

Fusion de données multimodales et structurées par apprentissage profond

Stage de Master 2 (5-6 mois)

Laboratoire : LaBRI, UMR CNRS 5800
351 cours de la Libération, 33400 Talence, France

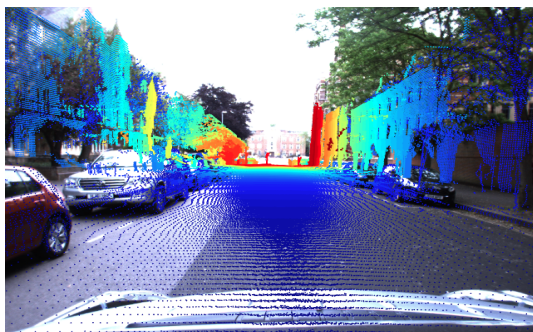
Encadrants : Michaël Clément / LaBRI / Image et Son (michael.clement@labri.fr)
Rémi Giraud / IMS / Signal et Image (remi.giraud@ims-bordeaux.fr)

Mots-clés : *Fusion de données ; Représentations structurelles ; Apprentissage profond*

Description du sujet

Contexte

Les véhicules conçus pour la modélisation urbaine, ou bien pour la conduite autonome, sont généralement équipés de divers capteurs : caméras haute-résolution stéréoscopiques, centrale inertielle, signal GPS, et LiDAR (capteur qui permet de mesurer la distance à une cible et ainsi fournir un nuage de points 3D de la scène). Toutes ces données de différentes natures peuvent être utiles pour analyser et interpréter des scènes d'extérieur urbaines. Par exemple, elles peuvent permettre de détecter et localiser les différents objets qui composent la scène, ou même de segmenter de manière dense l'environnement 3D capturé par le système (voir la Figure 1). Cependant, fusionner de manière efficace toutes ces modalités représente toujours un défi essentiel, notamment afin d'améliorer la qualité et la fiabilité des résultats de détection d'objets ou de segmentation.



(a) Nuage de points projeté sur l'image optique



(b) Détection d'objets urbains en 3D



(c) Segmentation sémantique de scènes

Figure 1 – Exemples d'applications issues des jeux de données KITTI [1] et *RobotCar* [2].

Ces dernières années, plusieurs approches utilisant l'apprentissage profond ont pu obtenir de bons résultats pour la détection de voitures (environ 90% de bonnes détections et localisations), mais les objets plus complexes tels que les piétons ou les cyclistes restent difficile à traiter. De façon analogue, des approches performantes ont été proposées pour la segmentation sémantique d'images optiques, mais l'extension au traitement de nuages de points 3D reste encore particulièrement difficile. De plus, la plupart des méthodes actuelles emploient les différentes modalités de manière séparée, ou bien les combinent de manière *ad-hoc* (par *late fusion*) pour raffiner les résultats [3]. Une méthode par *early fusion* a récemment été proposée [4], mais là aussi se limite à la détection de voitures.

Objectifs du stage

Dans ce stage, nous proposons d'étudier de nouvelles approches de fusion de données multimodales, telles que les images optiques et le LiDAR, en se focalisant notamment sur l'exploitation de leurs structures spatiales. En effet, les capteurs LiDAR récents suivent, lors des captures, une certaine structure qui est intrinsèquement liée au balayage régulier de la scène, et qui peut être exploitée pour extraire de l'information [5, 6]. De même, les images issues de caméras présentent une structure spatiale de haut niveau dû au positionnement relatif des différents objets. Il paraît alors pertinent d'intégrer au sein des approches de fusion de données des représentations irrégulières telles que des décompositions en superpixels [7] ou des graphes décrivant les relations spatiales entre régions [8].

Ces représentations irrégulières pourraient apporter des informations additionnelles pertinentes aux chaînes de traitement actuelles reposant sur des réseaux de neurones convolutifs. Un défi sera notamment de proposer des architectures adaptées à la manipulation de données irrégulières (où le nombre d'éléments et leurs formes sont variables), afin qu'elles puissent être exploitées de manière complémentaires aux caractéristiques d'apparence classiques à l'échelle pixellique. L'approche pourra également s'inspirer de travaux récents sur les réseaux de neurones non-locaux [9], qui permettent de capturer des interactions à différents niveaux dans les images.

Profil recherché

Étudiant-e en Master 2 ou dernière année d'école d'ingénieurs, spécialisé-e en informatique, image et/ou intelligence artificielle. Des connaissances et expériences solides en traitement d'images, apprentissage profond et programmation (Python, TensorFlow/PyTorch, C/C++) sont demandées. La maîtrise de l'anglais scientifique et des qualités rédactionnelles sont également très importantes.

Candidature : Pour candidater, envoyer un dossier avec CV, lettre de motivation, relevés de notes, ainsi que toute pièce susceptible de renforcer la candidature (lettre de recommandation, etc.). Pour l'envoi des pièces demandées, ou pour toute question sur le sujet : michael.clement@labri.fr

Références

- [1] Geiger, A., Lenz, P., and Urtasun, R. "Are we ready for Autonomous Driving? The KITTI Vision Benchmark Suite". *Conference on Computer Vision and Pattern Recognition (CVPR)*. 2012.
- [2] Maddern, W., Pascoe, G., Linegar, C., and Newman, P. "1 Year, 1000km: The Oxford RobotCar Dataset". *The International Journal of Robotics Research (IJRR)* 36.1 (2017), pp. 3–15.
- [3] Chen, X., Ma, H., Wan, J., Li, B., and Xia, T. "Multi-view 3D object detection network for autonomous driving". *Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017.
- [4] Liang, M., Yang, B., Chen, Y., Hu, R., and Urtasun, R. "Multi-task multi-sensor fusion for 3D object detection". *Conference on Computer Vision and Pattern Recognition (CVPR)*. 2019.
- [5] Biasutti, P., Aujol, J.-F., Brédif, M., and Bugeau, A. "Range-Image: Incorporating sensor topology for LiDAR point cloud processing". *Photogrammetric Engineering & Remote Sensing* 84.6 (2018), pp. 367–375.
- [6] Biasutti, P., Lepetit, V., Aujol, J.-F., Brédif, M., and Bugeau, A. "LU-Net: An efficient network for 3D LiDAR point cloud semantic segmentation based on end-to-end-learned 3D features and U-Net". *Computer Vision for Road Scene Understanding and Autonomous Driving (CVRSUAD, ICCV workshop)*. 2019.
- [7] Giraud, R., Ta, V.-T., Bugeau, A., Coupé, P., and Papadakis, N. "SuperPatchMatch: An algorithm for robust correspondences of superpixel patches". *IEEE Transactions on Image Processing* 26.8 (2017), pp. 4068–4078.
- [8] Clément, M., Kurtz, C., and Wendling, L. "Learning spatial relations and shapes for structural object description and scene recognition". *Pattern Recognition* 84 (2018), pp. 197–210.
- [9] Wang, X., Girshick, R., Gupta, A., and He, K. "Non-local neural networks". *Conference on Computer Vision and Pattern Recognition (CVPR)*. 2018, pp. 7794–7803.