

## **Adaptation des méthodes de reconnaissance visuelle pour divers points de vue**

### **Présentation du laboratoire d'accueil**

Basé à Paris-Saclay, le CEA List est l'un des quatre instituts de recherche technologique de CEA Tech, direction de la recherche technologique du CEA. Dédié aux systèmes numériques intelligents, il contribue au développement de la compétitivité des entreprises par le développement et le transfert de technologies.

L'expertise et les compétences développées par les 800 ingénieurs-chercheurs et techniciens du CEA List permettent à l'Institut d'accompagner chaque année plus de 200 entreprises françaises et étrangères sur des projets de recherche appliquée s'appuyant sur 4 programmes et 9 plateformes technologiques. 21 start-ups ont été créées depuis 2003.

Le Laboratoire de Vision et Apprentissage pour l'analyse de scène (LVA) mène ses recherches dans le domaine de la Vision par Ordinateur (Computer Vision) selon quatre axes principaux :

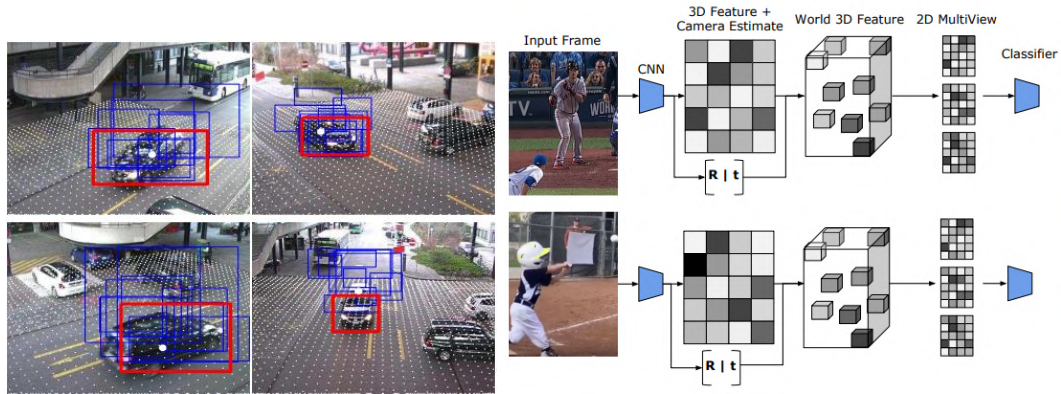
- La reconnaissance visuelle (détection et/ou segmentation d'objets, de personnes, de patterns ; détection d'anomalies ; caractérisation)
- L'analyse du comportement (reconnaissance de gestes, d'actions, d'activités, de comportements anormaux ou spécifiques pour des individus, un groupe, une foule)
- L'annotation intelligente (annotation à grande échelle de données visuelles 2D/3D de manière semi-automatique)
- Les modèles de perception pour l'aide à la décision.

### **Description du stage**

Les méthodes de reconnaissance visuelle (détection d'objets, segmentation sémantique et d'instances, etc.) ont connu un grand essor ces dernières années atteignant de très bonnes performances sur la plupart des datasets publics. Cependant, leurs performances baissent drastiquement quand elles sont testées sur les données différentes de celles d'apprentissage. Les méthodes d'adaptation de domaine ont pour but de réduire la différence entre les domaines, mais leur application est souvent limitée quand la différence entre les domaines est dû aux changements de points de vue. En effet, les transformations géométriques qui apparaissent sur les images quand la caméra change de position représentent un verrou pour les méthodes d'apprentissage automatique actuelles. L'objectif du stage est d'explorer les méthodes de reconnaissance visuelle qui prennent en compte les transformations géométriques provoquées par le changement de point de vue de la caméra afin de rendre les modèles invariants ou robustes à de tels changements.

Nous nous intéressons dans ce stage aux modèles d'auto-calibration [1] capables de prédire, à partir d'une image, les paramètres extrinsèques de la caméra. Ces paramètres peuvent être utilisés dans les apprentissages afin d'obtenir des *features* 3D qui ne dépendent pas du point de vue [2]. Ces *features* serviront dans les tâches finales de reconnaissance visuelle, telles que la détection d'objet, reconnaissance d'action et d'interaction [2,5].

L'objectif du stage est, dans un premier temps, d'étudier plusieurs méthodes de l'état de l'art de l'auto-calibration et adaptation de domaine en utilisant l'information de calibration estimée. Ensuite, le candidat devra évaluer leurs performances sur des datasets multi-vues publics [3,4]. Le candidat sera également invité à proposer des améliorations aux méthodes de l'état de l'art pour pallier un ou plusieurs problèmes identifiés ou adapter les méthodes à d'autres tâches de reconnaissance visuelle [5]. Les travaux menés durant le stage pourront faire l'objet de publications scientifiques.



L'image de gauche : exemples d'images du dataset multiview multi-classe pour la détection d'objets [3]. L'image de droite : la représentation des features 3D utilisée pour la reconnaissance d'actions invariante aux points de vues [2].

### Keywords

Viewpoint estimation, domain adaptation, camera calibration, camera self-calibration, object detection, action recognition, interaction detection.

### Références

- [1] G. Iyer, R. K. Ram., J. K. Murthy, K. M. Krishna. CalibNet: Geometrically Supervised Extrinsic Calibration using 3D Spatial Transformer Networks. In *Proceedings of the IEEE International Conference on Intelligent Robots and Systems (IROS)*, 2018.
- [2] AJ Piergiovanni, M. S. Ryoo. Recognizing Actions in Videos from Unseen Viewpoints. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021.
- [3] G. R. Noguera; X. B. Bosch; H. B. Shitrit; P. Fua, Conditional Random Fields for Multi-Camera Object Detection, In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2011.
- [4] G. Vaquette, A. Orcesi, L. Lucat, C. Achard. The DAily Home Life Activity Dataset: A High Semantic Activity Dataset for Online Recognition. . In *Proceedings of the IEEE 12th International Conference on Automatic Face & Gesture Recognition*, 2017.
- [5] S. Chafik, A. Orcesi, R. Audigier, B. Luvison. Classifying All Interacting Pairs in a Single Shot. In *Proceedings of the IEEE Winter Conference on Application of Computer Vision*, 2020.

<b>Niveau demandé :</b>	Ingénieur, Master 2
Ce stage ouvre la possibilité de poursuite en thèse et ingénieur R&D dans notre laboratoire.	
<b>Durée :</b>	6 mois
<b>Rémunération :</b>	entre 700 € et 1300 € suivant la formation.
<b>Compétences requises :</b>	
<ul style="list-style-type: none"> <li>- Vision par ordinateur</li> <li>- Apprentissage automatique (deep learning)</li> <li>- Reconnaissance de formes</li> <li>- Python, C/C++</li> <li>- Maîtrise d'un framework d'apprentissage profond (en particulier Tensorflow ou PyTorch)</li> </ul>	