



Gunshot Recognition using Low Level Features in the Time Domain

HRABINA, M.; SIGMUND, M.

Proceedings of the 2018 28th International Conference Radioelektronika (RADIOELEKTRONIKA)

pp. 1-5

eISBN: 978-1-5386-2485-2

DOI: <http://dx.doi.org/10.1109/RADIOELEK.2018.8376372>

Accepted manuscript

©2018 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. Martin Hrabina, Milan Sigmund, " Gunshot recognition using low level features in the time domain ", Proceedings of the 2018 28th International Conference Radioelektronika (RADIOELEKTRONIKA), pp. 1-5, 2018. DOI: 10.1109/RADIOELEK.2018.8376372. Final version is available at <https://ieeexplore.ieee.org/document/8376372/>

Gunshot Recognition using Low Level Features in the Time Domain

Martin Hrabina, Milan Sigmund

Brno University of Technology
Department of Radio Electronics
Technicka 12, 61600 Brno, Czech Republic
hrabina@phd.feec.vutbr.cz, sigmund@feec.vutbr.cz

Abstract— This paper explores the possibility of using scarcely used time-domain features for the task of gunshot recognition. A set of 11 features derived from temporal characteristics (waveform) of signals is calculated from a mixed dataset of gunshots and non-gunshots. The features leverage the impulsive nature of gunshots and their dissimilarity to other, especially more stationary signals. The paper includes a description of feature extraction, distribution of features and their recognition performance on a selected audio dataset. A subset achieves promising results in comparison with more frequently used spectral-domain features. This makes them a valuable addition to other frequently used features, especially for tasks of impulsive sound recognition.

Keywords—*gunshot detection, feature extraction, time-domain features, audio processing*

I. INTRODUCTION

The aim of this paper is to explore some less frequently used features, some of which, to our knowledge, were not explored before. Most authors dealing with sound recognition mainly use features derived from a spectrum (such as LPC [1-2] and MFCC [3], spectral energy and MPEG-7 descriptors [4] and spectrogram features [5]) with temporal features being rare (e.g. zero-crossing rate). Some of the more frequently used features, along with the task for which they are used (such as simple detection, localization, classification etc.) and reference to formula for calculation, are listed in Table 1.

TABLE I. SOME FEATURES USED BY OTHER AUTHORS

Feature	Goal	Paper	Calculation
Linear prediction coefficients (LPC)	Gunshot detection, shout detection,	[1], [2]	[6]
Mel-frequency cepstral coefficients (MFCC)	Gunshot detection, event classification, acoustic scene classification	[1], [2], [3]	[6]
Perceptual linear prediction coefficients (PLP)	Shout detection	[2]	[7]
Zero-crossing rate	Gun type recognition, detection	[4]	[4]
Spectral energy	Gun type recognition	[4]	[4]
MPEG-7 descriptors	Gun type recognition	[4]	[4]
Spectrogram image features (SIF)	Audio event detection	[5]	[5]

In our pursuit of temporal features, we suppose that gunshots have an unusual waveform but are similar enough to each other. Some of the features might be susceptible to noise, however at this point we are working with recordings without added noise. Preliminary results show usability of some of the presented features. If further research proves their viability, investigation on the effects of various types of noises (white noise, impulsive, natural noises, etc.) and possible mitigation of these effects will follow.

The rest of the paper is organized as follows. Section 2 briefly describes the audio dataset we are working with and the preprocessing of input audio. Section 3 defines the proposed features and provides some explanatory figures. Section 4 summarizes the results from a statistical viewpoint, provides detection results using a simple neural network and compares them to other papers using different features.

II. DATASET AND PREPROCESSING

This work uses audio recordings from multiple sources. Gunshots are from the Free Firearm Library [8], from which we filtered out far-away recordings (defined by energy criterion described at the end of the paragraph) and subsequently created two subsets for testing. The first includes only recordings of an AK-47 – which is the most frequent weapon in the library (374 occurrences). The second, a greater subset, includes all weapons (1483 occurrences). For non-gunshot sounds, a subset of the Urban Audio Dataset [9] was chosen, recordings of public spaces from the Airborne website [10], a small part of recordings comes from Freesound [11], as well as our own recordings. Non-gunshot set contains 9147 segments. During the selection process, all segments with energy below 0.1 (defined as a sum of squared samples) were omitted, in order to exclude any silent segments. Recordings contain no additional noise, except for noise present during recording, as this paper does not investigate influence of noise on feature performance.

All audio frames have a length of 11 ms (486 samples at sampling frequency of 44.1 kHz) and have been normalized so that the maximum absolute value is 1. The frame length was selected in accordance with our previous research [12].

Matlab 2016b was used for all calculations and audio processing.

III. INVESTIGATED FEATURES

Gunshots are usually distinctly N-shaped, however the most dominant extreme of some gunshots is positive and of others negative. Fig. 1 and 2 depict both these cases.

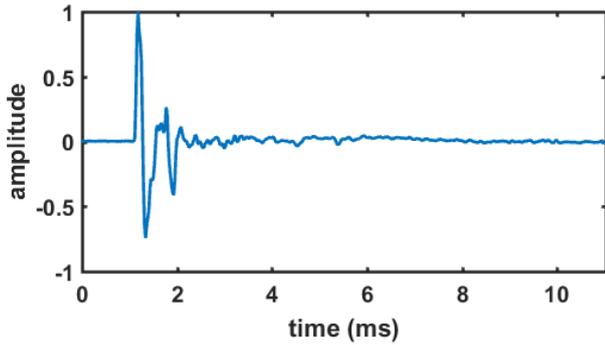


Fig. 1. AK-47 gunshot waveform – positive dominant peak

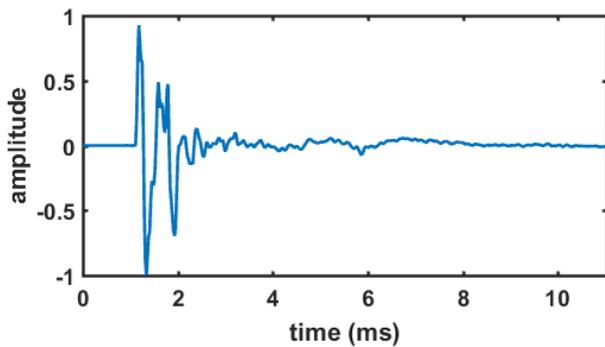


Fig. 2. AK-47 gunshot waveform – negative dominant peak

The first investigated features are the relative position of zero-crossings before (RP-) and after (RP+) the most dominant peak, as well as distance between these zero-crossings (ZDist). These features are depicted in Fig. 3. Fig. 4 presents a histogram of zero-crossing positions and Fig. 5 shows a histogram of distances between these positions for an AK-47.

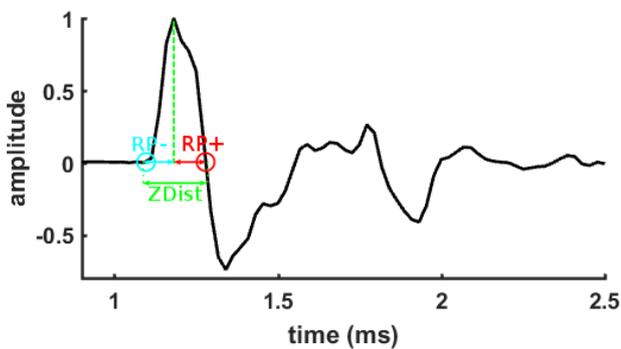


Fig. 3. Feature illustration: zero-crossing points (RP- and RP+)

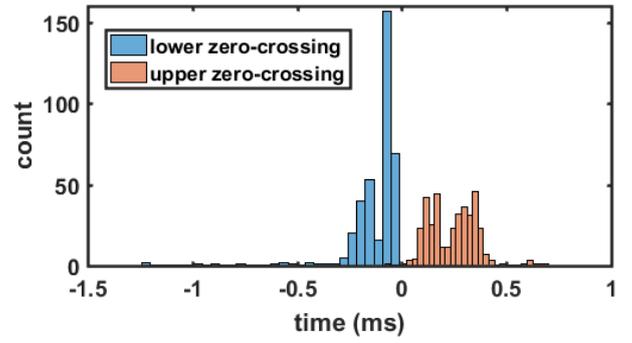


Fig. 4. Distribution of zero-crossing points (RP- and RP+)

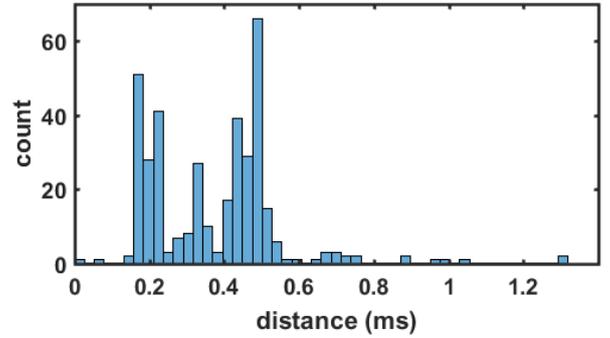


Fig. 5. Distribution of the distance between zero-crossings (ZDist)

More features, some of which are illustrated in Fig 6., are defined as: time distance between maximum and minimum values (PDist), PDist can acquire positive as well as negative values, as can be seen in Fig. 10. This is caused by the fact that minimum sometimes precedes maximum. Next features are, distance between these points in two dimensions (PIDist), angle between horizontal line and line connecting maximum and minimum points (Ang). In order to calculate the angle, the horizontal line is measured in seconds. The area of the triangle delimited by the 2 highest peaks and a minimum (referred to as Area). Area is illustrated in Fig. 8 and Fig. 11. The last 4 features, shown in Fig. 9, are coefficients (A and B in equation 1) of exponential fit to both positive local extremes (Ap, Bp) and negative local extremes (An, Bn).

$$f(t) = A \cdot \exp(B \cdot t) \quad (1)$$

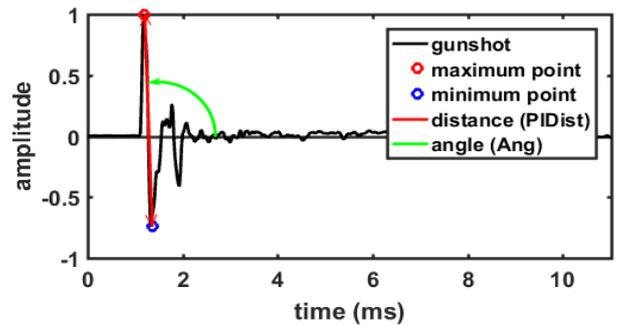


Fig. 6. Feature illustration: peak distance (PIDist) and Angle (Ang)

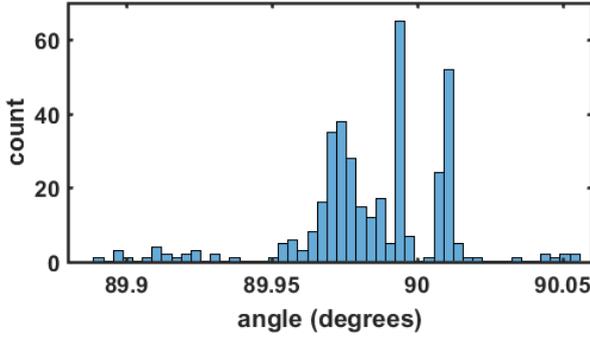


Fig. 7. Distribution of angles (Ang)

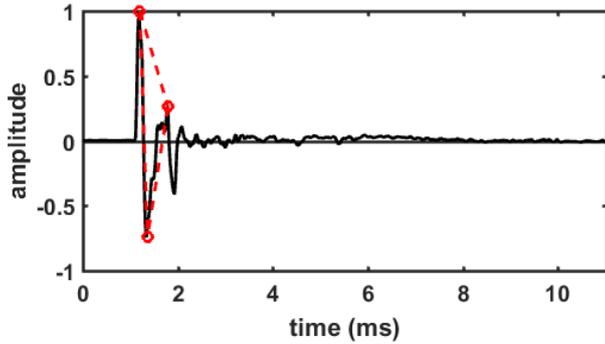


Fig. 8. Feature illustration: area between 3 extremes (Area)

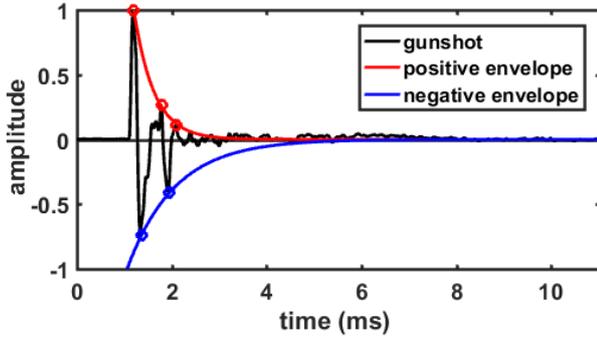


Fig. 9. Feature illustration: envelope approximation by exponential fit (An, Ap; Bn, Bp)

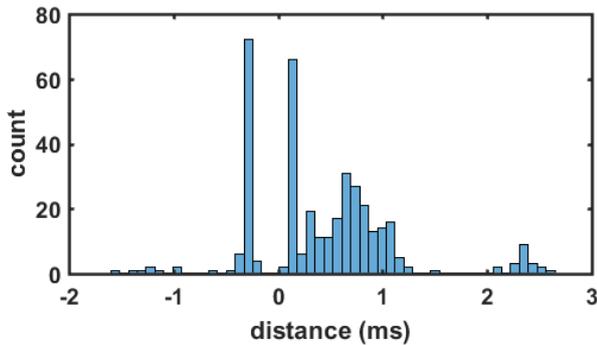


Fig. 10. Distance between maximum and minimum points (PDist)

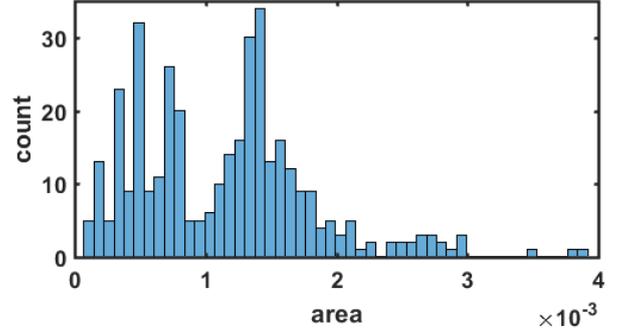


Fig. 11. Area between points max-min-max (Area)

Table 2 below shows mean values and standard deviation of individual features extracted from AK-47 recordings as well as values extracted from all gunshots. Mean value is denoted by μ , standard deviation by σ and an overall measure $|\mu|/\sigma$ is introduced (calculated as a ratio between absolute mean value and standard deviation).

TABLE II. TABLE 1 MEAN AND STD. DEVIATION OF EXAMINED FEATURES

Feature	AK-47			All gunshots		
	μ	σ	$\frac{ \mu }{\sigma}$	μ	σ	$\frac{ \mu }{\sigma}$
RP-	-0.13	0.13	1	-0.13	0.13	1
RP+	0.24	0.12	2	0.21	0.59	0.36
ZDist	0.37	0.17	2.18	0.35	0.61	0.57
PDist	0.43	0.68	0.63	0.32	1.37	0.23
PDist	0.61	0.52	1.17	0.82	1.14	0.72
Ang	89.99	0.03	2999	89.99	0.05	1800
Area	0.001	0.001	1	0.001	0.001	1
An	-6.52	98	0.07	-3.63	65.82	0.06
Ap	8.65	65	0.13	4.75	42.99	0.11
Bn	-465	517	0.9	-320	361	0.89
Bp	-737	801	0.92	-531	648	0.82

IV. RESULTS

To compare performance of proposed features, they were extracted from both gunshot categories (AK-47-only and all weapons) and non-gunshot signals. Subsequently, they were fed to a neural network for recognition. This paper uses default Matlab implementation of a neural network with one hidden layer with 10 neurons. Since individual features by themselves (dimensionality 1), or even a combination of two features were not able to recognize gunshots, we resorted to compare recognition performance when increasing the number of features. In order to achieve the best results with the least possible amount of features, we considered ordering them according to several metrics. Firstly, features were ordered from expected best to expected worst according to the μ/σ measure presented in Tab. 2 (assuming: the greater the better). The second ordering was dependent on mutual information between feature values and class labels. Mutual information was calculated using the Matlab function `kernelmi` [13] and greater mutual information meant the feature was supposed to perform

better at recognition (we disregarded effects of mutual information between individual features in ordering for now). The last criterion was based on two sample t-tests, as calculated by the Matlab function `ttest2`, measuring similarity of distributions of two datasets. Ordering was done according to the p-value of the function which reflects confidence in the null hypothesis, i.e. that samples are from one distribution. The compared distributions were those of features from gunshot categories against those of features from non-gunshot categories. From that follows that the lesser the p-value, the more dissimilar the sets should be and more convenient for recognition. Although all criteria produced similar results, better results were obtained after slight manual rearranging of proposed orders. The final order we settled on is mentioned in the next paragraph, just below the figures. Features that are not mentioned are deemed unfit for recognition. Fig. 12 shows performance (recall (2) and precision (3)) of different numbers of features for all gunshots; the subsequent figure, Fig. 13, only represents AK-47 assault rifle recognition (whose features were sorted in the same manner as “all gunshots”). For better illustration, Fig. 14 shows the ROC curve for the 7 best features when using all gunshots (which achieved the best results in terms of F-score).

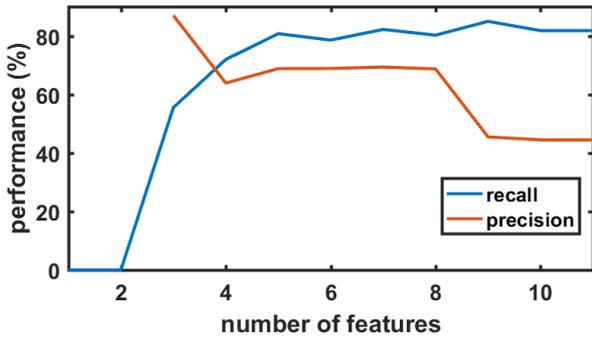


Fig. 12. Recall and precision for all gunshots and different number of features

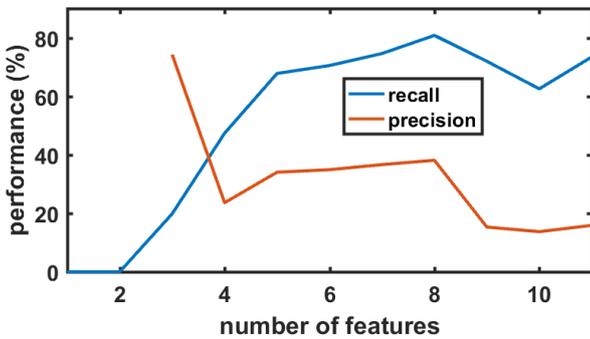


Fig. 13. Recall and precision for AK-47 gunshots and different number of features

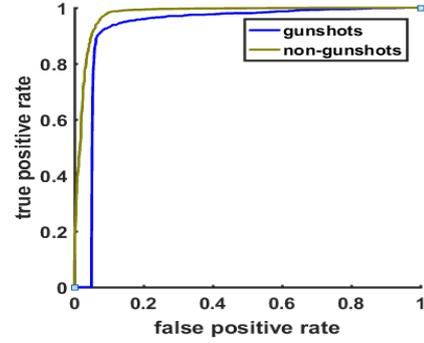


Fig. 14. ROC curve for all gunshots and 7 features

Achieved results indicate, that although other features (such as LPC or MFCC) might be more effective, there are still many possibilities for feature extraction in time. We conclude that the first 7 features (8 in the case of AK-only) might be useful for gunshot, and possibly other impulsive signal recognition. These features are: RP- and RP+ (zero crossing before and after the most dominant peak), Bn and Bp coefficients of envelope approximation, PIDist (distance between maximum and minimum), angle between horizontal line and line connecting maximum and minimum, and possibly distance between zero-crossings (ZDist). The results achieved using the first 7 features (8 in the case of AK-only) are summarized in Table 3. Formulae for the metrics used in Tab. 3 are the following:

$$recall = \frac{tp}{tp+fn} \quad (2)$$

$$precision = \frac{tp}{tp+fp} \quad (3)$$

$$F - score = 2 \cdot \frac{precision \cdot recall}{precision+recall} \quad (4)$$

Where tp are true positives (gunshots classified as gunshots), fn are false negatives (misclassified gunshots) and fp are false positives (non-gunshots flagged as gunshots).

TABLE III. BEST RECOGNITION RESULTS

Subset	Recall (2)	Precision (3)	F-Score (4)
AK-47	80.8 %	38.1 %	51.8
All gunshots	82.2 %	69.3 %	75.2

A. Comparison with other works

This chapter presents the results achieved by other, similar works. It is important to note that the aims of these papers varies, some focus on gunshot detection, others on specific weapon recognition and yet others have multiple objectives such as general (and/or multiple) event recognition/classification and scene classification. When there are multiple objectives, the one with the closest conditions and best results was chosen. Tab. 4 presents results achieved by previously mentioned papers. TPR means “True Positive Rate” (i.e. Recall), FPR “False Positive Rate”, ACC stands for Accuracy.

TABLE IV. RESULTS OF OTHER WORKS

Paper	Dataset	Feature Dimensionality	Results
[1]	332 gunshots / 102 non-gunshots	9	99% TPR 0% FPR
[3]	320 samples in 16 categories	63	60% ACC
[3]	320 samples in 16 categories	26	56% ACC
[4]	372 gunshots (4 handguns)	30	87% ACC 88% TPR
[5]	4000 sounds in 50 categories	SIF resolution: 52x40	97% ACC

Most papers are performance oriented, and thus frequently use high feature dimensionality. The focus of this paper is to explore the usability of new features, so it does not use optimal combinations of existing features. This explains the comparatively worse results in performance, but also testifies to the fact that some of the features presented have the potential to boost recognition performance when used in combination with other conventional features. Further research into their compatibility with conventional features is required, but their different nature (as they are not derived from a spectrum) might hint to low mutual information between the two groups.

ACKNOWLEDGMENT

Research described in this paper was financed by the Ministry of Education, Youth and Sports within the frame of the National Sustainability Program under grant LO1401. For research, infrastructure of the SIX Center was used. The research was also financially supported by the Brno University of Technology Internal Grant Agency under project no. FEKT-S-17-4707.

REFERENCES

- [1] T. Ahmed, M. Uppal, and A. Muhammad, "Improving efficiency and reliability of gunshot detection systems," in 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Canada, 2013, pp. 513–517.
- [2] J. L. Rouas, J. Louradour, and S. Ambellouis, "Audio Events Detection in Public Transport Vehicle," in 2006 IEEE Intelligent Transportation Systems Conference, Canada, 2006, pp. 733–738.
- [3] G. Roma, P. Herrera, and W. Nogueira, "Environmental sound recognition using short-time feature aggregation," *Journal of Intelligent Information Systems*, pp. 1–19, Aug. 2017.
- [4] E. Kiktova, M. Lojka, M. Pleva, J. Juhar, and A. Cizmar, "Gun type recognition from gunshot audio recordings," in 3rd International Workshop on Biometrics and Forensics (IWBF 2015), Norway, 2015, pp. 1–6.
- [5] H. Zhang, I. McLoughlin, and Y. Song, "Robust sound event recognition using convolutional neural networks," in 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Australia, 2015, pp. 559–563.
- [6] T. F. Quatieri, "Discrete-Time Speech Signal Processing: Principles and Practice". London: Prentice Hall, 2001
- [7] H. Hermansky, "Perceptual linear predictive (PLP) analysis of speech", *Journal of the Acoustical Society of America*, vol. 87, pp. 1738-1752, April 1990.
- [8] B. Jaszczak and B. Nelson, (2014). Free Firearm Sound Effects Library. [online]. Available at: <http://www.airbornesound.com/downloads/free-firearm-sound-effects-library-expanded-edition> [Accessed january 2018].
- [9] J. Salamon, C. Jacoby, and J. P. Bello, "A dataset and taxonomy for urban sound research", in Proc. of 22nd ACM International Conference on Multimedia, USA, 2014, pp. 1041-1044.
- [10] Airborne Sound - downloadable sound effects collections [online]. Available at: <http://www.airbornesound.com/> [Accessed january 2018].
- [11] Freesound [online]. Available at: <http://www.freesound.com/> [Accessed january 2018].
- [12] M. Hrabina and M. Sigmund, "Analysis of Linear Predictive Coefficients for Gunshot Detection Based on Neural Networks", in Proc. of the 2017 IEEE International Symposium on Industrial Electronics, United Kingdom, 2017, pp. 1961-1965
- [13] Kernel estimate for (Conditional) Mutual Information. Matlab File Exchange [online]. Available at: <https://www.mathworks.com/matlabcentral/fileexchange/30998-kernel-estimate-for--conditional--mutual-information?focused=5183650&tab=function> [Accessed march 2018].