

# Sujet de thèse en Informatique / Biomécanique - Vision par ordinateur

## Titre :

Reconnaissance de comportements humains par apprentissage profond hybride : application au cas des passagers d'un véhicule autonome

## Contexte :

Le projet, nommé AutoBehave, est un projet précurseur sur l'étude des comportements des personnes dans les véhicules autonomes (VAs). Alors que la plupart des recherches menées au sujet des VAs se concentrent sur le guidage automatique des VAs et les acteurs aux alentours du véhicule, nous nous intéressons aux comportements à l'intérieur du véhicule : des passagers, leurs interactions et leurs conséquences pour les milieux urbains.

## Description du sujet :

Cette thèse, à la frontière entre l'informatique et la biomécanique, a pour objectifs principaux d'étudier et de développer de nouvelles techniques de vision par ordinateur et d'intelligence artificielle appliquées à la reconnaissance automatique des comportements humains, dans des milieux confinés comme l'habitacle des véhicules, à différentes échelles dans l'espace et dans le temps.

Dans premier temps, cette thèse étudiera l'intégration d'informations *a priori* dans les systèmes de reconnaissance, provenant notamment des modèles biomécaniques du corps humain. Ces éléments pourront être pris en compte comme de connaissances préalables dans un modèle de réseaux des neurones profonds par exemple. L'objectif est ici de réduire le volume de données nécessaire pour entraîner ce type de modèle en profitant des contraintes biomécaniques connues pour évaluer/corriger la prédiction des réseaux de neurones concernant les positions admissibles pour les parties du corps humain (Sabour *et al.*, 2017). Il s'agit de contrecarrer l'un des plus gros problèmes de ce type d'outil qui, quoique très performant dans un grand nombre de contextes, nécessite en principe une base de données d'apprentissage avec des milliers d'exemples (Krizhevsky *et al.*, 2012).

Dans un deuxième temps, cette thèse s'attaquera aux méthodes permettant la modélisation des dépendances spatio-temporelles et sémantiques des comportements humains. Les méthodes existantes pour la modélisation de dépendances temporelles sont bien adaptées pour le traitement du texte et le langage parlé (modèle de Markov caché, N. Grams, réseaux des neurones récurrents). Cependant, elles ont une performance beaucoup plus faible pour les comportements humains, particulièrement pour ceux qui contiennent des sous-éléments séparés dans le temps de façon non uniforme. De plus, ces méthodes présentent des difficultés avec des vidéos d'une durée plus élevée que quelques secondes et qui contiennent plusieurs comportements, comme démontré par Crispim-Junior *et al.*, 2016b. Également, les méthodes actuelles ne profitent pas de la sémantique des comportements pour apprendre des relations d'appartenances pertinentes entre les sous-éléments d'un comportement (objets, postures, interactions entre personnes) ou pour différencier des comportements qui sont visuellement similaires, mais sémantiquement différents. Nous voulons concevoir des modèles logiques et probabilistes (Crispim-Junior *et al.*, 2016a, Crispim-Junior *et al.*, 2017) plus adaptés à ces conditions, qui correspondent plus aux vidéos de la vie réelle qu'aux vidéos Youtube qui sont actuellement utilisées dans les bases de données de référence de la communauté.

## Profil recherché :

Nous recherchons un.e candidat.e motivé.e ayant des bases solides en informatique et mathématique, avec une expérience en traitement d'images, vision par ordinateur, apprentissage. Étant donné que le projet s'attaque à un problème qui n'a pas encore été exploré, le doctorant aura la possibilité de participer à de grandes conférences dans le domaine de la vision par ordinateur et intelligence artificielle, comme ICCV, ECCV, CVPR, ECAI et IJCAI. Ces participations lui donneront la possibilité d'interagir avec des acteurs majeurs de l'industrie et de tisser des liens professionnels qui pourront être utilisés comme point de départ pour de futures opportunités de travail.

**Compétences attendues :**

Nous recherchons un.e candidat.e connaissant :

- Langage C++ et Python
- Bibliothèque OpenCV / Qt
- Outil de versioning (GIT)

Il devra également maîtriser le SE Linux et le langage BASH

La connaissance des éléments suivants serait un plus :

- Framework PyTorch et TensorFlow.
- Framework Python Anaconda
- Maîtrise de langue anglaise.

**Salaire :**

1 842 euros brut/mois

**Lieu de travail :**

LIRIS - Université Lumière Lyon 2 (Bron) / LBMC - IFSTTAR (Bron)

**Contact :**

Carlos Crispim, carlos.crispimjunior@univ-lyon2.fr, +33(0) 4 78 77 31 15

---

## **PhD thesis in Computer Science / Biomechanics – Computer Vision**

**Title:** Recognition of human behaviors by hybrid deep learning methods: an application to passengers of autonomous vehicles

**Context:**

The project AutoBehave is a pioneering project on the study of behaviors of people in an autonomous vehicle (VA). Most VA research focuses on the automatic guidance of VAs and people around the vehicle. We are interested in the behaviors inside the vehicle: passengers, their interactions and the consequences for the environment.

**Topic:**

This thesis sits at the intersection of Computer Science and Biomechanics. It has as main objectives to study and to develop new techniques of Computer Vision and Artificial Intelligence applied to the automatic recognition of human behaviors in confined environments, like the interior of vehicles, at different scales in space and time.

In the beginning, we will study the integration of external knowledge into recognition systems, more specifically, information about biomechanical models of the human body. These elements would be considered as prior knowledge for a deep neural network, for instance. Although very powerful in many contexts, this type of model requires in principle a database with thousands of examples (Krizhevsky and al., 2012). The aim here is to reduce the amount of data needed to train this type of model by taking advantage of known biomechanical constraints to evaluate/correct the prediction of neural networks concerning the admissible positions for human body parts (Sabour et al. 2017).

In a second step, we will study methods to model spatiotemporal and semantic dependencies present in human behaviors. Existing methods that address temporal dependencies are well suited for text processing and spoken language (hidden Markov model, N. Grams, recurrent neural networks). However, they have a much lower performance on human behaviors, especially for those behavioral classes that contain sub-elements separated over time in a non-uniform fashion. In addition, such methods cannot properly handle videos lasting longer than a few seconds and/or that contain several behaviors (Crispim-Junior et al., 2016b). Finally, current methods do not take advantage of the semantics of behaviors to learn relevant relationships between sub-elements of a behavior (objects, postures, interactions between people) or to distinguish behaviors that are visually similar, but semantically different. We seek to design logical and probabilistic models (Crispim-Junior et al., 2016a, Crispim-Junior et al., 2017) more adapted to these conditions which reflect more accurately real-life videos than Youtube videos, that are currently used in public databases of computer vision community.

**Candidate Profile:**

We are looking for a motivated candidate with a solid foundation in computer science and mathematics, with experience in image processing, computer vision, and deep learning. As the project tackles a problem that has not yet been explored, the Ph.D. student will have the opportunity to participate in major conferences in the field of computer vision and artificial intelligence, such as ICCV, ECCV, CVPR, ECAI, and IJCAI. These participations will give him the opportunity to interact with major players in the industry and to build professional relationships that can be used as a starting point for future work opportunities.

**Skill Requirements:**

We are looking for a candidate that knows :

- C++ and Python languages
- OpenCV / Qt libraries
- Code versioning tools (GIT)
- LINUX OS
- BASH language

The following skills will be also appreciated:

- PyTorch et TensorFlow frameworks
- Python Anaconda framework
- Mastering of English language.

**Salary:**

1 842 euros gross/month

**Workplace:**

LIRIS - Université Lumière Lyon 2 (Bron) / LBMC - IFSTTAR (Bron)

**Contact:**

Carlos Crispim, carlos.crispimjunior@univ-lyon2.fr, +33(0) 4 78 77 31 15

---

**Références bibliographiques/ Bibliographic references**

Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks, In Proceedings of Advances in Neural Information Processing Systems 25, pp. 1097--1105, 2012.

Sabour, S.; Frosst, N.; Hinton, G.E. Dynamic Routing Between Capsules. CoRR, URL : <http://arxiv.org/abs/1710.09829>, 2017.

Crispim-Junior, C.; Avgerinakis, K.; Buso, V.; Meditskos, G.; Briassouli, A.; Benois-Pineau, J.; Kompatsiaris, Y.; Bremond, F. Semantic Event Fusion of Different Visual Modality Concepts for Activity Recognition. Transactions on Pattern Analysis and Machine Intelligence, vol.38(8), pp.1598 -1611, August, 2016a

Crispim-Junior, C.F. ; Vlasselaer, J. ; Dries, A.; Bremond, F. BEHAVE - Behavioral Analysis of Visual Events for Assisted Living Scenarios. Oral presentation @ Assistive Computer Vision and Robotics Workshop. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017, pp. 1347-1353.

Crispim-Junior, C.F.; Koperski, M.; Cosar, S.; Bremond, F. Semi-supervised understanding of complex activities from temporal concepts. In Proceeding of 13th IEEE International Conference on Advanced Video and Signal-Based Surveillance 2016 (AVSS 2016), August 23-26, 2016b, Colorado Springs – Colorado - United States of America.