

Offre de thèse

Localisation visuelle à long terme pour le véhicule intelligent.

Encadrants : Pr. Rémi BOUTTEAU (LITIS, Université de Rouen Normandie) et Dr. Hind LAGHMARA (LITIS, INSA de Rouen Normandie)

Lieu : LITIS UR4108 Avenue de l'Université, 76800 Saint Etienne du Rouvray (Rouen), France
<https://www.litislab.fr/>

Date de début : Septembre/octobre 2024

Mots-clés : Localisation visuelle, Vision par Ordinateur, SLAM, Deep Learning, Véhicule Intelligent

Contexte

Ces dernières années, le SLAM (*Simultaneous Localization and Mapping*) visuel a été un axe de recherche particulièrement actif dans le domaine de la robotique pour la localisation en temps réel des systèmes mobiles [Mur-Artal et al, 2015] [Mur-Artal et al, 2017]. Le SLAM visuel permet d'estimer, à partir d'une ou plusieurs caméras embarquées, la trajectoire du mobile (robot, véhicule autonome, etc) et la structure tridimensionnelle de la scène dans laquelle il opère.

Le principal avantage du SLAM visuel est qu'il exploite un capteur à bas coût pour estimer la position du véhicule : la caméra. Le coût d'une caméra est en effet très faible par rapport aux autres capteurs habituellement utilisés pour la localisation des véhicules autonomes (quelques dizaines de milliers d'euros pour un lidar ou un système GNSS à haute précision).

Pour garantir le meilleur compromis entre vitesse et précision, la majorité des algorithmes de SLAM visuel sont basés sur une représentation éparsée de l'environnement. En mettant en correspondance des amers visuels sur des images successives, ou entre une carte de l'environnement et l'image courante de la caméra, il est possible d'estimer la pose (position, orientation) de la caméra et donc du véhicule. Ces amers sont généralement extraits grâce à des détecteurs de points d'intérêts (Harris, SIFT, SURF, ORB) auxquels on associe un descripteur qui sera nécessaire pour la mise en correspondance.

Quelques algorithmes ont montré d'excellentes performances et sont parfois proposés à la communauté en open source. Nous pouvons en particulier mentionner l'algorithme ORB-SLAM [Mur-Artal et al, 2015] et son évolution ORB-SLAM2 [Mur-Artal et al, 2017]. Ces algorithmes sont particulièrement adaptés aux environnements intérieurs, ou extérieurs lorsqu'il se passe peu de temps (et donc de changements) entre la phase de construction de la carte et la phase de relocalisation.

Cependant, la localisation par SLAM visuel présente encore de nombreux défis dans un environnement extérieur sujet à des changements d'apparence, notamment :

- Les changements de saisons
- Les changements d'illumination
- Les changements météorologiques
- Les changements de points de vue
- Les changements dus aux objets dynamiques, c'est-à-dire pouvant être présents lors de la création de la carte mais absents lors de la relocalisation, ou vice versa.

La relocalisation dans des zones précédemment cartographiées peut donc être une tâche extrêmement difficile en environnement extérieur puisque l'apparence de l'environnement ne cesse de changer. Dans le contexte du véhicule autonome, la tâche de localisation est critique car si celle-ci est inexacte, voire totalement erronée, elle peut entraîner de mauvaises décisions sur la conduite du véhicule.

Objectifs

Les travaux proposés dans le cadre du sujet de cette thèse se concentreront sur la localisation visuelle à long terme pour les véhicules autonomes dans des environnements extérieurs où l'apparence est soumise à des changements substantiels.

Quelques travaux ([Milford et al, 2012], [Bonardi et al, 2017], [Hausler et al, 2021]) ont été proposés sans la littérature pour tenter de résoudre ce problème mais ils traitent alors principalement du problème de la reconnaissance de lieu (place-recognition), c'est-à-dire retrouver l'image la plus proche visuellement dans une base d'images pré-enregistrée. Ce problème, bien qu'indispensable pour obtenir une première localisation « grossière », ne permet pas d'obtenir une localisation métrique nécessaire pour le contrôle du véhicule.

Le premier objectif de cette thèse sera consacré à la création de cartes permettant ensuite la relocalisation du véhicule. La plupart des travaux sur la localisation visuelle construisent bien évidemment la carte à partir des informations issues de caméras, c'est notamment le cas de ORB-SLAM [Mur-Artal et al, 2017]. Bien que populaire par sa simplicité de mise en œuvre, cette méthode présente néanmoins plusieurs inconvénients : déformation des cartes à cause de la dérive accumulée, estimation de l'information 3D à partir d'images 2D, absence de référencement GNSS. Nous proposons dans cette thèse de fusionner plusieurs modalités de perception et de localisation dans la phase de construction de la carte : lidar, caméra, GNSS.

Le deuxième objectif de cette thèse est que la localisation soit invariante aux changements d'apparence. Pour cela, la carte devra être enrichie et filtrée grâce à de multiples passages du véhicule de mapping (différentes saisons, différentes conditions d'illumination, etc). Deux pistes seront explorées pour rendre la carte plus robuste aux changements d'apparence :

- La proposition d'un descripteur invariant aux changements d'illumination et de saisons. Les approches hand-crafted ont montré leurs limitations dans ce domaine et la tendance actuelle est plutôt aux descripteurs appris par des réseaux de neurones profonds [Piasco et al, 2019] [Gridseth et al, 2022].
- L'intégration d'informations sémantiques à la carte. Ce type d'information permettrait de ne retenir que les points d'intérêts qui n'appartiennent pas à des objets dynamiques (voitures, piétons, nuages, etc) ou statiques mais susceptibles de disparaître de la carte (voitures garées, poubelles, etc). Cela permettrait également de renforcer le poids des amers jugés comme intéressants pour la localisation visuelle : il est par exemple plus intéressant de prendre des amers sur des bâtiments ou sur la signalisation plutôt que sur la végétation car ces amers seront moins sujets à des changements. L'information sémantique peut également être intéressante pour la description des points d'intérêts puisque moins sensible aux changements d'illumination qu'un descripteur basé sur les variations d'intensités (SIFT, ORB, etc). Cela pourrait aider à réduire le nombre d'erreurs d'appariement et donc augmenter la robustesse globale du système.

Enfin, l'objectif final de ces travaux de thèse est d'obtenir une localisation métrique, à partir d'informations visuelles, en exploitant également les capteurs présents de série dans les véhicules (notamment l'odométrie et les informations inertielles) dans la carte que nous aurons construite

précédemment. Les principales approches probabilistes permettant d'estimer cette localisation sont les filtres de Kalman (KF, EKF, UKF) et les filtres d'histogrammes [Thrun et al, 2005]. Cependant, les filtres de Kalman supposent que la distribution de l'état puisse s'approximer par une gaussienne unimodale. Les filtres d'histogrammes discrétisent l'espace d'état et ne peuvent pas s'appliquer à des espaces de grandes dimensions. Pour pallier ces problèmes, les filtres particuliers ont vu le jour et ont déjà démontré leurs performances dans les approches basées lidar [Merriault et al, 2018]. Peu de travaux les exploitent cependant pour la localisation visuelle. Nous proposons dans cette thèse d'exploiter cette approche pour la localisation visuelle. Le principal verrou scientifique à lever dans le cadre de ces travaux sera de déterminer la meilleure fonction de vraisemblance afin d'exploiter pleinement le potentiel des cartes que nous aurons construites.

Le doctorant aura des moyens expérimentaux lui permettant l'acquisition de données et l'évaluation de ses algorithmes, notamment un véhicule instrumenté équipé d'un système de navigation GNSS-RTK lui permettant d'obtenir une vérité terrain fiable.

Profil nécessaire :

- Diplôme d'ingénieur ou de Master en informatique, traitement du signal.
- Solides compétences en informatique (apprentissage profond) et/ou traitement de signal, développement informatique (C/C++, Python, Linux) et des environnements de programmation et simulation (Matlab, ROS, etc.).
- 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 remi.boutteau@univ-rouen.fr et hind.laghmara@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 :

[Bonardi et al, 2017] Bonardi, F., Ainouz, S., Boutteau, R., Dupuis, Y., Savatier, X., Vasseur, P., (2017) : PHROG: A multimodal Feature for Place Recognition, *Sensors*, 17, 1167

[Gridseth et al, 2022] M. Gridseth and T. D. Barfoot, "Keeping an Eye on Things: Deep Learned Features for Long-Term Visual Localization," in *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 1016-1023, April 2022, doi: 10.1109/LRA.2021.3136867.

[Hausler et al, 2021] Hausler, S., Garg, S., Xu, M., Milford, M., & Fischer, T. (2021). Patch-netvlad: Multi-scale fusion of locally-global descriptors for place recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 14141-14152).

[Merriault et al, 2018] Merriault, P., Dupuis, Y., Boutteau, R., Vasseur, P., Savatier, X., (2018) : Robust Robot Localization in a Complex Oil and Gas Industrial Environment, *Journal of Field Robotics (JFR)*, vol. 35, Issue 2, March 2018, pp 213-230.

[Milford et al, 2012] M. J. Milford and G. F. Wyeth, "SeqSLAM: Visual route-based navigation for sunny summer days and stormy winter nights," 2012 *IEEE International Conference on Robotics and Automation*, 2012, pp. 1643-1649, doi: 10.1109/ICRA.2012.6224623.

[Mur-Artal et al, 2015] Raúl Mur-Artal, J. M. M. Montiel and Juan D. Tardós. ORB-SLAM: A Versatile and Accurate Monocular SLAM System. *IEEE Transactions on Robotics*, vol. 31, no. 5, pp. 1147-1163, 2015

[Mur-Artal et al, 2017] Raúl Mur-Artal and Juan D. Tardós. ORB-SLAM2: an Open-Source SLAM System for Monocular, Stereo and RGB-D Cameras. IEEE Transactions on Robotics, vol. 33, no. 5, pp. 1255-1262, 2017

[Piasco et al, 2019] Piasco, N., Sidibé, D., Gouet-Brunet, V., & Demonceaux, C. (2019, May). Learning scene geometry for visual localization in challenging conditions. In 2019 International Conference on Robotics and Automation (ICRA) (pp. 9094-9100). IEEE.

[Thrun et al, 2005] Thrun, S., Burgard, W., Fox, D., (2005) : Probabilistic Robotics, MIT Press.