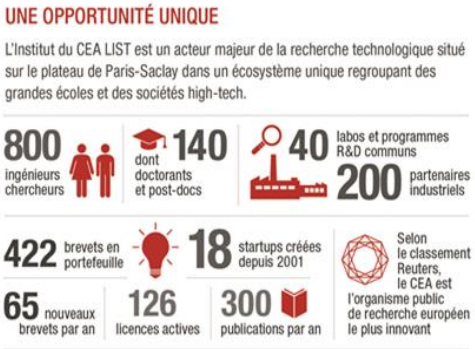
Proposition de stages 2019

L'équipe Perception 3D & Mobilité du Laboratoire de Vision et Ingénierie des Contenus du CEA LIST propose 8 sujets de stage

CEA Tech LIST

Les activités de recherche du CEA Tech LIST sont centrées sur les systèmes à logiciel prépondérant. Ces activités s'articulent autour de trois thématiques: les Systèmes Embarqués, les Systèmes Interactifs et les Capteurs et traitement du signal.

Le CEA Tech LIST collabore avec de grands acteurs industriels du nucléaire, de l'automobile, de l'aéronautique, de la défense et du médical pour étudier et développer des solutions innovantes adaptées à leurs besoins. Il réalise une recherche qui va du concept de système jusqu'au démonstrateur, contribuant au transfert de technologies et à l'innovation par l'émergence de nouvelles entreprises.



Laboratoire de Vision et de l'Ingénierie des Contenus (LVIC)

Le *Laboratoire de Vision et de l'Ingénierie des Contenus* (LVIC) du CEA Tech LIST mène des recherches dans les domaines de la vision par ordinateur et de l'analyse automatique de texte, et se compose de 3 équipes:

- **Perception 3D & Mobilité** *(Réalité Augmentée, Localisation, Vision pour la robotique)*
- Analyse de scènes par vision *(Détection/Suivi/Classification, Analyse d'activité)*
- Multimédia *(Reconnaissance d'images large échelle, Analyse conjointe texte/image, Multilinguisme, Filtrage, Extraction, Moteur de recherche)*

Opportunités

Les stages ouvrent des opportunités de poursuite en thèse ou en CDD dans notre laboratoire. Le LVIC entretient également des liens étroits avec de nombreux industriels, de la start-up aux plus grands groupes, qui n'hésitent pas à recruter les meilleurs profils que nous leurs recommandons.

Les partenaires industriels du LVIC :

Informations générales

Formation/Niveau d'études	Ingénieur, Master 2 / Bac+5
Durée	6 mois
Lieu	Palaiseau (91) – Centre d'intégration de Nano-INNOV
Indemnités de stage	Entre 700 € et 1300 € suivant formation. Aide au logement / transport / restauration.
Candidatures	Joindre CV+Lettre de motivation au contact indiqué avec la référence du stage Ne pas hésiter à préciser vos expériences de projets Préciser les dates de début et fin de stage envisagées

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

1- Localisation 3D d'objet par combinaison d'approches apprentissages (Deep Learning) et géométriques (SLAM)

Application à la Réalité Augmentée

Contact: boris.meden@cea.fr
Référence: *st3D-1*

Contexte du stage

Au cours des dernières années, les approches de localisation géométrique de type SLAM (Simultaneous Localization and Mapping) et les approches de localisation basées apprentissage profond (Deep Learning) ont connu de grands succès. Il s'agit d'approches fondamentalement différentes, chacune offrant son lot d'avantages et d'inconvénients.

L'objectif de ce stage sera de combiner ces deux approches au sein d'un algorithme de localisation d'objets 3D, ceci afin de tirer profit des avantages de ces deux approches. Pour y parvenir, l'étudiant se basera sur l'algorithme de SLAM contraint [1] ainsi que de l'algorithme Deep Manta [2], tous deux développés au laboratoire LVIC du CEA-LIST. L'étudiant sera amené à étudier les deux approches afin de les adapter en vue de leur collaboration. La fiabilité de la méthode obtenue et la facilité de déploiement seront au cœur des préoccupations et des choix réalisés.

Un démonstrateur de Réalité Augmentée sera mis en place pour illustrer les résultats de localisation obtenus.

Objectifs

Prise en main de l'algorithme de détection d'objets par Deep Learning et application sur un objet de référence.

1. Intégration de la localisation Deep Learning comme contrainte dans le SLAM
2. Adaptation de la méthode Deep Learning et/ou SLAM afin d'améliorer la fiabilité et la facilité de déploiement
3. Réalisation d'un démonstrateur de Réalité Augmentée mettant en avant l'approche mixte.

Compétences requises :

Le candidat devra disposer d'une bonne maitrise du langage C++, de connaissances en vision par ordinateur, et idéalement de pratique en apprentissage profond.

Compétences développées au cours du stage :

Ce stage permettra à l'étudiant d'acquérir des compétences sur deux technologies connaissant un essor important aujourd'hui : le SLAM et le Deep Learning. Ces compétences seront aisément valorisables à l'issu du stage puisque ces techniques sont au cœur des révolutions technologiques que sont la Réalité Augmentée (Microsoft Hololens, Google ARCore, Apple ARKit,…) et la Robotique Autonome (véhicules autonomes de Google, Uber,…).

[1] Real-Time Accurate Localization in a Partially Known Environment:Application to Augmented Reality on textureless 3D Objects, M. Tamaazousti, V. Gay-Bellile, S. Naudet Collette, S. Bourgeois, M. Dhome, ISMAR 2011
[2] Deep MANTA: A Coarse-to-fine Many-Task Network for joint 2D and 3D vehicle analysis from monocular image, Florian Chabot, Mohamed Chaouch, Jaonary Rabarisoa, Céline Teulière, Thierry Chateau. CVPR 2017

2- 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-2*

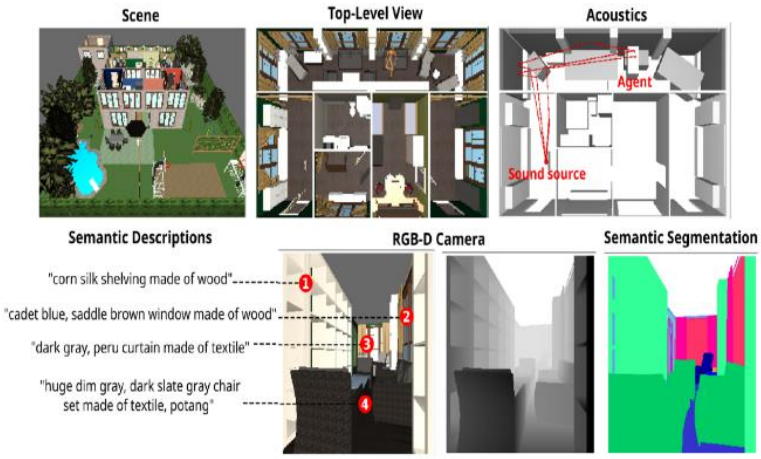
Contexte du stage

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 :

- 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

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

3 – Suivi d'objets médicaux pour la réalité virtuelle

Contact: romain.dupont@cea.fr
Référence: *st3D-3*

Sujet du stage

En collaboration avec la Faculté de Médecine Paris Sud, le laboratoire LVIC du CEA LIST développe un système de réalité mixte pour former le corps médical aux gestes d'urgence en néo-natalité. Dans cette formation, l'apprenant est équipé d'un casque de réalité virtuelle et évolue dans une salle constituée d'instruments médicaux et mannequins. Ce système de réalité mixte doit offrir un affichage à la fois virtuel et conforme à la réalité, de sorte que les interactions tangibles entre les éléments affichés dans le casque et les instruments médicaux réels manipulés par l'apprenant soient naturelles.
Pour une expérience de réalité mixte réussie, il faut que les instruments médicaux soient précisément localisés en temps-réel dans la salle. Or ceux-ci sont de petites tailles et rendent difficile leur localisation par des méthodes traditionnelles tels que les marqueurs (Aruco par exemple) ou ondes électromagnétiques par exemple. Durant ce stage, l'étudiant devra fabriquer des instruments médicaux adaptés et réalistes par impression 3D de sorte qu'on puisse y apposer des marqueurs sans gêner leur manipulation, améliorer l'implémentation du suivi de marqueurs (précision, robustesse et vitesse) en les fusionnant avec des méthodes bas niveau à faible latence (suivi de points d'intérêt/segments/cercles par exemple [1]). Des essais grandeur nature seront menés tout le long du stage avec un casque HTC Vive.

[1] Online Object Tracking: A Benchmark. Yi Wu et al.

Mots clés : suivi d'objets par marqueur, réalité virtuelle et mixte, optimisation C++, impression 3D, contexte médical

Compétences

Le candidat devra disposer d'une bonne maitrise du langage C++, d'un logiciel de modélisation 3D (Blender par exemple), des notions en réalité virtuelle. Des connaissances en vision par ordinateur sont un (vrai) plus.

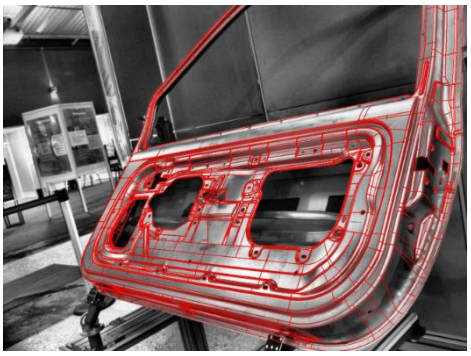
4- Localisation 3D multi-objets pour la Réalité Augmentée

Contact: steve.bourgeois@cea.fr
Référence: *st3D-4*

Contexte du stage:

Aujourd'hui, le CEA List dispose d'une technologie de localisation par vision d'objets 3D, temps réel et sans marqueurs, exploitée dans l'industrie pour des applications de Réalité Augmentée. Cette solution offre à la fois une robustesse élevée et tout en permettant de localiser de façon précise la caméra réelle dans le repère du modèle 3D de l'objet suivi.

Cependant, cette solution se limite actuellement à la localisation d'un seul objet à la fois. Une perspective d'évolution naturelle est donc d'étendre cette méthode au suivi multi-objets.



Exemple de recalage du modèle CAO sur l'objet réel

Objectifs:

L'objectif de ce stage sera donc d'étendre la méthode de localisation actuelle au suivi multi-objets. Les travaux menés se dérouleront selon deux principales étapes :

- La localisation multi-objets mobiles : il s'agira de développer une solution permettant de suivre plusieurs objets se déplaçant librement les uns aux autres. Une difficulté majeure sera d'éviter l'explosion des temps de calcul. Une mutualisation entre objets des étapes de traitement sera donc à mettre en place.
- La localisation multi-objets statiques : il s'agira de développer une solution permettant de suivre plusieurs objets statiques les uns par rapports aux autres, mais dont la position orientation relative des uns par rapport aux autres est inconnue. En plus du suivie, la méthode devra estimée la pose relative des objets les uns par rapport aux autres. Cette étape nécessitera d'optimiser les trajectoires des différents objets simultanément et non plus séparément.

Les résultats obtenus seront illustrés au travers de démonstrateurs de Réalité Augmentée.

Compétences requises :

Idéalement, le candidat devra disposer de bonnes connaissances en mathématiques appliquées et en vision par ordinateur, ainsi que d'une bonne maitrise du langage C++.

Compétences développées au cours du stage :

Ce stage permettra à l'étudiant d'acquérir de solides compétences en vision par ordinateur, en particulier sur les techniques de SLAM (« Simultaneous Localization and Mapping »).
Ces techniques connaissent un essor important dans les domaines de la Réalité Augmentée (*eg. Hololens de Microsoft, ARCore de Google, ARKit d'Apple, Magic Leap One de Magic Leap,…*), de la robotique et du véhicule autonomes.

5- Fusion de capteurs dans le cadre du SLAM visuel sur périphérique mobile

Contact: richard.guillemard@cea.fr
Référence: *st3D-5*

Contexte du stage

Les algorithmes de SLAM (*Localisation et cartographie simultanée*) visuel permettent d'estimer en temps réel la trajectoire d'une (ou plusieurs) caméra(s) tout en reconstruisant sous forme d'un nuage de points 3D l'environnement observé. Cependant les approches existantes ne fonctionnent que si le déplacement de la caméra ne présente pas de mouvements brusques et si l'environnement observé est très texturé. Pour combler ces lacunes, la fusion avec les données provenant d'autres capteurs permet d'obtenir une localisation lorsque la vision est mise en défaut.

Afin de répondre à cette problématique, nous développons un algorithme de SLAM visuel couplé à d'autres capteurs. Les approches de fusion avec une centrale inertielle (IMU) connaissent un essor récent[1], mais s'agissant d'un ensemble de capteurs relatifs, l'estimation de la trajectoire dérive forcément au cours du temps. Il est alors pertinent d'utiliser des capteurs absolus comme le GPS ou le baromètre en complément[2].

Objectifs du stage

Ce stage a pour objectif d'améliorer l'algorithme de SLAM visuel couplé à une centrale inertielle déjà existant et d'y ajouter le baromètre, et ce, afin d'améliorer la robustesse de celui-ci dans des contextes défavorables à la vision.

Pour cela, il sera nécessaire d'étudier le comportement du baromètre (bruit, modèle physique, changement de pièce), d'obtenir un calibrage précis de l'ensemble IMU/baromètre et d'intégrer l'algorithme de type SLAM visuel/IMU/baromètre sur une plate-forme d'acquisition du laboratoire.

Le récent dataset ADVIO[3] sur Iphone sera particulièrement utile pour étudier la fusion de capteur caméra/IMU/baromètre.

Compétences

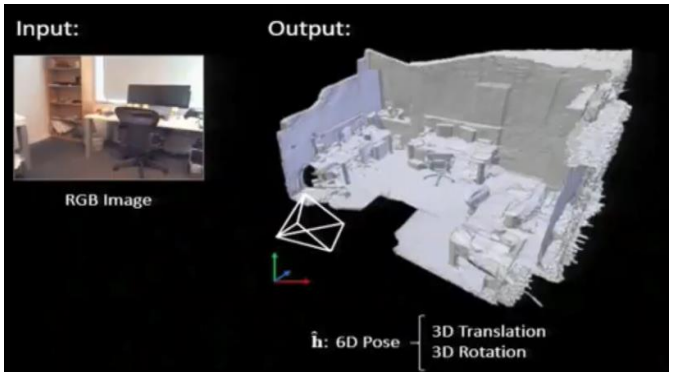
Le candidat devra disposer d'une bonne maîtrise du langage C++, de connaissances en vision par ordinateur et idéalement de compétences en fusion de capteurs (IMU, baromètre, …).

---

[1] **OrbSlam, Visual-Inertial Monocular SLAM with Map Reuse,** Raul Mur-Artal and Juan D. Tardos
[2] **Inertial Odometry on Handheld Smartphones**, Arno Solin, Santiago Cortes , Esa Rahtu, and Juho Kannala
[3] **ADVIO, An Authentic Dataset for Visual Inertial Odometry**, Arno Solin, Santiago Cortes , Esa Rahtu, and Juho Kannala

6 – Relocalisation d'une caméra par apprentissage profond

Contact: laetitia.gond@cea.fr
Référence: *st3D-6*



Contexte du stage:

La relocalisation d'une caméra, qui consiste à déterminer la pose d'une caméra dans une scène précédemment observée, est un problème fondamental de la vision par ordinateur et représente une brique essentielle dans de nombreuses applications telles que la navigation autonome de véhicules, la robotique, la réalité augmentée, etc.

Pour traiter ce problème, les approches basées sur les points d'intérêt, consistant à mettre en correspondance des descripteurs de l'image requête (SIFT, ORB, etc) avec des points 3D de la scène reconstruite (par un algorithme de type SLAM), ont longtemps dominé en raison de leur rapidité et de leur précision. Plus récemment, de nombreuses solutions basées sur l'entrainement de réseaux convolutifs ont été proposées, souvent avec une précision qui reste en dessous des méthodes classiques, mais elles laissent entrevoir la possibilité d'obtenir une plus grande robustesse aux variations des conditions d'acquisition (luminosité, changements de point de vue, éléments instables de la scène…) par le choix judicieux des données d'apprentissage et de la structure du réseau.

Objectifs:

L'objectif de ce stage est d'implémenter et tester une méthode de relocalisation se basant sur un réseau de neurone siamois entrainé à estimer la pose relative entre deux caméras observant une même scène*. L'idée est d'utiliser ce réseau comme un extracteur de caractéristiques pour représenter les images et ainsi rechercher les images similaires dans une base de données d'images localisées.
Le stage consistera à :

- Entrainer et tester un réseau de neurone estimant la pose relative entre deux images et déterminer la précision qu'il est possible d'obtenir sur cette pose relative,
- Implémenter et valider une méthode de relocalisation basée sur les caractéristiques apprises par ce réseau, et comparer ses performances avec celles d'une méthode basée sur les points d'intérêt,
- Proposer une méthode de calcul de la pose d'une image requête à relocaliser à partir des deux briques algorithmiques précédentes.

Compétences:

Le candidat devra disposer de bonnes compétences en vision par ordinateur et en apprentissage profond et d'une bonne maitrise des langages C++/Python.

*Inspirée de *Camera Relocalization by Computing Pairwise Relative Poses Using Convolutional Neural Network, Z. Laskar, I. Melekhov, S. Kalia and J. Kannala, 2017 IEEE International Conference on Computer Vision Workshop (ICCVW)*

7 – Odométrie visuelle renforcée par Deep Learning

Contact: romain.dupont@cea.fr
Référence: *st3D-7*

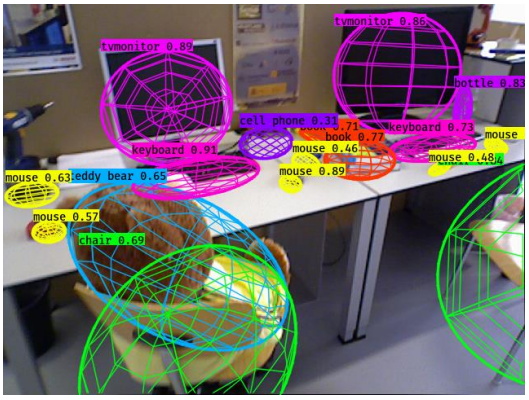
Contexte du stage

L'odométrie visuelle vise à estimer la position d'une caméra en mouvement à partir d'une succession d'images et s'appuie, entre autres, sur une détection de points d'intérêt dans chacune des images.

Le CEA LIST développe des applications de localisation *indoor* et *outdoor* en exploitant ce type d'algorithme, en particulier une méthode de SLAM (*Simultaneous Localisation And Mapping*) enrichie depuis plusieurs années. La méthode visant à être générique pour fonctionner dans des milieux très variés, les différentes scènes rencontrées peuvent comporter de nombreux éléments qui mettent en difficulté voire en échec les algorithmes classiques: surfaces réfléchissantes, vitrines transparentes, piétons ou véhicules en mouvement, feuillages, grillages, etc…

Objectif du stage

L'objectif de ce stage est de mettre en place un système basé apprentissage via un réseau de neurones profonds permettant de filtrer les images via une classification sémantique (*inspirée de* [1]), de sorte à ne fournir au SLAM que les parties pertinentes dans lesquels les points d'intérêts extraits seront valables pour assurer une localisation performante.



Compétences requises

Le candidat devra disposer de notions étendues en Apprentissage et Intelligence Artificielle, ainsi que d'une première expérience dans la mise en œuvre d'un réseau de neurones. Des connaissances en vision par ordinateur seraient appréciées.

Compétences développées au cours du stage :

Ce stage permettra à l'étudiant de renforcer ses compétences en Deep Learning et d'acquérir une connaissance approfondie de la méthode de SLAM visuel, et plus largement de la vision par ordinateur.

Mots clés : Apprentissage, Réseau de neurones profond, SLAM, Sémantique

[1] Davison et al. SemanticFusion: Dense 3D Semantic Mapping with Convolutional Neural Networks. ICRA 2017

8- Création de plans de bâtiment par système de vision

Contact: mathieu.carrier@cea.fr
Référence: *st3D-8*

Contexte du stage

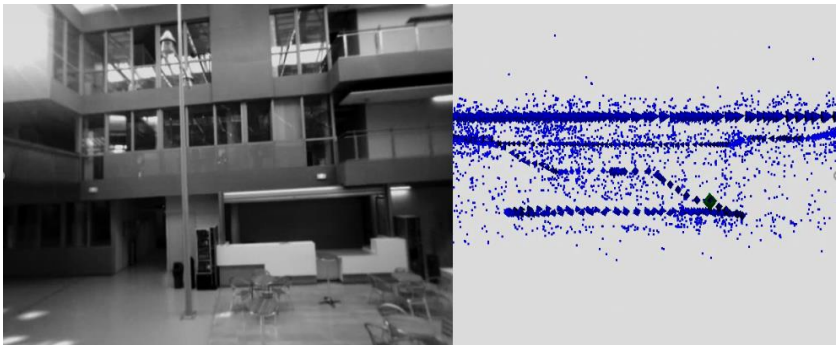
Les algorithmes de SLAM (*Localisation et cartographie simultanée*) visuel permettent d'estimer en temps réel la trajectoire d'une (ou plusieurs) caméra(s) tout en reconstruisant sous forme d'un nuage de points 3D l'environnement observé.

Ces méthodes sont utilisées dans le cadre d'applications de localisation, notamment en intérieur, dans des contextes multiples (sécurité, entretien, inspection…). Lorsque l'environnement est inconnu, notamment pour des interventions de sécurité ou de secours, la reconstitution des plans des bâtiments parcourus peut se révéler une aide précieuse pour le suivi des opérations.

Ce stage s'inscrit dans le cadre d'un projet en cours au laboratoire de vision du CEA LIST et sera mené en collaboration avec le Laboratoire de Simulation Interactive qui développe des outils de modélisation automatique de structures 3D.

Objectifs du stage

L'objectif du stage est donc de développer une solution permettant d'obtenir les plans 2D, puis 3D, d'un bâtiment avec une forte contrainte de temps de calcul et une puissance disponible limitée (calculateur embarqué). Ces plans seront reconstitués à partir des informations extraites à partir des images et algorithmes de localisation mis en œuvre.



Compétences

Le candidat devra disposer d'une bonne maîtrise du langage C++, de connaissances en vision par ordinateur et idéalement de compétences en fusion de capteurs (Caméras, IMU, …).

Mots-clés : SLAM, modélisation, 2D, 3D, localisation, caméra.