Département Intelligence Ambiante et Systèmes Interactifs
Laboratoire Vision et Ingénierie des Contenus
3D & Mobilité

Proposition de stage 2020

5- Etude de l'impact de différents types de données pour la localisation 3D d'un robot par *deep-learning*

Contact: mohamed.tamaazousti@cea.fr

Référence: *st3D-5*

Contexte du stage

Le développement des systèmes d'apprentissage automatique est un domaine de recherche en pleine effervescence. Ces systèmes nécessitent un important jeu de données, souvent annotées par des humains : on parle alors de supervision humaine. Certaines approches permettent également un apprentissage via le monde environnant (supervision physique) [1]. Des études cognitives [2,3] proposent aussi de s'appuyer sur l'utilisation de plusieurs modalités (la vision, le son, etc.) en exploitant la synchronisation entre modalités. Par exemple, la chute d'un objet va associer la vision du mouvement et le son de l'impact.

Objectifs du stage

Un robot apprenant automatiquement de son environnement doit être immergé dans un contexte riche et réaliste pour atteindre son plein potentiel. Des travaux antérieurs ont démontré les bénéfices d'un apprentissage interactif et multimodal [4, 5, 6]. L'objectif du stage est d'aller plus loin dans l'apprentissage d'un modèle d'environnement réaliste, interactif, dynamique et multimodal. En s'appuyant sur les récents progrès en simulation numérique d'environnements virtuels, plus précisément la plateforme HoME (*Household Multimodal Environment*) introduite par Brodeur *et al.* [7] qui permet aux agents de naviguer et d'interagir dans plus de 45 000 maisons conçues à partir du jeu de données SUNCG [8], il s'agira d'étudier l'impact de différents types de données utiles à la localisation 3D par réseaux-de-neurones en adressant les tâches suivantes :

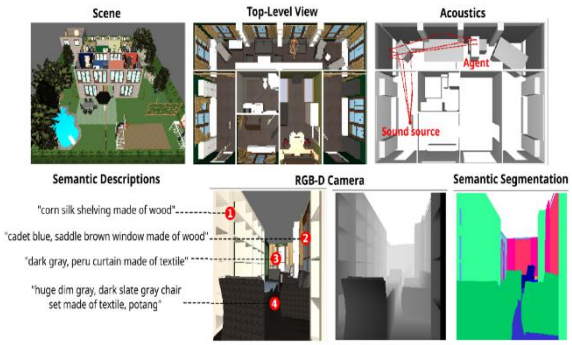

Illustration de la plateforme HoME.

- Trouver comment formaliser l'environnement pour obtenir des tâches intermédiaires;
- Utiliser ces tâches pour apprendre (via *deep learning* ou *reinforcement learning*) une représentation uni-modale et multimodale de la scène;
- Utiliser ces représentations issues de combinaisons des modalités pour résoudre la tâche d'intérêt, à savoir, la localisation 3D dans la scène du robot.

Compétences requises :

L'étudiant devra disposer de solides compétences en mathématiques appliquées, en apprentissage et en vision par ordinateur. Une connaissance du langage de programmation Python, et des frameworks d'apprentissage automatique (Pytorch, TensorFlow, Keras ou Caffe), ainsi que des librairies d'apprentissage automatique (Scikit-learn) et de structures de données (Pandas) sont nécessaires.

Références

[1] Kelly Fisher, Kathy Hirsh-Pasek, and Roberta M Golinkoff. *Fostering mathematical thinking through playful learning*. Contemporary debates on child development and education, pp. 81–92, 2012

[2] Barbara Landau, Linda Smith, and Susan Jones. *Object perception and object naming in early development*. Trends in cognitive sciences, 2(1):19–24, 1998.

[3] Linda Smith and Chen Yu. *Infants rapidly learn word-referent mappings via cross-situational statistics*. Cognition, 106(3):1558–1568, 2008.

[4] Karl Moritz Hermann, Felix Hill, Simon Green, Fumin Wang, Ryan Faulkner, Hubert Soyer, David Szepesvari, Wojtek Czarnecki, Max Jaderberg, Denis Teplyashin, et al. *Grounded language learning in a simulated 3d world*. arXiv preprint arXiv:1706.06551, 2017.

[5] Junhyuk Oh, Satinder Singh, Honglak Lee, and Pushmeet Kohli. *Zero-shot task generalization with multi-task deep reinforcement learning*. In ICML, 2017.

[6] Harm de Vries, Florian Strub, Sarath Chandar, Olivier Pietquin, Hugo Larochelle, and Aaron C. Courville. *Guesswhat?! visual object discovery through multi-modal dialogue*. In CVPR, 2017.

[7] Brodeur, Simon and Perez, Ethan and Anand, Ankesh and Golemo, Florian and Celotti, Luca and Strub, Florian and Rouat, Jean and Larochelle, Hugo and Courville, Aaron. *HoME: A household multimodal environment*. ICLR 2018.

[8] Song, Shuran and Yu, Fisher and Zeng, Andy and Chang, Angel X and Savva, Manolis and Funkhouser, Thomas. *Semantic Scene Completion from a Single Depth Image*. CVPR 2017

Formation / Niveau d'étude	Ingénieur, Master 2 / Bac+5
Possibilité poursuite	Oui, en thèse ou CDD selon profil.
Durée	6 mois
Lieu	Palaiseau (91) – Centre d'intégration de Nano-INNOV
Indemnités de stage	Entre 700 € et 1400 € suivant formation. Aide au logement / transport / restauration.

Candidatures

- Joindre CV + lettre de motivation à **mohamed.tamaazousti@cea.fr** avec le nom du stage auquel vous postulez
- Ne pas hésiter à détailler les projets ou cours auxquels vous avez participé
- Indiquer les dates de début/fin de stage envisagées.
- Ce stage pourra prendre une orientation recherche ou industrie en fonction du profil du candidat