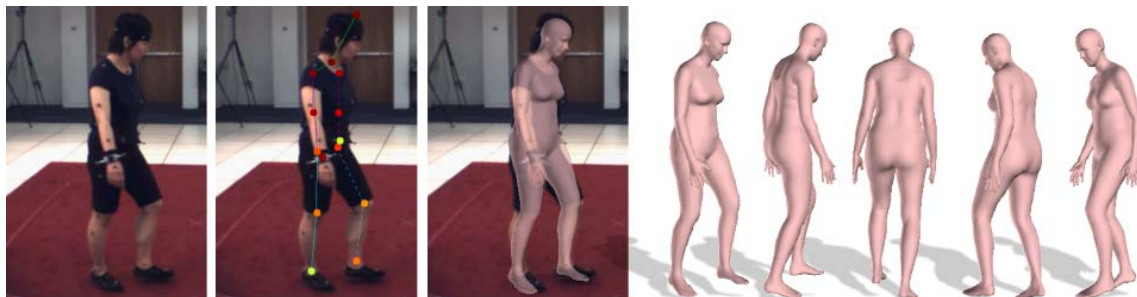


Sujet de stage Master 2 :  
**Reconstruction de squelette 3D de mouvement quotidien à partir de vidéo monoculaire**



*Figure 1: Une méthode basée sur un modèle peut contraindre efficacement l'espace de solution pour la reconstruction monoculaire du corps humain [1].*

### Organisme d'accueil

- Laboratoire ICube: Le laboratoire des sciences de l'ingénieur, de l'informatique et de l'imagerie (The Engineering science, computer science and imaging laboratory), <http://icube.unistra.fr/>
- CNRS, Délégation Alsace, France, <https://www.alsace.cnrs.fr/>

### Lieu de travail et salaire

Place de l'hôpital, Strasbourg (67), France. Le stage se déroulera dans l'équipe de recherche MLMS (Machine Learning, Modélisation & Simulation) située sur le site hospitalier du laboratoire, à 10 min à pied du cœur du centre-ville de Strasbourg, classé au patrimoine mondial de l'UNESCO.

Salaire : 600€/mois environ pour une durée de 6 mois.

### Encadrants

- [Hyewon Seo](mailto:seo@unistra.fr) (ICube, Univ. Strasbourg), [seo@unistra.fr](mailto:seo@unistra.fr)
- [Diwei Wang](mailto:d.wang@unistra.fr) (ICube, Univ. Strasbourg), [d.wang@unistra.fr](mailto:d.wang@unistra.fr)

### Date de début

Janvier – mars 2022.

### Contexte

L'analyse de mouvement quotidien est fréquemment utilisée dans plusieurs applications, telle que la détection des anomalies, et la reconnaissance des actions/intentions. Les évaluations reposant sur des capteurs portables sont coûteuses et parfois intrusives, alors que les systèmes commerciaux d'analyse de mouvement 3D nécessitent des caméras soigneusement calibrées pour collecter des données vidéo multi-vues, et ne sont donc pas pratiques. Cependant, compte tenu de la complicité dans les activités quotidiennes et de la variabilité des conditions de capture vidéo (lumière de la pièce, vue de la caméra, etc.), les

représentations directes à partir d'images 2D (silhouette, points d'intérêt spatio-temporels) peuvent ne pas être suffisantes. D'où l'intérêt de reconstruire le modèle humain 3D à partir de vidéos monoculaires.

Les approches d'apprentissage basées sur des modèles (e.g. SMPL [2]) sont de bonnes alternatives à ce problème, comme l'ont confirmé un nombre considérable de travaux antérieurs sur la reconnaissance du visage et du corps humain à partir d'images basées sur des modèles 3D préconstruits (Figure 1).

Dans ce projet, nous visons à atteindre la performance de l'état de l'art dans la reconstruction humaine 3D basée sur la vision. En particulier, nous nous intéressons aux mouvements quotidiens dont toutes les vidéos sont captées par une caméra fixe et des occlusions ou troncatures importantes du corps humain sont présents [3][4].

## Objectifs

L'objectif est de développer des modèles de réseau de neurones profonds (modèle de reconstruction ci-après) avec une masse de données de vidéos 2D, pour lesquelles les informations sur les squelettes 3D sont disponibles (fortement supervisé). Ces vidéos que nous traiterons contiennent déjà des positions articulaires 2D estimées par OpenPose [5].

Les travaux s'articulent autour des tâches suivantes :

1. Raffiner/compléter les positions articulaires 2D estimées sur les vidéos, si nécessaire.
2. Collectionner les données 2D (vidéo monoculaire) -- 3D (position des marqueurs) en pair, à partir de bases de données publiques.
3. Contributions pour améliorer/comparer des modèles de reconstruction : Calcule de paramètres de caractéristique, encodeur spatio-temporelle, par exemple.
4. Expérimentes (Ablation study, analyses comparatives), avec l'optimisation de hyperparameters
5. Construire une base de données du squelette 3D en utilisant le modèle de reconstruction entraîné.

## Compétence attendue

- Master en informatique, apprentissage automatique ou traitement du signal
- Expérience de la programmation en Python et Pytorch/Tensorflow
- Bonne notion de la modélisation cinématique est un plus
- Bonnes compétences en communication

## Application

Envoyez votre CV et vos relevés de notes (licence et master) aux encadrants.

## Bibliographie

- [1] F. Bogo, A. Kanazawa, C. Lassner, P. Gehler, J. Romero, and M. J. Black. Keep it SMPL: Automatic estimation of 3D human pose and shape from a single image. In European Conf. on Computer Vision. Springer International Publishing, 2016.
- [2] Loper, Matthew, et al. SMPL: A skinned multi-person linear model. ACM transactions on graphics (TOG) 34.6 (2015): 1-16.

- [3] K. Muhammed, N. Athanasiou, and MJ. Black. Vibe: Video inference for human body pose and shape estimation. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020.
- [4] K. Muhammed, et al. PARE: Part attention regressor for 3D human body estimation. Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021.
- [5] Cao, Zhe, et al. Realtime multi-person 2d pose estimation using part affinity fields, Proc. IEEE conference on computer vision and pattern recognition, 2017.