

**LIMOS – Axe SIC**

**Directeur de thèse : Vincent Barra ([vincent.barra@uca.fr](mailto:vincent.barra@uca.fr))**

**Collaboration : Freddy Maso, Directeur de la performance et de l'innovation, ASM Omnisports**

**Titre du sujet de thèse : Analyse comportementale de vidéos par modèles de langage multimodaux**

**Résumé du sujet de thèse :**

L'analyse vidéo sportive basée sur l'IA exploite l'apprentissage automatique et la vision par ordinateur pour transformer des séquences brutes en informations exploitables pour les athlètes, les entraîneurs et les diffuseurs.

L'objectif de cette thèse est de développer des méthodes innovantes d'analyse comportementale des sportifs de haut niveau à partir de vidéos enregistrées en conditions réelles. Les pistes de recherche envisagées concernent le développement de modèles :

- d'analyse comportementale basés sur des vidéos, en présence de contraintes inhérentes aux conditions de prise et aux spécificités du sport étudié,
- de langage multimodaux, adaptés au contexte de l'étude, et plus particulièrement d'un modèle de fondation qui pourrait être ensuite décliné par ajustement (finetuning).

Ce sujet se propose de traiter quatre problématiques, permettant de lever deux verrous scientifiques et de répondre à deux questions applicatives :

- 1- Comment ajuster un modèle de fondation vidéo (ViFM) à la problématique spécifique. Si certains modèles existent (e.g. ActivityNet [1]), ils ne peuvent en l'état répondre aux enjeux et contraintes du domaine abordé.
- 2- Comment construire un modèle de langage multimodal (MLM) permettant d'interroger efficacement des séquences vidéo à partir de prompts ? Aujourd'hui, la tâche de fouille (retrieval) dans les vidéos (e.g. T2V [2]) n'utilise pas, à notre connaissance, de MLM, et n'exploite donc pas le potentiel de ces algorithmes ayant prouvé leur efficacité et leur performance dans de nombreux domaines.
- 3- Comment ces modèles peuvent aider à automatiser l'analyse du jeu produit par les deux équipes ?
- 4- Comment au travers d'indicateurs, ces modèles peuvent analyser la performance des joueurs dans des optiques de progression et de recrutement

Les modèles développés dans ce travail s'alimenteront de nombreuses données fournies par le partenaire, pour lesquelles la vérité terrain est disponible. L'expertise applicative du partenaire permettra de plus de valider, a posteriori, les résultats des algorithmes produits.

Si la cible applicative principale est l'étude de vidéos sportives, les algorithmes produits pendant cette thèse pourront être déclinés par ajustement sur de nombreuses autres problématiques (détection d'attaques dans des DAB, Détection de chutes, ...).

[1] F. Caba Heilbron, V. Escorcia, B. Ghanem, and J. Carlos Niebles. Activitynet: A large-scale video benchmark for human activity understanding. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2015.

[2] Q. Ye, G. Xu, M. Yan, H. Xu, Q. Qian, J. Zhang, and F. Huang. Hitea: Hierarchical temporal-aware video-language pre-training. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pages 15405–15416, October 2023