sensing Team, Defence Science and Technology Laboratory, Porton Down, Wilts, SP4 0JQ, UK.,*

†Engineering Dept, University College London, WC1E 7JE, UK

Keywords: FMCW Radar, Micro-Doppler, Machine Learning, Classification, Autonomy

Abstract

This paper evaluates the classification performance of a dual polarised on receive, 24 GHz Frequency Modulated Continuous Wave (FMCW) radar system to autonomously identify micro-Doppler signatures of unique hand gestures. We employ an Eigen subspace feature selection technique on the calculated signal subspace in order to classify each gesture. Measurements using the dual polarised radar, permitting simultaneous recording of both the co-pol and cross-pol returns, are evaluated with this processing technique and results are reported herein. Our analysis displays the challenges presented by the high variance in individuals gestures and the limited additional information the cross polarised returns have provided to the classifier. Classification performance comparisons are presented when co, cross and dual polarised data are provided to the classifier. With this technique we achieve autonomous classification performance of up to 84.6% when Eigenvalue derived features are used for classification.

1 Introduction

The use of radio frequency (RF) systems to autonomously interpret human actions is an emerging area of research. This new opportunity for human machine interface (HMI) shows significant potential as a viable solution for machine control in a number of scenarios. One of the phenomena that this technology exploits is the Micro-Doppler effect. That is, the effect describing the resulting additional frequency components on top of a bulk Doppler motion generated by motion within the object such as rotating or vibrating components [1]. Large technology companies are currently investing significant funds into these technologies which are on the verge of wide scale commercialisation of RF sensors for this purpose, for example the Google Soli Project [2]. This HMI method also has applications in automotive vehicle interaction, allowing hand gestures in front of a steering wheel column for vehicle controls without the need of taking your eyes off the road.

We note that optical sensors have previously been used in numerous examples to perform hand gesture recognition [3].

However, RF sensors have several advantages over optical sensors as they are unaffected by background lighting conditions, can measure velocity of the actions directly, and produce more anonymous data in comparison to an optical camera sensor. For these reasons, there have been a number of research papers published on this topic within recent years. This emphasis has also been driven by the miniaturizations and reduced cost of high frequency RF sensors that can produce the Doppler and range measurement fidelity required to quantify the more subtle human hand gestures.

FMCW radar has previously been shown to be suitable for gesture recognition within [4], where a prototype sensor was used to detect and track hand movements. It was also demonstrated that sparse signal processing with Gabor atom libraries can be used in order to classify different human gestures [5]. Large scale experiments of full body micro-Doppler signatures were shown within [6] but this research focused on full gait analysis of an individuals walk, not the more subtle hand gesture features evaluated here.

Within this paper a FMCW radar system, developed by *An-cortek*, is used to gather a database of human hand gestures from multiple individuals. We process the data, see Section 2 for details, making it suitable for classification and report the result of the performance in Section 3. We note the novel aspects of this collection are the dual polarised returns for this frequency band as well as the Eigen subspace analysis technique. Conclusions and future suggested work are found in Section 4.

2 Experimental Setup and Theory

In the section to follow we describe the experimental set-up by which representative micro-Doppler data was collected for processing. We then outline the complete signal processing chain applied to the raw data highlighting key post-processing steps along the way, defining the features that were extracted and reviewing the classifier output results for a combination of different scenarios.

2.1 FMCW Radar Experimental setup

The radar system operates at 24 GHz with a bandwidth set to 750 MHz (generated via linear chirp), transmitting at +13

dBm power using a horizontal polarisation transmit horn antenna. The two receiver channels were set to receive both the co and the cross polarised returned signals at adjacent independent antennas. The radar sensor was mounted approximately 25 cm above desk height and horizontally aligned with the test subject to be measured.

The gestures were performed at a distance of no more than 0.5 m from the sensor at a constant height while the individual was sitting down using their dominant hand. The four gestures performed were a pinch (between thumb and index finger), click (single action between thumb and 2nd finger), swipe (a single wave of the hand/arm from left to right) and wave (three waves of the hand palm towards the radar system). We note, at this bandwidth, the range resolution of the system is 20 cm. This is larger in size than the subtle hand based gestures being made hence the requirement for micro-Doppler processing in order to differentiate the different gestures. These were selected as they were either already used as part of existing HMIs or have good potential for common applications. For example, the Xbox Kinetic sensor uses a wave action to initialise communication between the sensor and the user and the Google Soli™ sensor observes pinch movements to emulate clicking a button.

2.2 Measurement Data Pre-Processing

To prepare the measured data for classification several pre-processing steps are applied, the complete list of these steps are shown in Algorithm 1. Each data capture represented a recording of multiple repeats of the same gesture and individual over a period of 30 seconds, which was then manually spliced into individual gestures. The measurement database consisted of 10 sets of 30 second recordings of each gesture from each individual, some pre-processing steps were required to extract the features from the data for classification. The total quantity of measurements of all four gestures from the five individuals was 1153 for each the feature sets. As there was a frequency and time domain hermitian feature set for both receive channels this generates a total feature set size of 4412.

We begin with suitably sampled in-phase and quadrature data samples from the receiver which are ingested into Matlab. The first step, then, is to format the data into the Range Time Intensity (RTI) domain which was then converted into a Moving Target Indication (MTI) signal. The MTI processor applied was a high-pass 4th order Butterworth filter along the slow-time dimension. We note the MTI processing removes background static clutter leaving a signature that is dominated by the moving gesture. This processing step is important for this analysis as it removes the dominant static clutter component that can degrade classification performance. The very narrow Butterworth MTI filter is useful here as the gestures pass through 0 Hz and having a narrow filter maintains the continuity of the signals of interest. However, this pre-processing limits the classifier to non-zero Doppler gesture recognition. For example, a static gesture such as holding the hand open and still would not be recognized.

Algorithm 1 Processing Steps

- 1: **for all** Measurements Sets **do**
 - 2: Read Range Time Intensity Format (2-Channel)
 - 3: Filter Moving Target Indication 4th order Butterworth
 - 4: Apply Short Time Fourier Transform
 - 5: Splice Individual Gestures of Each Measurement Set
 - 6: Perform Feature Extraction (1 of 3 subspace methods)
 - Singular value feature extraction
 - Eigenvalue subspace feature extraction
 - 7: Train Discriminant Analysis Classifier
 - 8: Perform Classification
 - 9: **end for**
-

The next pre-processing step is to apply a Short Time Fast Fourier Transform (STFT) with window length of 0.2 sec, overlap 95% and padding of a factor of 4 on the MTI processed data. These selected processing parameters have been successfully applied in previous papers classifying human micro-Doppler [6, 7]. This processing is performed on the coherent summation of the ranges bins of interest (i.e, where the signatures were observed), typically up to 3 or 4 in our experiments. Example spectra generated with and without MTI processing are shown in Figure 1(a) and Figure 1(b) respectively. The MTI processing improves signal-to-noise ratio (SNR) in this domain to produce data that is dominated by the signal of interest, the hand gesture motion, instead of static clutter. It is this Doppler spectrogram that is used to extract information (i.e. features) in order to classify the different actions. Previous published classification techniques, [7], have employed methods that manually extracted empirical features from these signals, such as the period of walking gait, but in this case the focus was on generating a fully automated processing and feature extraction data pipeline which has greater potential for practical implementation.

2.3 Signal Subspace Determination

After MTI and Fourier processing we perform two feature extraction methods on the data. In the first method, we compute the singular value decomposition (SVD) on the data matrix $\mathbf{X} \in \mathbb{R}^{m \times n}$ where m, n are the number of rows and columns, corresponding to the Doppler and time dimensions. The second method performs an Eigen decomposition on the hermitian matrix of the received data $\tilde{\mathbf{X}} = \frac{1}{2}(\mathbf{X}^* \cdot \mathbf{X})$ and derives features from the entire Eigen spectrum. The Eigen spectrum is then decomposed into the frequency and time domain hermitian components of the received data. This is restricted to a signal subspace, whose rank is determined by the computed Minimum Description Length (MDL) as described in [8]. This is to observe the consequence of removal of noise on the classification performance.

The second technique is applied to observe only the component of the signal that contains the target response. If the signal and noise can reasonably be considered complex, independent,

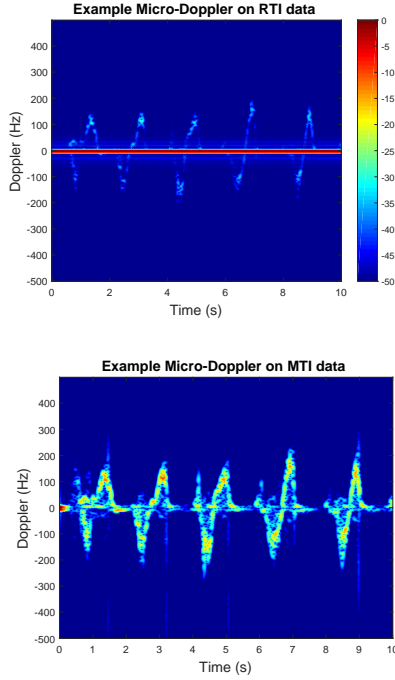


Fig. 1: Micro-Doppler spectrogram generated from a swipe gesture generated from the (a) Pre-MTI (b) MTI processed data, showing multiple repeats of the gesture.

stationary and ergodic Gaussian random processes then these assumptions would give a general Gauss-Gauss model. This is a common assumption in signal processing as it allows the model to be mathematically tractable. However, other distributions on the signal and noise are permitted, but would require a reformulation of the maximum likelihood estimate used to compute the rank of the signal subspace with this technique.

As is the case in many signal processing problems involving radar, detection of the number of signals (embedded in additive noise) is an important issue. For our concerns, when given a sample spectrogram we desire to make a classifier decision using information dominated by the signal of interest. Accurate estimation of the rank of the signal subspace allows for separation of the signal from the noise in order to use suitable limited features for classification from the measured data.

Within Figure 2 the largest N eigenvalues are highlighted in green which have been defined by the MDL model order determination. This method has been shown to be good at isolating the key components of the micro-Doppler data and hence should be suitable for feature extraction and then classification. The example shown has a clear delineation between signal and noise which makes the task of isolating the key components easier, but this will not always be the case and through the automated application of this proposed MDL processing this decision can be made using sound information theoretic concepts.

The four features taken from the signal subspace eigenvalue

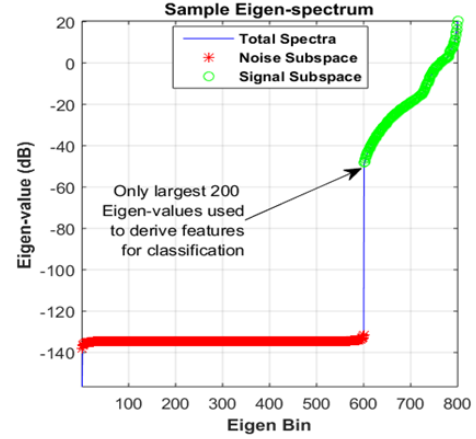


Fig. 2: Representation of eigenvalues of the length of the model order.

were :

- Rank of data
- Sum of the signal subspace eigenvalues (analogous to the signal power)
- The first and second moments of the sample eigenvalues.

As described above, in order to extract these features from the micro-Doppler signatures the data is transformed using Eigen-decomposition processing which takes the initial Doppler spectrogram from a single gesture and evaluates the generalised eigenvalues for its Hermitian matrix in both dimensions (time and frequency e.g. $\mathbf{X}^*\mathbf{X}$ or $\mathbf{X}\mathbf{X}^*$). The eigenvectors represent the key components of the data and the MDL threshold method enables a selection of only a subset of these values in-order to define features by which the original data are represented. Hence there are 2 lots of these 4 features from both the time and frequency projected Hermitian matrices. As there is also two channels of data (co and cross polarised) this results in a total possible feature set of 16 different features.

For comparison a set of Singular Value Decomposition (SVD) features were also extracted from micro-Doppler signatures. The application SVD processing allows the data to be decomposed from it's original matrix to a reduced dimensional space. SVD analysis performed on the original data matrix \mathbf{X} produces $\mathbf{X} = \mathbf{U}\mathbf{S}\mathbf{V}^T$, where \mathbf{S} is a diagonal matrix with the singular values of \mathbf{X} , and \mathbf{V} and \mathbf{U} are the matrices containing the right and left singular vectors of \mathbf{X} .

The features selected to extract from \mathbf{U} and \mathbf{V} have been previously applied within [7, 9]. These five empirically selected SVD features were the sum of the first 10 columns of \mathbf{U} , the standard deviation of first 10 columns of \mathbf{U} and \mathbf{V} and the mean and standard deviation of the diagonal component of \mathbf{U} . These have been applied successfully to micro-Doppler clas-

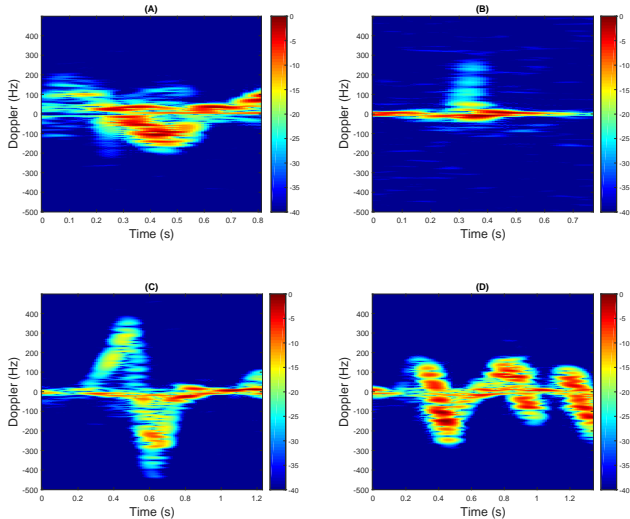


Fig. 3: Example of the Micro-Doppler signatures for the four actions (a) Click (b) Pinch (c) Swipe (d) Wave

sification previously and will be used to baseline the performance of the MDL feature sets in this paper.

An example of an extracted signal action spectrogram for each of the four actions can be seen in Figure 3. There are clear differences in frequency spread, duration and overall shape from one action to the next. There was also a level of variation between repeats of the same action and indeed in how each test subject performed the actions. It is important to extract fundamental features that can be robust to these changes.

3 Classification

A series of different classification tests were performed on the total dataset generated. This includes all 5 individuals and all 4 gestures, only 2 of the gestures from all the individuals and data from just one individual performing all gestures. In the full gesture classification examples this challenge represents a 4 class problem while the two gesture examples are the less challenging 2 class problem. The classifier that has been applied, for both MDL and SVD approaches, is a K-Nearest Neighbour (K-NN) classifier using 5 nearest neighbours. This is a simple classifier which measures the Euclidean distance between a given feature space sample and those surrounding it. Further details on this classifier can be found here [10]. Other options included linear or quadratic discriminant analysis but due to the irregular clustering of the samples present in the feature space these methods were found to be not as effective as the K-NN. The K-NN classifier was trained on a randomly selected subset percentage of the whole feature set and then tested on the remaining data. 50 Monte Carlo repetitions were used to obtain an average classifier result for a given training set size, each with a randomly selected training feature subset.

The K-NN classifier was initially applied to all datasets using the 16 feature array of data from all 4 MDL features for both the time and frequency domain eigenvalues from both the co and cross polarised datasets, as well as the 10 comparative SVD features. The results from this are shown in Figure 4. These results for four gesture recognition show a relatively low classification success rate between 45% and 60%, in comparison to a random chance success of 25%.

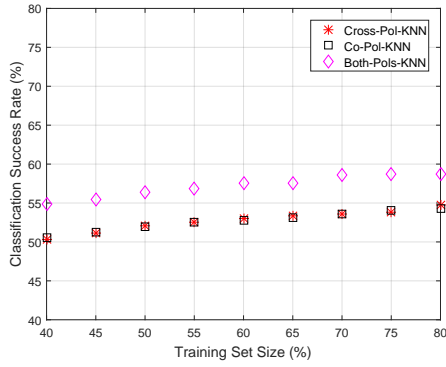
These first results show that the Eigen features defined by the MDL have a consistently improved performance over the SVD based feature sets by an average of 5%. The MDL extracted features results for the independent co and cross polarised feature sets were comparable, while the best result was obtained when using both the co and cross polarised feature sets jointly in the single classifier. For the SVD features the cross polarised results were 5% lower than the independent co or both polarisations.

We believe the relatively low classification rate is partly due to the large intra-class variance, which was produced by the different way the 5 individuals performed the gestures. This shows that RF gesture recognition may require user specific training on the gestures to ensure it is recognising that specific person's action, or improved selection and pre-processing on the extracted features.

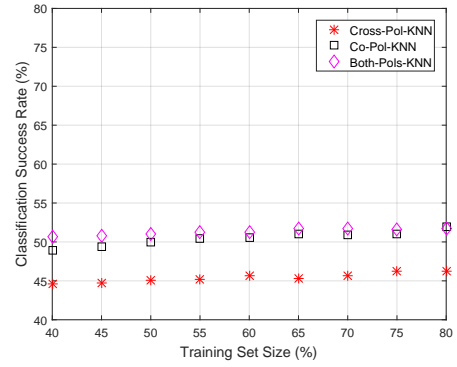
Figure 5 shows the result when the database was limited to a two gesture problem (the click and wave gestures) but still using data from all 5 individuals. This improved the success rate significantly although it is still lower than that seen in other full body motion human micro-Doppler classification problems [9]. The MDL Eigen values again outperformed the SVD features by up to 10%. The difference between co, cross and both polarisations was very limited for these features showing that the cross polarised features were not adding much additional information for the classifier to make a decision.

The data from a single person over all 4 gestures was then evaluated in Figure 6. This showed the greatest difference between the MDL Eigen features and the SVD, with the MDL based features showing > 20% in some results.

In all the results shown so far the input features sets from the MDL extracted Eigen space data are taken from both the frequency and time domain space. Within Figure 7 the mean classification success across 40% to 80% was evaluated for the time, frequency and joint time-frequency Eigen space features. These results were generated by separating out features that were derived solely from either the time or frequency domain Eigen space or both used jointly. This result shows that the worst performance was obtained from the frequency Eigen space feature set of the cross polarised data, and the best results was from the joint set of time and frequency feature Eigen feature sets from both polarisations; although the difference across these results was limited to only 6%.

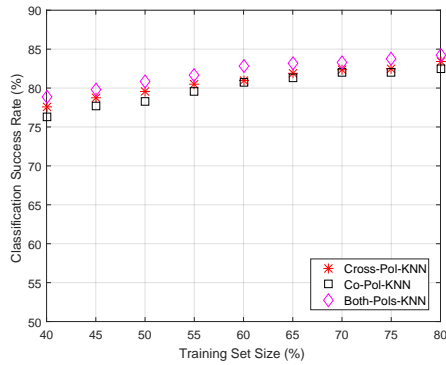


(a) MDL selected Eigen Features

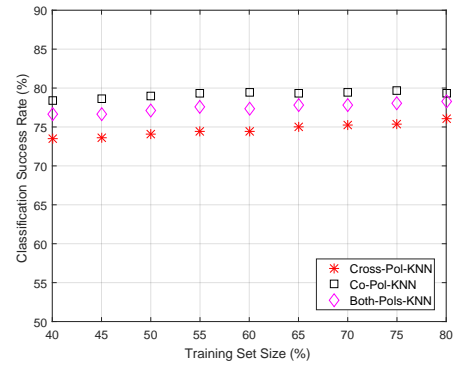


(b) SVD Features

Fig. 4: 5 person 4 gesture K-NN Classifier results



(a) MDL selected Eigen Features



(b) SVD Features

Fig. 5: 5 person 2 gesture K-NN Classifier results

4 Conclusions

This paper has demonstrated the classification potential of a 24 GHz FMCW radar sensor in recognising various hand gestures. A novel application of an information theoretic approach using features derived from the MDL limited eigen-space metric of the micro-Doppler data was demonstrated. The results from a K-NN classifier have shown that similar classification performance was obtained with co, cross and both polarization datasets.

The results have shown that using both co and cross polarised data can provide slightly higher classification rates but this additional gain was found to have less influence in comparison to other factors such as selected feature sets, or whether Eigen-MDL or SVD inputs are used. This is likely to be due to the high level of mutual information within the two channels, which could be reduced by separating the receiver antenna location to provide a more diverse perspective of the gesture.

In the data generated, the signatures of the micro-Doppler actions had SNR levels of $> 30\text{dB}$ which does allow for clear observation of the movements however for the MDL subspace extraction this may not be its optimum use case. In the case where there is a low SNR signal embedded in noise then us-

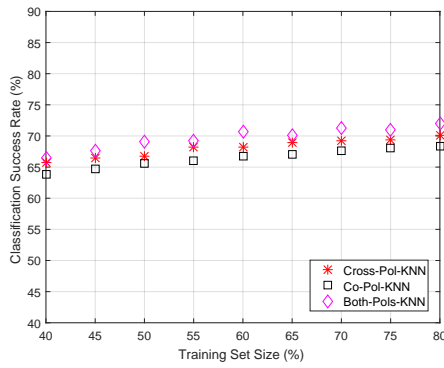
ing this MDL method to define the area of the subspace should show greater benefits over traditional techniques.

The extraction of the gesture action in time is an important aspect of this processing; the methods used did not attempt to align the gestures in time for each spliced section that was extracted thus making a greater variability and a more challenging dataset to classify on. Future research will look into resilience to temporal alignment. In a practical real-time system the classification algorithms will need to be operative very quickly over a sliding window of data which is constantly streaming, future work will look into the implementation of these methods.

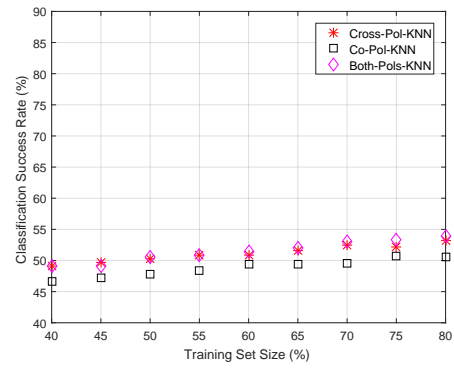
Acknowledgments

The authors would like to acknowledge both Dr. Francesco Fioranelli and Prof. Gang Li for their helpful discussions and collaborations within this domain, as well as Dr. Joseph Spencer, Mr. Richard Walkerdine and Ms. Emma Taylor for their assistance.

Content includes material subject to DSTL/WP103242 © Crown copyright (2017), Dstl. This material is li-



(a) MDL selected Eigen Features



(b) SVD Features

Fig. 6: 1 person 4 gesture K-NN Classifier results

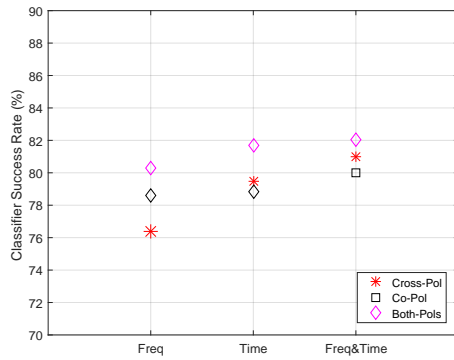


Fig. 7: 5 person 2 gesture K-NN classifier average result when using time domain, frequency domain or both as inputs to the classifier

censed under the terms of the Open Government Licence except where otherwise stated. To view this licence, visit <http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3> or write to the Information Policy Team, The National Archives, Kew, London TW9 4DU, or email: psi@nationalarchives.gsi.gov.uk

References

- [1] V. C. Chen, F. Li, S. S. Ho, and H. Wechsler, "Micro-doppler effect in radar: phenomenon, model, and simulation study," *IEEE Transaction on Aerospace and Electronic Systems*, vol. 42, pp. 2–21, 2006.
- [2] J. Lien, N. Gillian, M. E. Karagozler, P. Amihoud, C. Schwesig, E. Olson, H. Raja, and I. Poupyrev, "Soli: Ubiquitous gesture sensing with millimeter wave radar," *ACM Trans. Graph.*, vol. 35, pp. 142:1–142:19, July 2016.
- [3] S. S. Rautaray and A. Agrawal, "Vision based hand gesture recognition for human computer interaction: a survey," *Artificial Intelligence Review*, vol. 43, pp. 1–54, 2015.
- [4] P. Molchanov, K. Gupta, S. Kim, and K. Pulli, "Short-range fmcw monopulse radar for hand-gesture sensing," in *Proceeding of IEEE Radar Conference*, pp. 1491–1496, 2015.
- [5] G. Li, R. Zhang, M. Ritchie, and H. Griffiths, "Sparsity-based dynamic hand gesture recognition using micro-doppler signatures," in *Proceeding of 2017 IEEE Radar Conference*, 2017.
- [6] F. Fioranelli, M. Ritchie, S. Z. Gurbuz, and H. Griffiths, "Feature diversity for optimised human micro-doppler classification using multistatic radar," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 53, pp. 640–654, 2017.
- [7] 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," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, pp. 1933–1937, Sept 2015.
- [8] M. Wax and Kailath, "Detection of signals by information theoretic criteria," *IEEE Transactions Acoust. Speech Signal Processing*, vol. 33, pp. 387–392, April 1985.
- [9] F. Fioranelli, M. Ritchie, and H. Griffiths, "Performance analysis of centroid and svd features for personnel recognition using multistatic micro-doppler," *IEEE Geoscience and Remote Sensing Letters*, vol. 13, pp. 725–729, May 2016.
- [10] T. Hastie, R. Tibshirani, and F. J., *The Elements of Statistical Learning: Data Mining Inference and Prediction*. Berlin, Germany: Springer Series in Statistics, 2009.