

INSTITUT DES SYSTEMÈS INTELLIGENTS ET DE ROBOTIQUE*

OFFRE DE THESE

« Apprentissage de la saisie d'objets »

Ce projet de thèse s'insère dans le projet fédérateur de l'ISIR sur la robotique agricole. Il porte sur la saisie en robotique, en particulier dans un contexte agricole (saisie de fruits ou de légumes). Ce sujet sera abordé par une approche pluri-disciplinaire entre informatique (encadrement S. Doncieux et A. Coninx), mécanique et automatique (encadrement F. Ben Amar).

Contexte : Ce projet de thèse porte sur la saisie d'objets par un préhenseur robotique. Le but est, à terme, de concevoir un système de cueillette automatique de fruits et de légumes. Il s'agit donc de saisir des objets fragiles, de formes et couleurs variées, objets qui peuvent de plus être au moins partiellement occultés par le feuillage de la végétation. Ce projet de thèse portera sur une approche basée apprentissage dans laquelle une expertise significative côté modélisation mécanique sera utilisée à différents niveaux de la démarche méthodologique, de la conception du système apprenant à l'analyse de sa performance.

La saisie d'objets est une tâche emblématique en robotique car c'est un prérequis pour de nombreuses autres tâches plus avancées tels que leur manipulation. Du point de vue de l'apprentissage, elle représente un défi particulièrement difficile car elle correspond à un cas de récompense rare : très peu d'interactions sont susceptibles de mener à une saisie correcte. De nombreux mouvements ne toucheront même pas l'objet concerné. Le robot peut donc passer beaucoup de temps à explorer des comportements qui ne sont pas pertinents. Ce problème est fréquemment abordé en fournissant des démonstrations au robot ou des primitives de mouvements qui, sans être parfaites, réussissent à générer des saisies correctes avec une probabilité suffisantes pour que de tels mouvements soient observés pendant la phase d'exploration [1]. Ce projet vise à résoudre ce problème au travers de méthodes d'exploration appropriées et qui viendront compléter les méthodes d'apprentissage profond existantes.

Objectif scientifique : Ce projet de thèse vise à définir une approche autonome de l'apprentissage en robotique dans le cas de récompenses et d'interactions rares. Les développements réalisés seront testés sur le problème de la saisie d'objets en visant en particulier des applications dans le domaine agricole. Ils s'appuieront sur les travaux des encadrants sur des méthodes d'exploration de type recherche de nouveauté et algorithmes de qualité-diversité [2, 3, 4] ainsi que sur des résultats préliminaires sur l'apprentissage de comportement de saisie d'objets avec ces méthodes. Se plaçant dans le cadre défini dans projet Européen DREAM, l'approche développée sera itérative [5, 6]. Elle visera tout d'abord à générer des comportements de saisie par une approche mixte entre exploration en simulation et test en réalité. Cette étape s'appuiera sur un apprentissage de politiques en boucle ouverte. Chaque saisie d'un objet à une position particulière nécessitera donc un apprentissage dédié. La répétition de tels apprentissages permettra de générer une base d'exemples suffisante pour entraîner un système d'apprentissage profond qui associera un comportement adapté à la

*L'Institut des Systèmes Intelligents et de Robotique (Isir) est une Unité Mixte de Recherche (UMR7222) sous tutelle de Sorbonne Université, du CNRS et de l'Inserm (ERL-U1150). Ce laboratoire de recherche pluridisciplinaire rassemble des chercheurs et enseignants-chercheurs relevant de différentes disciplines des Sciences de l'Ingénieur et de l'Information ainsi que des Sciences du Vivant.

perception visuelle du robot. L'objectif principal de la thèse sera de générer ces bases d'exemples grâce à des méthodes d'exploration adaptées.

Cet objectif peut se décomposer en 3 sous-objectifs :

1. Définition d'espaces comportementaux adaptés à la saisie.

Les méthodes d'exploration basées sur des algorithmes de recherche de nouveauté ou de qualité diversité visent à couvrir, de façon aussi exhaustive que possible, un espace comportemental qu'il est pertinent d'explorer. La diversité des solutions ainsi trouvées permet d'identifier les comportements possibles et de disposer de nombreuses alternatives pour atteindre un même but. Cet apprentissage étant réalisé en simulation, la disponibilité des différentes alternatives augmente les chances de trouver des solutions transférables et efficaces sur le système réel [4]. De plus, cela multiplie les exemples et donc la taille de la base d'apprentissage. Plusieurs approches seront comparées, de la conception experte en s'appuyant sur des modèles mécaniques de la saisie à l'apprentissage de ces descripteurs [7].

2. Définition de politiques appropriées pour des gestes de saisie.

Les méthodes d'exploration mentionnées nécessitent de définir des politiques paramétrées. Le choix de la politique va déterminer l'espace des possibilités et des façons de saisir les objets. Il convient donc de choisir une représentation polyvalente permettant de tirer le meilleur parti des capacités des différents robots utilisés, qu'ils soient classiques (bras robotisés de type Franka Emika) ou non (robot souple de type trompe d'éléphant en cours de développement à l'ISIR). La méthode d'apprentissage doit s'adapter à ces différentes configurations pour en tirer le meilleur parti pour la tâche de saisie considérée. Cette étape s'appuiera également sur une forte expertise en mécanique et en automatique pour définir des primitives de mouvements paramétrées suffisamment génériques.

3. Apprentissage de modèles de saisie. Les méthodes d'exploration évoquées nécessitent de nombreux tests de politiques.

Il n'est pas possible de tous les réaliser sur le robot réel qui serait trop vite endommagé. Il est donc nécessaire d'en réaliser la majeure partie en simulation, ce qui pose la question d'un modèle suffisamment précis pour qu'une part significative des comportements appris se transfère sans difficulté sur le système réel. Or les comportements de saisie impliquent des contacts qui sont souvent mal maîtrisés dans les modèles et les simulateurs robotiques utilisées pour l'apprentissage (mujoco ou pybullet, par exemple). Dans une perspective itérative, cette partie de la thèse s'appuiera sur les données acquises pour corriger un modèle existant ou apprendre un modèle complet de la physique sous-jacente.

Comme pour les sous-objectifs précédents, ce sujet sera abordé dans une perspective pluridisciplinaire entre informatique pour l'apprentissage et mécanique pour la modélisation de ces systèmes dynamiques en contacts multiples avec un environnement mal maîtrisé [12][13]. On tentera de répondre à une question fondamentale d'actualité : comment peut-on rapprocher les

**L'Institut des Systèmes Intelligents et de Robotique (Isir) est une Unité Mixte de Recherche (UMR7222) sous tutelle de Sorbonne Université, du CNRS et de l'Inserm (ERL-U1150). Ce laboratoire de recherche pluridisciplinaire rassemble des chercheurs.euse.s et enseignant.e.s-chercheur.euse.s relevant de différentes disciplines des Sciences de l'Ingénieur et de l'Information ainsi que des Sciences du Vivant.*

méthodes de modélisation basées données des méthodes de modélisation physiques, afin d'en combiner les avantages, à savoir l'efficacité des premiers et l'explicabilité des seconds ?

Justification de l'approche : L'apprentissage par renforcement recherche une politique, autrement dit une fonction associant un état à une action maximisant une récompense. Cela permet de trouver le comportement approprié pour atteindre un objectif qui n'est décrit que par la survenue de récompenses que le système cherche à maximiser. Les succès retentissants de l'apprentissage par renforcement profond [8,9] ont cependant peu d'équivalents en robotique. Ce domaine cumule en effet les difficultés pour l'apprentissage [10], avec des espaces d'états et d'actions de très grandes dimensions, les transitions entre continu et discret, des perceptions et actions bruitées, des récompenses rares, ... Plusieurs approches sont possibles pour faire face à ces défis [11]: (1) des approches "pas à pas", dans lesquelles un modèle est utilisé pour tester plusieurs actions possibles qu'un critère permettra de discriminer pour choisir la meilleure et (2) des approches basées épisodes dans lesquelles la politique est décrite par une fonction paramétrée qui est évaluée au cours d'un épisode complet, c'est-à-dire d'une séquence de perceptions-actions susceptible de mener à une récompense. Les méthodes "pas à pas" sont plus efficaces et généralisent mieux, mais ont du mal à gérer le cas de récompenses rares. Les approches basées épisodes peuvent faire face à ce problème, mais au prix de plus d'exploration et d'une faible capacité de généralisation. Ce projet de thèse vise à combiner le meilleur de ces deux types d'approches en gérant les récompenses rares avec une approche basée épisodes pour générer une base d'entraînement permettant de démarrer l'apprentissage d'un système "pas à pas".

Profil recherché : Étudiant en informatique avec un profil "apprentissage machine".

Compétences requises : La programmation en python doit être maîtrisée. Des compétences en robotique sont souhaitables.

- **Date limite des candidatures : 23 avril 2021**
- **Directeur de thèse : Stéphane Doncieux**
- **Codirection éventuelle :** Faiz Ben Amar, Alexandre Coninx
- **Laboratoire d'accueil :** Isir (Institut des Systèmes Intelligents et de Robotique), 4 Place Jussieu 75005, Paris
- **Contact : Stéphane Doncieux ; doncieux@isir.upmc.fr ;** Envoyer votre candidature par mail, avec [Thèse : Apprentissage de la saisie d'objets] en objet, un CV et une lettre de motivation.

Références (les publications des encadrants sont marquées par un '*') :

[1] Ibarz, J., Tan, J., Finn, C., Kalakrishnan, M., Pastor, P., & Levine, S. (2021). How to train your robot with deep reinforcement learning: lessons we have learned. The International Journal of Robotics Research, 0278364920987859.

[2]* Doncieux, S., Laflaquière, A., & Coninx, A. (2019, July). Novelty search: a theoretical perspective. In Proceedings of the Genetic and Evolutionary Computation Conference (pp. 99-106).

[3]* Paolo, G., Coninx, A., Doncieux, S., & Laflaquière, A. (2021). Sparse Reward Exploration via Novelty Search and Emitters. arXiv preprint arXiv:2102.03140.



- [4]* Kim, S., Coninx, A., & Doncieux, S. (2021). From exploration to control: learning object manipulation skills through novelty search and local adaptation. *Robotics and Autonomous Systems*, 136, 103710.
- [5]* Doncieux, S., Filliat, D., Díaz-Rodríguez, N., Hospedales, T., Duro, R., Coninx, A., ... & Sigaud, O. (2018). Open-ended learning: a conceptual framework based on representational redescription. *Frontiers in neurorobotics*, 12, 59.
- [6]* Doncieux, S., Bredeche, N., Goff, L. L., Girard, B., Coninx, A., Sigaud, O., ... & Duro, R. (2020). Dream architecture: a developmental approach to open-ended learning in robotics. *arXiv preprint arXiv:2005.06223*.
- [7]* Paolo, G., Laflaquiere, A., Coninx, A., & Doncieux, S. (2020, May). Unsupervised learning and exploration of reachable outcome space. In *2020 IEEE International Conference on Robotics and Automation (ICRA)* (pp. 2379-2385). IEEE.
- [8] Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., ... & Hassabis, D. (2017). Mastering the game of go without human knowledge. *nature*, 550(7676), 354-359.
- [9] Badia, A. P., Piot, B., Kapturowski, S., Sprechmann, P., Vitvitskyi, A., Guo, Z. D., & Blundell, C. (2020, November). Agent57: Outperforming the atari human benchmark. In *International Conference on Machine Learning* (pp. 507-517). PMLR.
- [10] Kober, J., Bagnell, J. A., & Peters, J. (2013). Reinforcement learning in robotics: A survey. *The International Journal of Robotics Research*, 32(11), 1238-1274.
- [11] Sigaud, O., & Stulp, F. (2019). Policy search in continuous action domains: an overview. *Neural Networks*, 113, 28-40.
- [12]* Bouton, A. and Grand, A. and Ben Amar, F. (2017). Obstacle Negotiation Learning for a Compliant Wheel-on-Leg Robot. *Int. Conf. on Robotics and Automation, IEEE/ICRA*. 2017.
- [13]* Bouton, A. and Grand, Ch. and Ben Amar, F. (2020). Design and Control of a Compliant Wheel-on-Leg Rover which Conforms to Uneven Terrain. *Transaction on Mechatronics*. Volume: 25, Issue: 5, Oct. 2020...