Plénière - 3 ans de thèse

Manon Mottier

Avantix - CentraleSupélec

February 10th, 2022

Content

Content

Overview

- Administrative
  - 3rd year thesis
  - Full time at the L2S
- Communication
  - Publication of a Conference paper: "Deinterleaving and Clustering unknown RADAR pulses" - Manon Mottier; Gilles Chardon; Frédéric Pascal
  - International Conference: IEEE RadarConf 2021, Atlanta, GA, USA
  - International Seminar: Atos Scientific Community

Content

Data description

- Work from simulated labeled data
- Each dataset is composed of RADAR pulses, described by only 4 features:
  - Frequency (F)
  - Pulse Width (PW)
  - Level (L)
  - Time of Arrival (TOA)

Data description

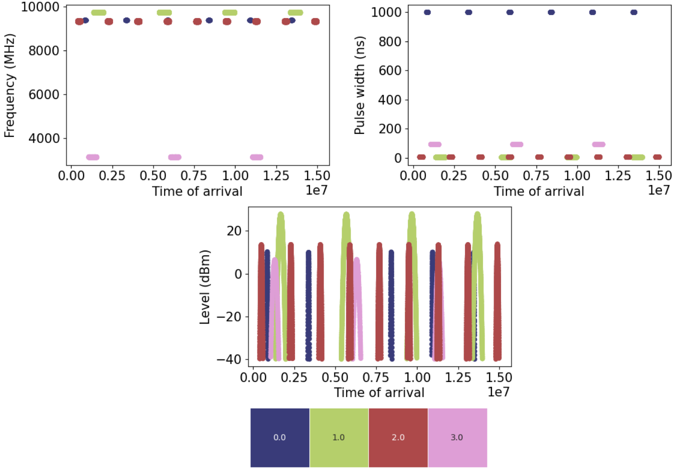


Figure: Example of a **simple** signal containing 4 RADARs.

Methodology

1. **RADAR pulses deinterlacing:**
   - **Clustering**: Cluster each observation of this signal under the constraint of not merging different radars in one cluster
   - **Cluster fusion**: Hierarchical agglomerative clustering combined with optimal transport distances
2. **RADAR Identification**: Extraction of group characteristics and comparisons to a reference database to identify the transmitter

Deinterlacing RADAR pulses

Clustering with HDBSCAN

Methodology:

- Application of clustering algorithm on two features: Frequency and Pulse Width
- Why HDBSCAN ?
  - Unsupervised algorithm that allows us to search for clusters of varying densities
  - Few parameters to optimize
- Over-estimation of the number of clusters
  - Allows to receive RADARs that do not transmit continuously
  - Each cluster must not contain the observations of different RADARs

Deinterlacing RADAR pulses

Output of HDBSCAN

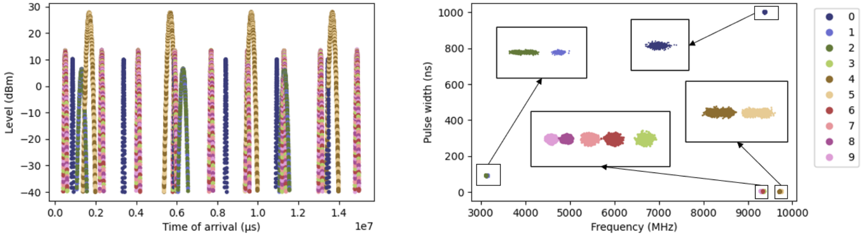


Figure: HDBSCAN Output. Colors represent the different clusters.

Deinterlacing RADAR pulses

Cluster fusion

Methodology:

1. Dimensionality reduction: each cluster is represented by a histogram of its TOA and LEVEL
2. Compute distances between cluster histograms using **optimal transport**
3. Aggregation of the 2 closest clusters i.e., the 2 clusters with the lowest transport costs
4. Computation of new costs between old clusters and the new ones
5. Repeat steps 2 to 4 until a single cluster is obtained

Deinterlacing RADAR pulses
Histograms

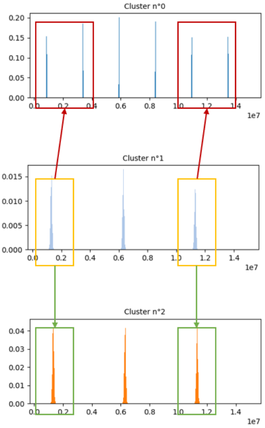


Figure: Histograms of the previous clusters.

Deinterlacing RADAR pulses

Optimal Transport explanation

- Transporting distribution to another has a cost:

$$C(\mathbf{P}) = \sum_{n=1}^{N} \sum_{m=1}^{M} C_{nm} P_{nm} = \langle \mathbf{C}, \mathbf{P} \rangle \qquad (1)$$

$C(\mathbf{P}) =$ Total cost

$C_{nm} =$ cost transporting unit mass from x_n to y_m (transport cost)

$P_{nm} =$quantity of mass taken from x_n to y_m (transport plan)

- The optimal transport plan is a simple minimization problem:

$$\mathbf{P}^\star = \underset{\mathbf{P} \in \mathbf{R}_+^{N \times M}}{\operatorname{argmin}} \langle \mathbf{C}, \mathbf{P} \rangle \text{ such that } \mathbf{P}\mathbf{1}_M = \mathbf{a}, \mathbf{P}^T \mathbf{1}_N = \mathbf{b}^T. \qquad (2)$$

Deinterlacing RADAR pulses

Dendrogram

- Decision model based on several unsupervised metrics to cut the dendrogram:
  - Silhouette score
  - Calinski-Harabasz score
  - Davies-Bouldin score

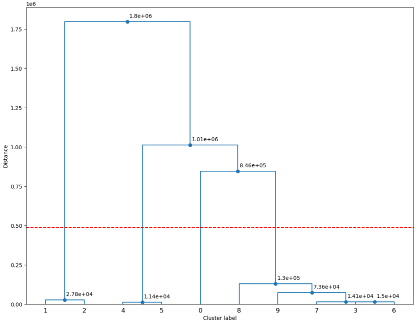
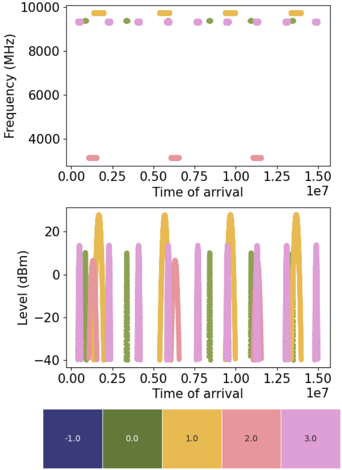


Figure: Cluster aggregations with Optimal Transport Distances.

Deinterlacing RADAR pulses

Final Grouping

- Perfect grouping
- Statistics:
  - Homogeneity = 1
  - Completeness = 1
  - Unclassified rate = 2%



RADAR Identification

Waveform analysis

Methodology:

1. Construction of a measure that describes each RADAR belonging to a reference database based on its frequencies and pulse width:

$$\mu_j = \sum_{n=1}^{N} \alpha_n \delta_{f_n, pw_n} \tag{3}$$

$N =$ the number of frequencies and pulse width on which the RADAR transmits
$\alpha =$ the weight of the frequency and pulse width
$j =$ the index of the transmitter in the database
$\delta =$ dirac mass (with $\sum \alpha_n = 1$)

RADAR Identification

Waveforms representation: 1D (Frequency)

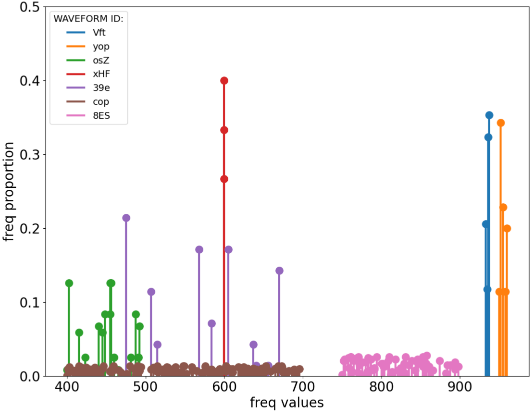


Figure: Referential Database representation with some **simulated** emitters from Frequency.

RADAR Identification

Waveforms representation: 2D (Frequency x Pulse Width)

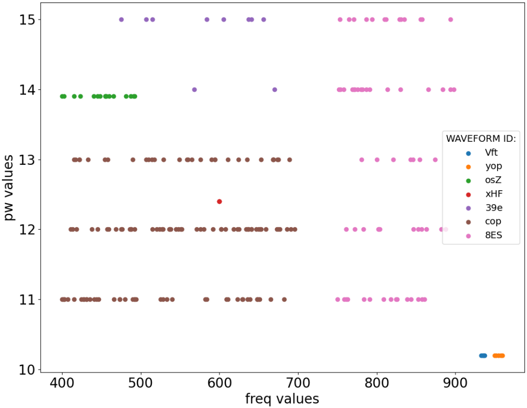


Figure: Referential Database representation with some **simulated** emitters from Frequency and Pulse Width.

RADAR Identification

Waveforms representation: 3D (PRI x Frequency x Pulse Width)

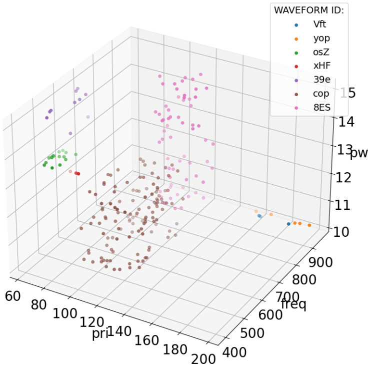
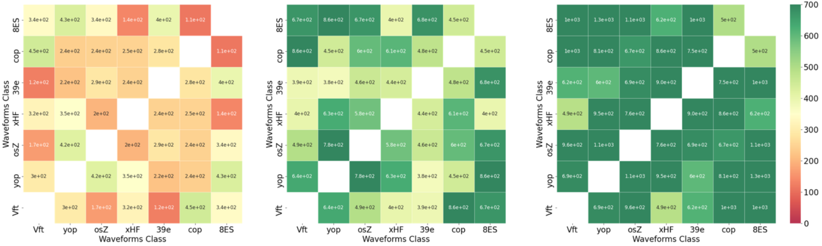


Figure: Referential Database representation with some **simulated** emitters from Pri, Frequency and Pulse Width.

RADAR Identification

Cost matrix between all waveforms

Distance between waveforms is defined as the optimal transport cost between their distributions: $d\left(\mu_i, \mu_j\right)$



Freq

Freq - Pw

PRI – Freq - Pw

Figure: Cost matrix between some simulated waveforms. Green indicates high transport cost which means that the waveforms are very different and red indicates low transport cost means that the waveforms have similar characteristics.

RADAR Identification

Groups analysis

2. Construction of a measure that describes each cluster from the frequency and the pulse width:

$$\gamma = \frac{1}{M} \sum_{m=1}^{M} \delta_{f_m, pw_m} \tag{4}$$

$M =$ the number of pulses in the cluster

3. Identification of the transmitter:

$$j^\star = \operatorname{argmin} d\left(\mu_j, \gamma\right) \tag{5}$$

RADAR Identification

Cost matrix

- The rows represent the number of groups and the columns of the waveforms class.
- Green indicates high transport cost, which means that the waveforms are very different, and red indicates low transport cost means that the waveforms have similar characteristics.

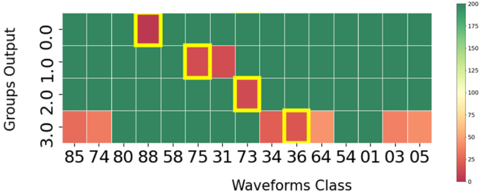


Figure: Cost matrix extract between groups and waveforms built from Frequency and Pulse Width.

RADAR Identification

Classification output in 2 dimensions: Group 0 & 1

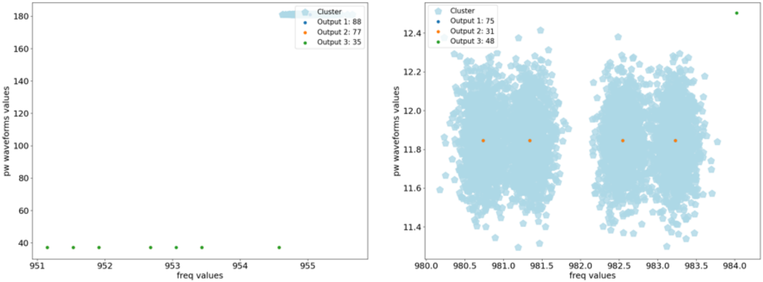


Figure: Classification results perform in 2 dimensions from the frequency and pulse width on group 0 (on the left) and group 3 (on the right). The plots show the three first outputs from the classification. The transmitters 77 and 35 for group 0 are indissociable because they are identical in Frequency and Pulse Width, so their distributions overlapped-the same behavior for transmitters 31 and 75 for group 1.

RADAR Identification

Classification output in 2 dimensions: Group 2 & 3

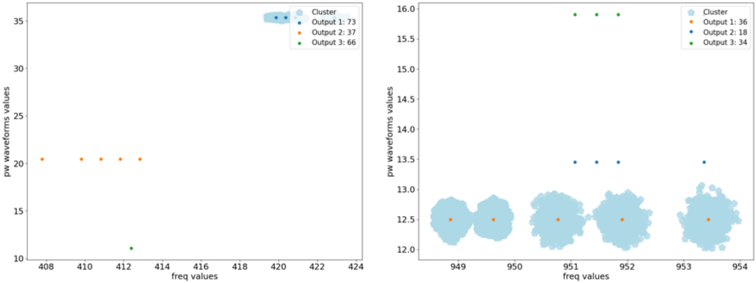


Figure: Classification results perform in 2 dimensions from the frequency and pulse width on group 2 (on the left) and group 3 (on the right). The plots show the three first outputs from the classification.

RADAR Identification

Global results

Table: Global results on simulated data

Correct classification position	Number of group tested	Percent
First	1 982	72.6%
Second	318	11.7%
Third	110	4%
More	318	11.7%
Total	2 728	.

- Correct identification
- Problem with mono frequency RADAR
- Difficulty to identify sub-modes
- Frequency and Pulse width not sufficient

Content

Application case

Please take the pdf "Annexes"

Content

Conclusion

- **Next publication/ communication**
  - EUSIPCO *(European Signal Procecssing Conference)*
    - Belgrade - 22 Aug / 2 Sep, 2022
    - Paper submission deadline: 20 Feb, 2022
  - GRETSI *(Colloque Francophone de Traitement du Signal et des Images)*
    - Nancy - 5 Sep / 9 Sep, 2022
    - Paper submission deadline: Feb, 2022
  - Article in a scientific journal
- **Methodology**
  - Improve Deinterlacing chain
    - Improve Differentiate identical RADARs in Frequency / Pulse Width
    - Get a better model decision to cut the dendogram
  - Add the PRI for the classification
  - Novelty detection

Thank you for your attention

For more information please contact:

Manon MOTTIER

manon.mottier@atos.net