

Désentrelacement et classification de signaux RADAR basés sur des distances de transport optimal

Manon MOTTIER, Gilles CHARDON, Frédéric PASCAL

Université Paris-Saclay, CNRS, CentraleSupélec
Laboratoire des signaux et systèmes
91190, Gif-sur-Yvette, France

manon.mottier@centralesupelec.fr, gilles.chardon@centralesupelec.fr
frederic.pascal@centralesupelec.fr

Résumé – L’identification d’émetteurs RADAR représente un enjeu primordial dans le domaine de la guerre électronique d’autant plus que ces émetteurs présentent des caractéristiques de plus en plus complexes. Ce papier présente une méthodologie de désentrelacement et d’identification d’émetteurs RADAR basée sur l’utilisation de distances de transport optimal pour désentrelacer les impulsions d’un nombre inconnu d’émetteurs, puis identifier ces émetteurs grâce à une base de données. La méthodologie est testée sur des données obtenues grâce à un simulateur de données.

Abstract – Identifying RADAR transmitters represents a key challenge in electronic warfare, especially since RADARs have increasingly complex characteristics. This paper presents a methodology for deinterleaving and identifying RADAR transmitters based on the use of optimal transport distances to deinterleave pulses from an unknown number of transmitters and then identify these RADARs using a database. The methodology is tested on data obtained using a data simulator.

1 Introduction

Les récentes évolutions technologiques ont permis de nombreuses innovations dans le domaine de la guerre électronique. La modernisation des équipements est un véritable défi ; les émetteurs sont plus sophistiqués et ont des spectres électromagnétiques plus similaires rendant leur différenciation complexe et devenant ainsi un enjeu crucial dans la prise de décision et l’engagement militaire. L’identification des émetteurs RADARs est primordiale car elle permet d’apporter une réponse stratégique et opérationnelle adaptée à la prise de position ennemie et ainsi éviter des potentiels conflits et menaces.

Le processus d’identification d’émetteurs RADAR est principalement divisé en 2 étapes ; la première étape consiste à désentrelacer un signal regroupant des impulsions mélangées d’un nombre inconnu d’émetteurs. Plusieurs méthodes de désentrelacement basées sur des modèles d’apprentissages profonds ou de clustering ont été construites pour séparer et regrouper les impulsions en groupes [4, 6–8]. Puis, à partir de ces résultats, un classificateur est construit pour comparer les ensembles d’impulsions à une base de données de référence pour d’identifier les émetteurs RADARs présents. L’identification RADAR est une étape sensible, car elle se base sur les résultats du désentrelacement d’un signal pour extraire les caractéristiques des groupes, les comparer à une base de données de référence et identifier l’émetteur. La plupart des méthodes sont basées sur l’application de l’apprentissage profond et gèrent peu de classes [5, 9]. Des comparaisons ont montré que les méthodes d’apprentissage profond ne sont pas néces-

sairement meilleures que des modèles plus conventionnels et plus simples.

Cet article présente une méthodologie simple basée sur l’utilisation de distances de transport optimal pour désentrelacer puis classifier les impulsions d’un signal RADAR afin d’identifier les émetteurs présents dans un signal. La méthodologie a été testée sur des données construites à partir d’un simulateur. Les données sont collectées par un récepteur qui écoute l’environnement électromagnétique sur une large bande de fréquences. Les impulsions sont ensuite segmentées, analysées et décrites par 4 caractéristiques :

- Fréquence, (f_n)
- Durée d’impulsion, PW (pw_n)
- Niveau, G (g_n)
- Temps d’arrivée, TOA (t_n)

La Fig. 1 représente un signal simulé rassemblant près de 6 000 impulsions provenant de 4 émetteurs différents. Chaque point représente une impulsion. Les paramètres f_n , pw_n et g_n sont tracés en fonction du temps d’arrivée t_n . Le plan (f_n, t_n) met en évidence les impulsions d’un RADAR émettant sur 3 bandes de fréquences différentes entre 945 et 956 MHz : un RADAR peut être représenté par de multiples fréquences et durée d’impulsion. L’étalement de la durée d’impulsion dans le plan (pw_n, t_n) est dû à une erreur de mesure des capteurs ; la valeur est sous-estimée compliquant le désentrelacement et particulièrement la classification. Les lobes correspondants aux différents RADARs peuvent être mélangés et se superposer comme le montre le plan (g_n, t_n) .

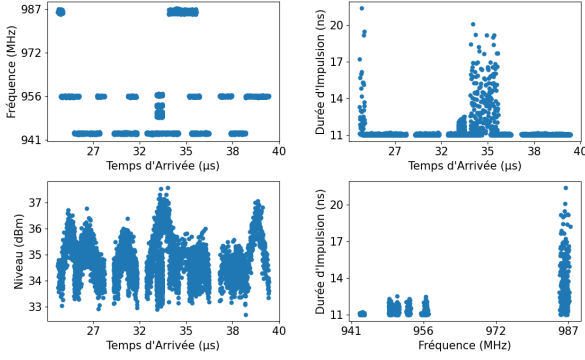


FIGURE 1 – Exemple d’un signal simulé regroupant les impulsions de 4 émetteurs différents.

L’article est organisé de la façon suivante ; les sections 2 et 3 décrivent respectivement la stratégie de désentrelacement d’un signal puis la méthode de classification proposée pour identifier les émetteurs RADAR. La partie 4 présente la conclusion et les perspectives de recherches. La méthodologie est illustrée en utilisant les données de la Fig. 1.

2 Désentrelacement d’un signal RADAR

La méthode de désentrelacement proposée se fait en deux étapes ; premièrement un algorithme de clustering est appliqué dans le plan (f_n, pw_n) afin de séparer les impulsions puis, à partir de ces clusters, un clustering agglomératif hiérarchique utilisant des distances de transport optimal est appliqué dans le plan (g_n, t_n) pour regrouper les clusters ; le bruit et la qualité d’estimation des variables impactent significativement les résultats du clustering en surestimant le nombre de clusters estimé. Le regroupement des clusters est une étape d’autant plus nécessaire qu’un RADAR peut également émettre sur de multiples bandes de fréquences et/ou de durée d’impulsion et donc être représentés par plusieurs clusters dans le plan (f_n, pw_n) comme le montre la Fig 1.

2.1 Séparation des impulsions avec HDBSCAN

L’algorithme HDBSCAN est appliqué dans le plan (f_n, pw_n) afin de séparer les impulsions. HDBSCAN est un algorithme de clustering non supervisé capable de détecter des clusters de densités différentes [2] ainsi que les valeurs aberrantes (par exemple de fausses alarmes lors de la détection d’impulsions). HDBSCAN a été paramétré pour surestimer le nombre de clusters retourné pour éventuellement capturer les RADARs ayant peu d’impulsions ou émettant très peu dans le temps ; sous-estimer le nombre de clusters pourrait conduire à regrouper les impulsions de plusieurs RADARs.

2.2 Agglomération des clusters

Comme précédemment expliqué, un émetteur peut être représenté par plusieurs clusters. Nous avons appliqué un clustering agglomératif hiérarchique utilisant des distances de transport optimal pour regrouper les clusters appartenant à un même émetteur RADAR. La méthode proposée est basée sur le développement d’une distance entre les clusters [1, 10] pour mesurer leurs similarités et dissimilarités. Nous montrerons par la suite que des distances de transport optimal sont tout particulièrement adaptées pour effectuer ces regroupements.

Dans cet article, nous nous concentrons sur la partie de cette théorie traitant des distributions de probabilités discrètes. Dans le cas de deux distributions de probabilité discrète $\nu = \sum_{n=1}^N a_n \delta_{x_n}$ et $\mu = \sum_{m=1}^M b_m \delta_{y_m}$, avec $\mathbf{a} = (a_1, \dots, a_N)^T \in \mathbf{R}_+^N$, $\sum_{n=1}^N a_n = 1$, et $\mathbf{b} = (b_1, \dots, b_M)^T \in \mathbf{R}_+^M$, $\sum_{m=1}^M b_m = 1$. Un plan de transport \mathbf{P} entre ν et μ est défini par ses coefficients P_{nm} , représentant la quantité de masse prise de x_n à y_m . Avec $c(\cdot, \cdot)$ une fonction de coût (nous prendrons ici $c(x, y) = \|x - y\|_2$), et $C_{nm} = c(x_n, y_m)$ le coût de transport d’une unité de masse de x_n à y_m , le coût total $C(\mathbf{P})$ d’un plan de transport est

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

La cohérence du plan de transport \mathbf{P} avec ν et μ est garantie par $\mathbf{P} \mathbf{1}_M = \mathbf{a}$, $\mathbf{P}^T \mathbf{1}_N = \mathbf{b}$. Le plan de transport optimal \mathbf{P}^* est défini comme le minimiseur de Eq. (1) sous les contraintes suivantes :

$$\mathbf{P}^* = \underset{\mathbf{P} \in \mathbf{R}_+^{N \times M}}{\operatorname{argmin}} C(\mathbf{P}) \text{ subject to } \mathbf{P} \mathbf{1}_M = \mathbf{a}, \mathbf{P}^T \mathbf{1}_N = \mathbf{b} \quad (2)$$

La distance de transport optimal entre ν et μ est alors définie par $d(\nu, \mu) = C(\mathbf{P}^*)$.

Nous faisons l’hypothèse que deux clusters appartenant au même RADAR sont actifs durant les mêmes intervalles de temps ; un RADAR pouvant émettre selon différents modes sera considéré comme étant deux émetteurs distincts ou plus. Chaque cluster est représenté par une mesure de probabilité décrivant sa distribution à partir du temps d’arrivée et du niveau :

$$\tau = \frac{1}{M} \sum_{n=1}^M \delta_{g_n, t_n} \quad (3)$$

Par conséquent, deux clusters du plan (f_n, pw_n) d’un même émetteur auront une distance entre leurs distributions très faible. Ces clusters sont fusionnés de façon hiérarchique, en agglomérant itérativement les clusters avec la plus petite distance de transport optimal dans le plan (g_n, t_n) . Après la fusion, la distance entre les clusters fusionnés et celle des autres clusters est mise à jour. Le processus est réitéré puis interrompu lorsqu’un critère d’arrêt est atteint.

Fig. 2 montre le résultat final de l’agglomération. Le modèle décisionnel identifie correctement la présence de 4 émetteurs distincts ainsi qu’une classe d’outliers (-1). Les couleurs représentent les différents ensembles d’impulsions.

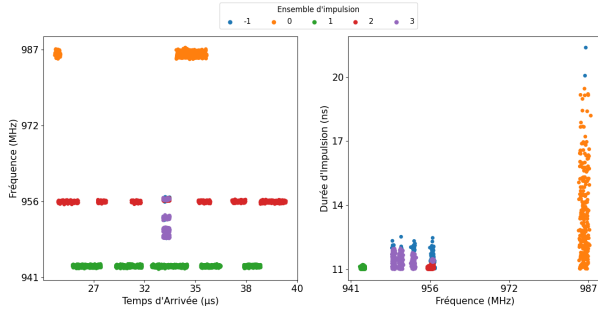


FIGURE 2 – Regroupements finaux. Chaque couleur représente un ensemble d’impulsion différent.

3 Classification des RADARs

Lorsque que le signal est désentrelacé, la phase de classification peut débiter pour identifier les différents RADARs présents dans le signal. La classification est basée sur l’élaboration d’une distance entre un ensemble d’impulsions RADARs reçues, et une description des caractéristiques d’un émetteur RADAR à partir d’une base de données de référence. La classification est faite en identifiant les émetteurs RADARs les plus proches des données reçues. Ici, les données et les classes reçues sont représentées sous forme de distributions de probabilité. Dans le cas d’émetteurs avec des caractéristiques simples (par exemple, une seule fréquence), des distances simples telles que la distance euclidienne entre les moyennes des caractéristiques des impulsions peuvent être utilisées. Cependant, de telles distances ne peuvent pas traiter des émetteurs plus complexes et agiles, pour lesquels leurs moyennes ne peuvent pas simplement décrire la distribution des caractéristiques. Il existe plusieurs façons de définir une distance entre des distributions plus générales. Par exemple, la divergence de Kullback-Leibler où la distance de variation totale sont fréquemment utilisées. Néanmoins, ces distances ou divergences ne peuvent pas être utilisées dans notre cas, car les distributions représentant les données sont discrètes et, en général, auront des supports disjoints. Nous montrerons dans la suite que des distances de transport optimales proposées sont bien adaptées au problème posé, car elles peuvent traiter des distributions représentant des émetteurs agiles et sont robustes au bruit.

3.1 Représentation des RADARs typiques

Notre base de données RADAR initiale comprend plus de 60 émetteurs différents, et certains ont des caractéristiques très similaires tandis que d’autres se distinguent facilement. Nous avons choisi de ne travailler qu’avec des caractéristiques spécifiques; la fréquence f_n et la durée d’impulsion w_n car elles sont très significatives et fiables. À partir de cette base de données, nous construisons une mesure μ_j décrivant chaque émetteur appartenant à la base de données de référence par :

$$\mu_j = \sum_{n=1}^N \alpha_n \delta_{f_n, p w_n} \quad (4)$$

avec N le nombre de fréquences et de durée d’impulsions sur lesquelles le RADAR émet, α la proportion de fréquence et de durée d’impulsions et δ , la masse de Dirac (avec $\sum \alpha_n = 1$).

3.2 Identification des RADARs

Afin de calculer la distance entre un ensemble d’impulsions et une classe RADAR, les impulsions reçues sont également modélisées sous forme de distribution de probabilité :

$$\nu = \frac{1}{M} \sum_{m=1}^M \delta_{f_m, p w_m} \quad (5)$$

avec M le nombre d’impulsions dans l’ensemble. Une classe est ensuite attribuée à l’ensemble d’impulsions en identifiant la classe RADAR la plus proche au sens de la distance de transport optimal :

$$j^* = \operatorname{argmin}_{j \in \{1, \dots, J\}} d(\mu_j, \nu). \quad (6)$$

avec J le nombre de classes d’émetteurs. Pour diminuer la complexité de calcul des distances, les impulsions reçues sont regroupées en intervalles de fréquences et de durée d’impulsion. Avec des tailles d’intervalles suffisamment petites, la perturbation de la distance est suffisamment faible pour que l’ordre de grandeur des distances entre les données et les classes soit conservé. La Fig. 3 montre le résultat de notre méthode de classification appliquée à l’ensemble d’impulsion 0 de la Fig. 2. Le graphique de gauche superpose les impulsions et les trois classes d’émetteurs les plus proches. Les points caractéristiques de la classe la plus proche rendent bien compte des données par rapport aux autres classes; en effet les classes 2 et 3 représentent des RADARs de plus basses fréquences. L’étalement de la durée d’impulsion est caractéristique d’une erreur de mesure. Malgré cette erreur, le classificateur identifie correctement l’émetteur présent dans les données. Le graphique de droite montre les plans de transport entre la distribution des données et les trois classes [3]. La classe la plus proche représente un émetteur multifréquence. Les points de données sont envoyés sur les différents points caractéristiques en respectant les proportions; c’est pourquoi les impulsions autour d’une fréquence donnée ne sont pas toutes envoyées au même point. Les deux autres classes représentent des émetteurs à fréquence unique; tous les points de données sont tous envoyés au même emplacement.

4 Conclusion et Perspectives

4.1 Conclusion

Nous avons appliqué une stratégie en 2 étapes pour désentrelacer et classifier les impulsions d’un signal RADAR. Tout

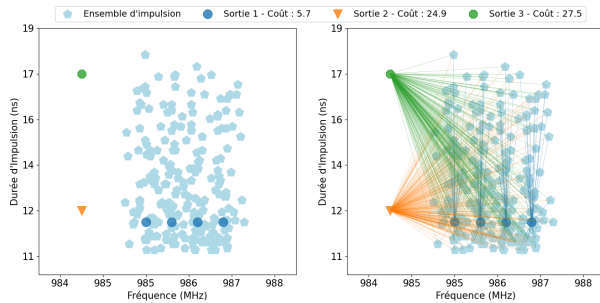


FIGURE 3 – Résultat de la classification de l'ensemble d'impulsion 0.

d'abord, un algorithme de clustering a été appliqué sur la fréquence et la durée d'impulsion pour séparer chaque impulsion du signal en plusieurs clusters avant de les regrouper grâce à un clustering ascendant hiérarchique combiné aux distances de transport optimal. Puis, à partir de ces clusters fusionnés, nous avons extrait des caractéristiques pour les comparer à une base de données de référence en utilisant la distance du transport optimal. Les résultats obtenus sur les données simulées sont très encourageants et permettent d'identifier avec grande confiance la classe de l'émetteur. La méthodologie peut gérer plus de 60 classes à identifier et peut facilement être agrémentée de nouvelles classes.

4.2 Perspectives

Lors du désentrelacement, nous supposons que les clusters retournés par HDBSCAN contiennent les observations d'un seul émetteur, mais il est possible que les RADARs puissent avoir les mêmes caractéristiques et être confondus en fréquence et en durée d'impulsion. C'est le cas dans les ports ou les aéroports où plusieurs modèles similaires sont utilisés. Afin d'améliorer les résultats de la classification, nous travaillons actuellement sur deux axes principaux : Premièrement, ajouter une troisième dimension dans le transport optimal pour mieux discriminer les RADARs. Comme expliqué précédemment, certains RADARs peuvent avoir des caractéristiques très similaires en fréquence et en durée d'impulsion. Il est donc nécessaire d'ajouter d'autres caractéristiques pour les différencier ; par exemple, avec la période de répétition des impulsions (PRI), qui est la différence du temps d'arrivée entre des impulsions successives ($\delta_n = t_n - t_{n-1}$). La modélisation des différents types de PRI (constante, aléatoire...) permettra d'extraire le motif d'émission du RADAR et ainsi mieux le discriminer. Néanmoins, tout comme la durée d'impulsion, la PRI ne peut pas être directement intégrée et nécessite un prétraitement en raison d'erreurs de mesure et d'impulsions non détectées. Enfin, pour proposer une méthode de classification complète, nous développerons sur une méthodologie capable de détecter de nouveaux émetteurs non référencés dans la base de référence pour l'enrichir automatiquement.

Remerciements

ATOS a soutenu ce travail en fournissant les données et son expertise RADAR.

Références

- [1] Nicolas Bonneel, Michiel Van De Panne, Sylvain Paris, and Wolfgang Heidrich. Displacement interpolation using lagrangian mass transport. In *Proceedings of the 2011 SIGGRAPH Asia conference*, pages 1–12, 2011.
- [2] Ricardo JGB Campello, Davoud Moulavi, and Jörg Sander. Density-based clustering based on hierarchical density estimates. In *Pacific-Asia conference on knowledge discovery and data mining*, pages 160–172. Springer, 2013.
- [3] Rémi Flamary, Nicolas Courty, Alexandre Gramfort, Mokhtar Z. Alaya, Aurélie Boisbunon, Stanislas Chambon, Laetitia Chapel, Adrien Corenflos, Kilian Fatras, Nemo Fournier, Léo Gautheron, Nathalie T.H. Gayraud, Hicham Janati, Alain Rakotomamonjy, Ievgen Redko, Antoine Rolet, Antony Schutz, Vivien Seguy, Danica J. Sutherland, Romain Tavenard, Alexander Tong, and Titouan Vayer. Pot : Python optimal transport. *Journal of Machine Learning Research*, 22(78) :1–8, 2021.
- [4] Xueqiong Li, Zhangmeng Liu, and Zhitao Huang. Deinterleaving of pulse streams with denoising autoencoders. *IEEE Transactions on Aerospace and Electronic Systems*, 56(6) :4767–4778, 2020.
- [5] Zhang-Meng Liu and S Yu Philip. Classification, denoising, and deinterleaving of pulse streams with recurrent neural networks. *IEEE Transactions on Aerospace and Electronic Systems*, 55(4) :1624–1639, 2018.
- [6] HK Mardia. New techniques for the deinterleaving of repetitive sequences. In *IEE Proceedings F (Radar and Signal Processing)*, volume 136, pages 149–154. IET, 1989.
- [7] DJ Milojević and BM Popović. Improved algorithm for the deinterleaving of radar pulses. In *IEE Proceedings F (Radar and Signal Processing)*, volume 139, pages 98–104. IET, 1992.
- [8] Manon Mottier, Gilles Chardon, and Frédéric Pascal. Deinterleaving and clustering unknown radar pulses. In *2021 IEEE Radar Conference (RadarConf21)*, pages 1–6. IEEE, 2021.
- [9] Mustafa Atahan Nuhoglu, Yasar Kemal Alp, and Fatih Cagatay Akyon. Deep learning for radar signal detection in electronic warfare systems. In *2020 IEEE Radar Conference (RadarConf20)*, pages 1–6. IEEE, 2020.
- [10] Cédric Villani. *Optimal transport : old and new*, volume 338. Springer, 2009.