Fusion asynchrone résiliente aux pannes pour la détection d'obstacles 3D fondée sur deep learning et propagation d'incertitudes

Mots-clés : Véhicule intelligent, détection obstacles 3D, fusion multimodale des données (a)synchrones, propagation d'incertitudes

---

Encadrement : A. Bensrhair (Pr, Directeur de thèse), A. Rogozan (MC, Encadrante)
Contact : abdelaziz.bensrhair@insa-rouen.fr et alexandrina.rogozan@insa-rouen.fr
Contexte : contrat doctoral INSA Rouen Normandie/LITIS
Début : Septembre/octobre 2024

Résumé : Afin de garantir la qualité des services et un fonctionnement fiable et robuste des véhicules intelligents, il est essentiel d'approfondir les étapes de perception, de localisation et de détection d'obstacles. En effet, ces aspects sont critiques, car ils fournissent les données nécessaires aux modules de décision et de prédiction voire de planification de trajectoires. La thèse proposée concerne la conception d'un système de détection d'obstacles routiers à partir des modalités différentes provenant des caméras de stéréovision, du Lidar, parmi d'autres, lorsque celles-ci sont peu ou pas synchronisées. Dans la thèse proposée, il faudrait appliquer les approches de Deep Learning à la construction d'une architecture de détection d'objets 3D par fusion de données multimodale. La plupart des méthodes existantes reposent sur l'hypothèse que les données sont parfaitement synchronisées, mais plus le nombre de modalités et de sources d'information augmente, plus la contrainte de synchronisation de ces sources est difficile à respecter et à garantir, créant ainsi un verrou scientifique avec un impact fort sur la sécurité des véhicules intelligents. La fusion par propagation d'incertitudes et par fonctions de croyances, au sein d'une architecture de réseaux de neurones profonds à différents niveaux (précoce, intermédiaire et/ou tardif) devrait assurer un système robuste dans des conditions complexes et défavorables (manque d'éclairage, saturation d'images) et résilient aux pannes des caméras et capteurs.

Abstract : In order to guarantee the quality of service and reliable and robust operation for intelligent vehicles, it is essential to improve the perception, localization and detection of obstacles. Indeed, these aspects are critical because they provide the data necessary for decision and prediction or even trajectory planning modules. The proposed thesis concerns the design of a road obstacle detection system using different modalities from stereovision cameras, Lidar, among others, when these are poorly or not synchronized at all. In the proposed thesis, Deep Learning approaches should be applied to the construction of a 3D object detection architecture by multimodal data fusion. Most existing methods rely on the assumption that the data is synchronized, but the more the number of modalities and sources of information increases, the more difficult the synchronization constraint of these sources is to respect and guarantee, thus creating a scientific objective with a strong impact on the safety of intelligent vehicles. Fusion by propagation of uncertainties and belief functions, within a deep neural network architecture at different levels (early intermediate and/or late) should ensure a robust system in complex and unfavorable conditions (lack of lighting, image saturation) and resilient to failures.

Présentation synthétique du sujet de la thèse :

Depuis plusieurs décennies, de nombreux travaux visent à proposer des solutions pour le développement et le déploiement de nouvelles formes de mobilités intelligentes. Ces dernières sont avant tout destinées à améliorer la sécurité routière des participants au trafic routier en réduisant le nombre d'accidents grâce aux systèmes d'aide à la conduite [3,4,5]. Ainsi, plusieurs laboratoires, aussi bien académiques qu'industriels, proposent, pour relever ce défi scientifique, des systèmes de perception de la scène routière, de détection d'obstacles, de prédiction des trajectoires et du risque de collision, (d'aide à la) décision et enfin d'actions automatiques lors des risques importants et imminents d'accidents routiers. Dans ce cadre, s'inscrivent les travaux de recherche liés aux véhicules intelligents voire autonomes. Le sujet de thèse proposé concerne les premières étapes allant de la perception à la prédiction de trajectoire en passant par la détection d'obstacles 3D à partir des modalités différentes provenant des caméras de stéréovision, du Lidar, parmi d'autres, lorsque celles-ci sont peu ou pas synchronisées.

Dans des conditions complexes et défavorables, les systèmes embarqués constitués de divers caméras et capteurs sur les véhicules intelligents montrent leurs limites [5]. Ceci est encore plus flagrant lorsqu'un ou plusieurs capteurs ou caméras sont défaillants. La propagation d'incertitudes et les fonctions des croyances [7, 8], parmi d'autres schémas de fusion, devrait aussi être abordée pour proposer une architecture fondée sur des réseaux de neurones profonds qui soit robuste et résiliente devant la défaillance d'un/des capteur(s) ou dans des conditions environnementales défavorables.

Il n'existe pas encore de paradigme concernant la fusion de données et/ou des sorties des réseaux de neurones profonds qui prouve son efficacité pour la perception et la détection d'objets routiers 3D. De plus, toutes les méthodes existantes de détection d'objets en 3D par fusion de données reposent sur l'hypothèse que les données en entrée sont synchronisées. Mais plus le nombre de modalités et de sources d'information augmente, plus la contrainte de synchronisation de ces sources est difficile à respecter et à garantir, créant ainsi un verrou fort pour l'efficacité de la perception et plus particulièrement de la détection des objets 3D.

Notre équipe STI au sein du laboratoire LITIS est particulièrement reconnue pour ses contributions dans ces trois axes perception-détection-prédiction, grâce aux résultats obtenus lors des travaux de recherche dans le cadre des projets tant au niveau national - collaboration avec l'équipe RITS de l'INRIA Paris [2] qu'au niveau international - collaboration avec l'équipe du Pr Alberto Broggi, Italie.

Le sujet de thèse est une suite logique des thèses passées que nous avons pu encadrer. Les travaux de thèse de Robin Condat [1] étudient l'effet de l'absence d'une des modalités (signal ou canal) durant l'inférence au sein d'un réseau de neurones profond et proposent une approche de type augmentation de données dans un cadre multimodal permettant d'améliorer la robustesse à ce type de défaillance dans le cadre de la détection des piétons. Plus récemment, dans sa thèse de doctorat, Haodi Zhang [2] propose une méthode de fusion hybride fondée sur un modèle de détection d'obstacles routiers 3D utilisant des images aussi bien synchrones qu'asynchrones provenant des sources différentes.

L'objectif de la thèse que nous proposons est d'améliorer les performances du système de perception réalisé suite aux 2 précédentes thèses en intégrant un nouveau processus de fusion plus performant fondé sur la propagation d'incertitude et par fonction de croyance.

Le sujet de thèse devrait aborder l'estimation des incertitudes par les méthodes fonctions de croyance [10], y compris les approches ensemblistes, parmi d'autres, où au lieu d'apprendre un seul réseau de neurones, plusieurs modèles seront entrainés chacun sur une modalité donnée et ensuite leurs prédictions seront agrégées par la règle de Dempster, après estimation des incertitudes. On obtient alors un réseau de neurones profond évidentiel entrainé sur différentes modalités. Celui-ci pourrait être éventuellement ré-entrainé simultanément sur l'ensemble des modalités. En particulier, nous nous intéresserons au formalisme récemment présenté dans [10] qui propose d'intégrer les fonctions de croyance et la théorie de Dempster dans un réseau de neurones profond. Ce formalise a montré récemment un grand intérêt pour représenter les incertitudes et les imprécisions [11], notamment dans divers domaines applicatifs tels que la segmentation en imagerie médicale [12-13]. Il est clair que le domaine de la mobilité intelligente, et en particulier les véhicules intelligents, n'ont pas encore exploré un tel formalisme notateur d'apprentissage profond fonctions de croyance, ni la théorie qui l'accompagne. Ce sujet est au cœur de collaboration naissante entre les équipes STI et Apprentissage et particulièrement les encadrants de cette thèse et du Professeur Paul Honeine.

La théorie de Dempster pourrait être utilisée pour la quantification des incertitudes de prédiction en apprentissage, en ajoutant une couche évidentielle à des réseaux de neurones profonds. Cette approche permet aussi de fusionner des réseaux de neurones, dont les sorties sont exprimées dans des cadres de discernement différents. Une autre possibilité devant être étudiée pour réduire l'incertitudes aléatoires serait l'augmentation des entrées pour une modalité imprécise dans un contexte environnemental particulier avec des données provenant d'une autre modalité. Enfin, il serait intéressant d'étudier la fusion intermédiaire par combinaison probabiliste ou évidentielle des vecteurs des caractéristiques extraites par les réseaux de neurones unimodaux.

Afin de pouvoir garantir une sécurité élevée et un risque minimum, les véhicules automatisés sont équipés de plusieurs capteurs et caméras de technologies différentes. Ces dispositifs sont généralement asynchrones et leur utilisation dans une architecture de perception nécessite de faire un recalage spatial et temporel pour exploiter efficacement leurs données. La majorité des méthodes de fusion existantes utilisent des données synchronisées [8]. Plusieurs stratégies de synchronisation existent : synchronisation sur le capteur le plus rapide, le capteur le plus lent, le capteur le plus fiable, ou sur une période fixe prédéfinie. Chaque méthode a ses avantages et ses inconvénients. Diminuer la fréquence de détection par une synchronisation sur le capteur le plus lent, par exemple, peut entrainer des risques de collisions en raison d'une perception tardive. Nous avons proposé [2] une méthode hybride fondée un modèle de détection d'obstacles routiers 3D utilisant des images aussi bien synchrones qu'asynchrones provenant des sources différentes (caméras stéréovision et LIDAR). En s'appuyant sur les données LiDAR projetées dans l'image 2D associée, ainsi que l'estimation du mouvement entre images, il est possible de synthétiser les données LiDAR manquantes. Le sujet de thèse proposé devrait étudier et proposer d'autres méthodes de fusion en relâchant la contrainte de synchronisation forte toute en prenant en compte la propagation des incertitudes et permettant d'avoir une forte robustesse et une forte résilience à la défaillance d'un capteur ou à des conditions environnementales défavorables ou dégradées.

<u>Déroulement de la thèse</u>

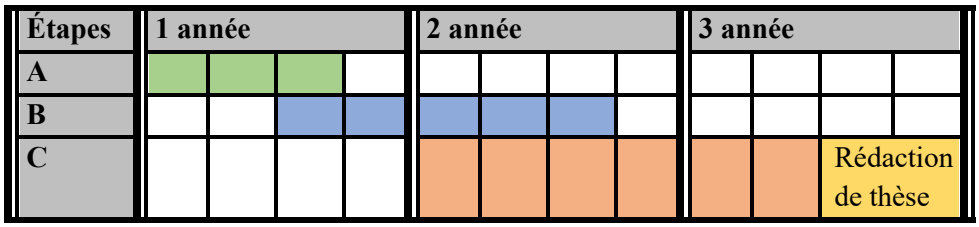
a) Dans un premier temps la doctorante ou le doctorant va commencer par faire un état de l'art sur les algorithmes utilisés pour la perception, la détection d'obstacles et la prédiction de trajectoires, ainsi

qu'étudier leurs limitations. Ensuite explorer les travaux de recherches qui proposent des solutions pour traiter la multimodalité dans des conditions complexes, ainsi que les techniques de fusion et de traitement des données asynchrones et/ou manquantes (pannes de caméras/capteurs). Une attention particulière devrait être apporter au choix de la base de données multimodale suite à l'étude et la comparaison des bases existantes. En fonction de l'objectif visé les bases de données choisies devront être labélisées, reformatées et ou complétées avec des données réelles ou simulées.

b) Dans une deuxième étape la doctorante ou le doctorant est amené à proposer de nouvelles approches fondées sur les réseaux de neurones et l'apprentissage profond pour la détection d'obstacles et/ou la prédiction de trajectoires avec fusion précoce, intermédiaire et tardive des données multimodales. Différentes situations devront être testées ou simulées allant des conditions environnementales difficiles à la panne d'un(e) ou plusieurs caméra(s)/capteur(s).

c) Dans une dernière étape la doctorante ou le doctorant doit mettre en place des approches originales permettant de traiter l'asynchronisme possible des données multimodales dans un cadre unifié avec les méthodes précédentes.

d) Chronologie :



e) Profil nécessaire :
- Diplôme d'ingénieur ou de Master en Signal et/ou Image.
- Solides compétences en Machine learning (apprentissage profond) et en traitement de signal et des images, développement informatique (C/C++, Python, Linux) et des environnements de programmation.
- Très bonnes capacités de communication écrite et orale en français et en anglais.
- Motivation, autonomie, rigueur, force de proposition

Candidature : **avant le 15/05/2024** en envoyant par email à abdelaziz.bensrhair@insa-rouen.fr et

Alexandrina.rogozan@insa-rouen.fr :

1. Lettre de motivation.

2. CV incluant deux références académiques.

3. Relevé de notes provisoires de Master/formation d'ingénieurs

4. Lettre(s) de recommandation.

Bibliographie :

1. R. Condat, A. Rogozan, S. Ainouz, A. Bensrhair. Identifying and deactivating unusable modalities to improve multimodal cnn robustness for road scene understanding, in *the Proceedings of the 25th IEEE International Conference on Intelligent Transportation Systems (ITSC)*, 2022.
2. H. Zhang. Multi-Modal Fusion Based 3D Object Detection, *PhD INSA Rouen (*29/12/2022)
3. Feng, C. Haase-Schütz, L. Rosenbaum, H. Hertlein, C. Glaeser, F. Timm, W. Wiesbeck, K. Dietmayer. Deep multi-modal object detection and semantic segmentation for autonomous driving: Datasets, methods, and challenge, IEEE Transactions on Intelligent Transportation Systems 22 (3) (2020) 1341–1360.
4. C. Sun *et al*. Toward ensuring safety for autonomous driving perception: Standardization progress, research advances, and perspectives, in *IEEE Transactions on Intelligent Transportation Systems*, doi: 10.1109/TITS.2023.3321309
5. P. S. Chib and P. Singh. Recent advancements in end-to-end autonomous driving using deep learning: A survey, in *IEEE Transactions on Intelligent Vehicles*, doi: 10.1109/TIV.2023.3318070,
6. Y. Wang, Q. Mao, H. Zhu, J. Deng and Y. Zhang. Multi-modal 3D object detection in autonomous driving: A survey, in *International Journal of Computer Vision*, 2023 – Springer
7. Z. Tong, P. Xu and T. Denœux. Fusion of Evidential CNN Classifiers for Image Classification. In T. Denoeux, E. Lefèvre, Zh. Liu and F. Pichon (Eds), *Belief Functions: Theory and Applications*, Springer International Publishing, Cham, pp 168–176, 2021
8. C. Xiang *et al*. Multi-Sensor Fusion and Cooperative Perception for Autonomous Driving: A Review, in *IEEE Intelligent Transportation Systems Magazine*, vol. 15, no. 5, pp. 36-58, Sept.-Oct. 2023, doi: 10.1109/MITS.2023.3283864.
9. P. Xu, F. Davoine, J.-B. Bordes, H. Zhao and T. Denœux. Multimodal Information Fusion for Urban Scene Understanding. *Machine Vision and Applications* 27(3) :331–349, 2019
10. Z. Tong, P. Xu, T. Denœux. An evidential classifier based on Dempster-Shafer theory and deep learning. *Neurocomputing*, vol. 450, p. 275-293, 2021
11. Z. Liu and S. Letchmunan. Representing uncertainty and imprecision in machine learning: A survey on belief functions. *Journal of King Saud University-Computer and Information Sciences,* 101904, 2024
12. L. Huang, S. Ruan, and T. Denœux. Application of belief functions to medical image segmentation: A review. *Information fusion* 91: 737-756, 2023
13. S. Xu, et al. Deep evidential fusion network for medical image classification. *International Journal of Approximate Reasoning* 150: 188-198, 2022