



Université Rennes 2 – EUR Digisport

Institut de Recherche en Informatique et Systèmes Aléatoires

**Annotation Automatique Markerless de la Performance au 110 Mètres Haies:
Vers Une Approche Fondée sur l'Estimation de Pose Humaine**

**Master Sciences du Numérique et Sport
Mémoire de 2ème année**

Année universitaire 2024-2025

Présenté par : Raphaël Peus

Encadré par : Laurent Guillo

Tuteur Universitaire : Françoise Rannou-Bekono



Remerciements :

Je tiens à remercier chaleureusement l'ensemble du laboratoire IRISA pour son accueil, ainsi que l'équipe COMBO pour son accompagnement tout au long de ce stage.

J'exprime également ma profonde gratitude à mon tuteur de stage, M. Laurent Guillo, pour sa disponibilité, ses conseils avisés et son expertise en intelligence artificielle, en particulier dans le domaine de l'estimation de pose humaine, qui m'ont été précieux pour la réalisation de ce mémoire.

Je remercie aussi la Fédération Française d'Athlétisme pour nous avoir permis d'échanger avec des entraîneurs – une expérience particulièrement enrichissante – et pour la mise à disposition des vidéos indispensables au développement de notre méthodologie.

Enfin, je remercie tous les enseignants de l'EUR DIGISPORT pour la qualité de leur enseignement et leur engagement à transmettre leur savoir.

Table des matières

| | |
|---|----|
| Table des matières | 2 |
| 1. Introduction..... | 3 |
| 2. Revue de littérature | 5 |
| 2.1 Définition et principes de l'estimation de la pose humaine..... | 5 |
| 2.2 Approches et méthodes d'estimation de la pose humaine | 6 |
| 2.2.1 DeepPose..... | 6 |
| 2.2.2 DeeperCut | 7 |
| 2.2.3 OpenPose | 7 |
| 2.2.4 MediaPipe..... | 7 |
| 2.2.5 YoloPose | 7 |
| 2.2.6 RTMPose | 8 |
| 2.2.7 VitPose..... | 8 |
| 2.3 Datasets | 9 |
| 2.4 Domaines d'applications | 10 |
| 2.4.1 Applications générales des HPE..... | 10 |
| 2.4.2 Intégration des HPE dans l'analyse et l'optimisation de la performance sportive | 11 |
| 2.5 Limites | 12 |
| 3. Problématique..... | 13 |
| 4. Méthodologie | 14 |
| 4.1 Sélection de l'HPE candidate | 14 |
| 4.2 Nettoyage des données..... | 16 |
| 4.3 Collecter les trajectoires des points d'intérêts pertinents | 17 |
| 4.5 Calcul/Sauvegarde des métriques..... | 22 |
| 4.6 Protocole expérimental..... | 22 |
| 5. Résultats | 25 |
| 5.1 Comparaison des métriques..... | 25 |
| 5.1.1 Résultat pour la vidéo 1..... | 25 |
| 5.1.2 Résultat pour la vidéo 2..... | 26 |
| 5.2 Temps de traitement..... | 27 |
| 6. Discussion | 28 |
| 7. Conclusion | 30 |
| 8. Bibliographie..... | 31 |
| 9. Annexes | 35 |
| 10. Résumé - Abstract | 38 |

1. Introduction

Dans le domaine du sport de haut niveau, l'analyse vidéo occupe une place centrale. Comprendre les points forts et les faiblesses d'un athlète, ainsi que la façon dont il se déplace, franchit un obstacle ou réalise une action technique, permet d'optimiser sa performance et d'affiner les stratégies d'entraînement. C'est pourquoi, depuis plusieurs années, la Fédération Française d'Athlétisme (FFA) s'attache à recueillir des métriques à partir de vidéos sur les disciplines techniques, notamment le 110 mètres haies, une épreuve qui demande une maîtrise fine du rythme de course et une coordination parfaite entre les foulées et les franchissements de haies.

Pour cela, les entraîneurs et les analystes vidéo s'appuient traditionnellement sur des techniques d'annotation manuelle réalisées à partir de séquences filmées. Cette méthode consiste à identifier visuellement des événements clés de la course (comme les phases de vol, les temps de contact au sol ou les intervalles entre les haies) et à les consigner de manière structurée. Si cette pratique a fait ses preuves, elle présente toutefois des limites : elle est chronophage, car non automatisée, et soumise à une variabilité inter-observateur qui peut nuire à la fiabilité des données collectées.

Parallèlement aux méthodes traditionnelles d'analyse vidéo, les avancées en vision par ordinateur ont conduit au développement de technologies d'estimation de la pose humaine (Human Pose Estimation – HPE). Ces techniques permettent d'identifier automatiquement les différents points d'intérêts du corps humain (chevilles, genoux, hanches, etc.) à partir d'images ou de vidéos. D'abord conçues pour répondre à des problématiques générales en robotique, surveillance ou interaction homme-machine (Heindl et al., 2019), elles ont progressivement trouvé des applications dans le domaine du sport, en particulier pour l'analyse du mouvement. Dans ce contexte, elles apparaissent comme une solution prometteuse aux limites des annotations manuelles précédemment décrites.

Lors de mon précédent stage, j'ai participé au développement de Meta-vidéo, un outil d'analyse vidéo reposant sur l'annotation manuelle. Cette expérience m'a permis de comprendre l'utilité de ces analyses tout en constatant les limites évoquées. Le projet actuel vise à franchir une nouvelle étape : intégrer des techniques d'estimation de pose humaine (HPE) afin d'automatiser l'analyse du 110 mètres haies, notamment en détectant les différentes phases de la course et en extrayant les données pertinentes à partir de vidéos, dans la perspective d'une intégration future par la FFA.

Ce travail s'inscrit dans une volonté d'explorer les capacités des outils de HPE pour détecter et analyser les différentes phases de la course sur 110 mètres haies. Il vise à évaluer la précision et la fiabilité de ces technologies en les comparant aux annotations manuelles afin de mesurer leurs performances et leurs limites. Une attention particulière sera portée aux écarts entre les résultats obtenus par ces méthodes et les références établies en annotation manuelle. L'objectif final est d'identifier les adaptations nécessaires

pour optimiser l'application de ces outils dans un contexte sportif de haut niveau, afin que les informations extraites puissent être exploitables par les entraîneurs et les athlètes pour améliorer leurs performances.

Dans une première partie, ce mémoire présentera une définition de l'estimation de la pose humaine (Human Pose Estimation – HPE), en revenant sur son fonctionnement, son origine, ainsi que ses principales applications. Une attention particulière sera portée à ses usages dans le domaine sportif, à ses apports mais aussi à ses limites, en s'appuyant sur les travaux existants présentés dans la revue de littérature. Il précisera ensuite la problématique spécifique de notre étude, les objectifs poursuivis, ainsi que la méthodologie adoptée. Enfin, il exposera les résultats obtenus, avant de discuter de l'impact, les limites et les perspectives d'application pour l'optimisation de la performance sportive.

2. Revue de littérature

2.1 Définition et principes de l'estimation de la pose humaine

L'estimation de la pose humaine est un domaine clé de la vision par ordinateur et de la capture de mouvement sans marqueurs, visant à détecter et localiser les différentes parties du corps humain pour reconstruire une représentation squelettique à partir d'images ou de vidéos (Zheng et al., 2020). Cette technologie génère un pseudo squelette, composé de points clés ou keypoints (KP) représentant les articulations principales (coudes, genoux, hanches...) ainsi que d'autres repères anatomiques (tête, yeux, nez...).

L'HPE permet d'extraire les coordonnées spatiales des points d'intérêt détectés par le modèle, fournissant ainsi des données exploitables. Il existe plusieurs modèles d'HPE qui se distinguent principalement par leur architecture et par les points d'intérêt qu'ils analysent en fonction du jeu de données sur lequel ils sont entraînés (Zheng et al., 2020). Chaque modèle génère un pseudo squelette spécifique avec des repères corporels propres à chacun, ce qui signifie que chaque approche présente à la fois des avantages et des limitations en fonction du contexte d'application. Par exemple, la Figure 1 et la Figure 2 présentent des pseudo-squelettes obtenus par les HPE RTMPose (Jiang et al.2023) et YoloPose (Maji et al., 2022).



Figure 1 : Pseudo-squelette de l'auteur généré par RTMPose. (Jiang et al.,2023)

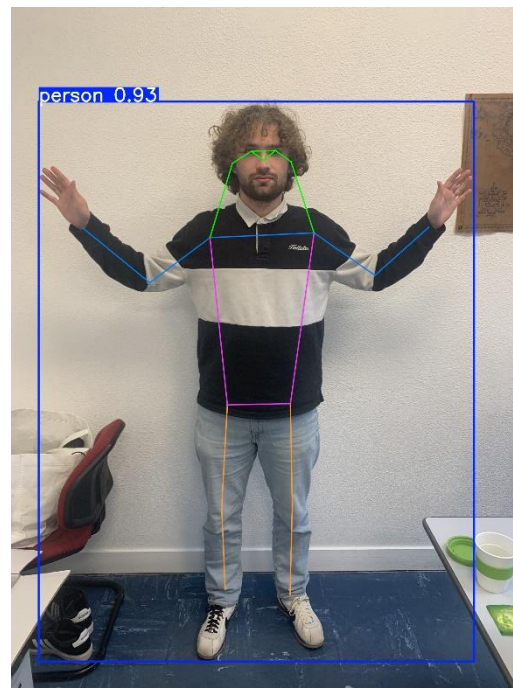


Figure 2 : Pseudo squelette de l'auteur généré par YoloPose. (Maji et al., 2022)

Parmi les nombreuses approches existantes, l'étude de Zheng et al. (2020) propose de classer les techniques d'estimation de pose humaine en deux grandes catégories : les approches 2D et les approches

3D. Les HPE 2D, utilisés lors de notre projet, consistent à détecter les points clés du corps humain dans un plan, en s'appuyant principalement sur des réseaux de neurones convolutifs (CNN). Elles peuvent être appliquées à une seule personne (single-person) ou à plusieurs individus (multi-person) présents dans une même scène. Les points d'intérêt sont généralement représentés par des coordonnées (u, v) correspondant aux positions horizontale et verticale des KP dans l'image, l'origine étant le coin supérieur gauche de l'image. Lorsqu'une vidéo est analysée image par image, la position des KP dans chaque image permet de suivre et d'analyser le mouvement au fil du temps. De leur côté, les HPE 3D cherchent à estimer la position des articulations dans l'espace en ajoutant une profondeur (z) , ce qui permet une représentation plus complète des mouvements corporels.

2.2 Approches et méthodes d'estimation de la pose humaine

Aujourd'hui, il existe de nombreux modèles d'HPE 2D qui s'appuient sur un large éventail de méthodes, allant des techniques classiques de vision par ordinateur aux approches les plus avancées de l'intelligence artificielle. Historiquement, les méthodes traditionnelles de vision par ordinateur se concentraient sur les calculs géométriques et des approches basées sur les caractéristiques pour estimer la pose d'objets ou de sujets humains (Hartley & Zisserman., 2004). Bien que ces méthodes aient fourni des informations précieuses et ouvert la voie à des avancées ultérieures, elles avaient des limites. Elles étaient souvent sensibles au bruit, aux occultations, et elles s'appuyaient fortement sur des paramètres de caméra connus. En conséquence, les approches basées sur des modèles ont commencé à gagner du terrain comme les « Active Shape Model » (ASM), les « Active Appearance Models » (AAM) (Cootes et al, 2001) ou les « Articulated model » (Eichner et al, 2012). Les approches basées sur des modèles utilisent des modèles prédéfinis d'objets ou de parties du corps pour estimer leurs poses. Plus récemment, au milieu des années 2010, les avancées majeures en deep learning ont permis de développer des modèles puissants d'HPE, capables de traiter des contextes variés avec une précision et une robustesse accrue. Ces approches modernes sont conçues pour répondre aux contraintes spécifiques de chaque application, en cherchant un équilibre optimal entre performance, vitesse d'exécution et complexité. Dans cette section, nous présenterons les principaux modèles étudiés dans le cadre de ce projet, en détaillant leur fonctionnement, ainsi que leurs spécificités.

2.2.1 DeepPose

DeepPose, introduit par Toshev et Szegedy en 2014 (Toshev, A., & Szegedy, C. 2014), représente une avancée majeure dans le domaine de l'estimation de la pose humaine, car il s'agit du premier modèle à utiliser des réseaux de neurones convolutifs (CNN) pour cette tâche. Ce modèle adopte une approche fondée sur la régression directe des coordonnées des points clés : il apprend à prédire les positions des articulations dans l'image via un processus itératif, permettant d'affiner progressivement la localisation des KP. DeepPose est un modèle single-person, il ne peut estimer la pose que pour une seule personne à la fois. Ce type d'approche repose généralement sur deux étapes : d'abord, la détection de la personne dans l'image, puis l'estimation de sa pose à l'intérieur de la zone détectée. Cependant, comme le

soulignent Sun et al. (2017), les méthodes basées uniquement sur la régression (comme DeepPose) présentent des limites en termes de précision.

2.2.2 DeeperCut

Pour répondre au besoin de traiter des scènes comprenant plusieurs individus, des modèles multi-person ont émergé, comme DeeperCut (Insafutdinov et al., 2016), l'un des premiers à adopter une approche bottom-up. Celle-ci repose sur la détection indépendante des différentes parties du corps, à partir de cartes de confiance, avant de les regrouper pour reconstruire la pose de chaque individu (Kresovic & Nguyen., 2021). Les cartes de confiance sont des cartes de chaleur dans lesquelles chaque pixel représente la probabilité de présence d'un point clé, tel qu'un coude ou un genou. Newell et al. (2016), avec leur modèle Stacked Hourglass, ont montré que ces cartes permettaient de localiser précisément les articulations. Grâce à des prédictions affinées de manière itérative (Qu et al., 2022), ces approches offrent des résultats plus robustes, même dans des environnements dynamiques. À savoir que, plus récemment, ces cartes sont souvent accompagnées de méthodes de régression pour améliorer la précision et la robustesse comme le suggère l'étude de Sun et al. (2018).

2.2.3 OpenPose

OpenPose, développé en 2017 par le Perceptual Computing Lab de la Carnegie Mellon University (Cao et al., 2017), est un HPE multi-personne qui a largement contribué à démocratiser l'estimation de la pose pour plusieurs individus, notamment dans des applications en temps réel, et il est intégré à la bibliothèque OpenCV (Culjak et al. 2012). Reprenant le principe bottom-up déjà introduit, il innove avec les Part Affinity Fields (PAFs), des champs vectoriels permettant de relier les KP en fonction de l'orientation et de la position des membres. Grâce à cette architecture, OpenPose atteint un bon compromis entre précision et vitesse, même dans des scènes complexes.

2.2.4 MediaPipe

MediaPipe Pose, développé par Google en 2019, est une solution robuste pour la détection des poses humaines « single person » en temps réel. Contrairement aux approches traditionnelles, MediaPipe utilise une architecture modulaire, divisant le système en plusieurs modules indépendants, chacun ayant une fonction spécifique. Cette modularité facilite la maintenance, l'évolution et la personnalisation du système, permettant aux développeurs de combiner différents composants pour créer des pipelines de traitement adaptés à leurs besoins spécifiques.

2.2.5 YoloPose

En 2022, une avancée significative dans le domaine de la détection de poses humaines a été réalisée avec l'introduction de YOLO-Pose (Maji et al., 2022), une innovation développée par l'équipe Ultralytics. YOLO-Pose se distingue par son approche hybride, combinant les avantages des méthodes top-down et bottom-up (Zhou et al., 2023). L'approche top-down (Nguyen & Kresovic., 2022), contrairement à l'approche bottom-up mentionnée précédemment, détecte d'abord les personnes avant de prédire leurs poses à la manière des HPE single person. YOLO-Pose, tout en évitant l'utilisation

traditionnelle des cartes de chaleur des méthodes bottom-up, permet une détection simultanée des boîtes englobantes et des poses. Fondé sur le modèle YOLOv5, il optimise ainsi le temps de traitement et élimine les étapes complexes de post-traitement des méthodes bottom-up.

2.2.6 RTMPose

RTMPose (Jiang et al., 2023) publié par le Shanghai AI Laboratory se distingue comme un modèle d'estimation de la pose humaine (HPE) de pointe, surpassant d'autres modèles grâce à une combinaison d'innovations architecturales, d'optimisations d'entraînement et de techniques d'inférence avancées. C'est un modèle qui utilise une approche top-down et qui contrairement aux approches traditionnelles qui utilisent des heatmaps ou des méthodes de régression directe, RTMPose adopte le modèle SimCC (Simultaneous Coordinate Classification) (Li et al., 2019), qui reformule la localisation des points clés comme un problème de classification. Cette approche réduit la complexité du modèle tout en améliorant la précision, permettant une localisation de précision sous-pixel sans post-traitement complexe. RTMPose combine une architecture efficace, des stratégies d'entraînement avancées et des techniques d'inférence optimisées pour offrir des performances supérieures en termes de précision et d'efficacité.

2.2.7 ViTPose

Publié en 2022, ViTPose (Xu et al., 2022) est le premier modèle d'HPE à avoir popularisé l'utilisation des « transformers ». Depuis 2014, les modèles d'HPE s'appuyaient principalement sur des réseaux de neurones convolutifs (CNN). Cependant, les CNN présentent des limites lorsqu'il s'agit de modéliser les relations globales entre les articulations, notamment dans des situations complexes où certaines parties du corps sont occultées ou en interaction étroite. C'est dans ce contexte que les transformers ont émergé comme une alternative prometteuse (Wang et al., 2023). ViTPose est ainsi le premier HPE à avoir popularisé les transformers, offrant une meilleure capacité à modéliser les relations globales entre les articulations. Toutefois, cet avantage a un coût : les transformers nécessitent des ressources computationnelles élevées, ce qui peut poser problème dans des applications en temps réel ou sur des dispositifs embarqués.

La Figure 3 illustre les principales méthodes HPE 2D, tandis que la Figure 4 retrace la chronologie des modèles présentés dans cette section.

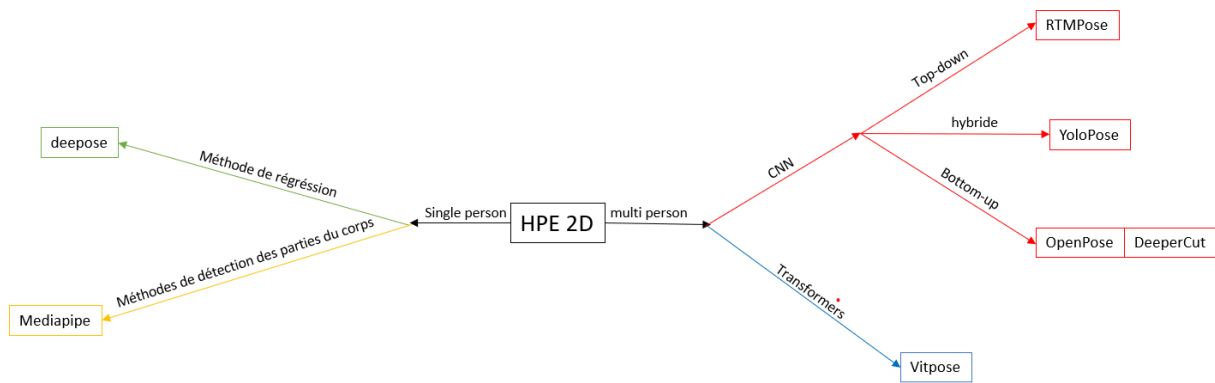


Figure 3 : Typologie des principales approches d'HPE 2D. Schéma réalisé à partir de (Cao et al., 2017), (Zheng et al., 2023), (Jiang et al., 2023), (Maji et al., 2022), (Insafutdinov et al., 2016), (Toshev et Szegedy, 2014)

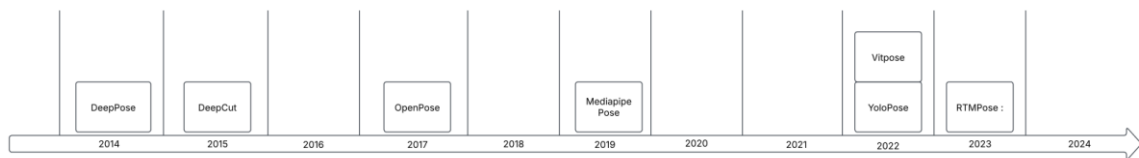


Figure 4 : Phrise chronologique des principaux modèles d'HPE 2D. Schéma réalisé à partir de (Cao et al., 2017), (Jiang et al., 2023), (Maji et al., 2022), (Insafutdinov et al., 2016), (Toshev et Szegedy, 2014)

2.3 Datasets

Les HPE sont entraînés à partir de jeux de données spécialisés contenant des annotations précises des positions des points d'intérêts corporels sur de nombreuses images. Parmi les bases de données les plus utilisées, on retrouve COCO, MPII Human Pose Dataset et Human3.6M. Ces datasets sont parmi les plus cités (cf. Figure 5). Il est important de noter qu'un même jeu de données peut exister en plusieurs versions, chacune proposant un schéma d'annotation des KP différent. Ces datasets fournissent des images ou des vidéos d'individus dans une grande variété de postures, de contextes et de conditions d'éclairage, avec des annotations détaillées pour chaque articulation visible.

L'apprentissage se fait généralement de manière supervisée : le modèle prédit la position des KP, puis cette prédiction est comparée aux annotations du dataset afin de calculer une erreur. Cette dernière permet ensuite d'ajuster les poids du réseau de neurones. Deux métriques couramment utilisées pour évaluer la précision des modèles sont le mAP (mean Average Precision) et le RMSE (Root Mean Square Error).

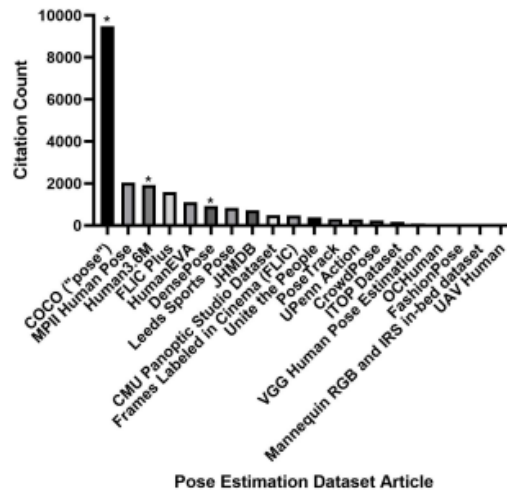


Figure 5 : Nombre de citations pour les articles présentant les jeux de données de pose les plus populaires (Lachance et al,2023)

Les annotations manuelles présentes dans ces datasets sont essentielles à l’entraînement des modèles HPE. Elles leur permettent d’apprendre à reconnaître avec précision les positions des articulations, en tenant compte des variations de posture, des contextes et des conditions d’éclairage. Toutefois, ces bases de données ne couvrent pas l’ensemble des situations humaines possibles. Par exemple, dans le cadre de notre étude, la détection des athlètes en position de départ dans les starting-blocks s’avère difficile, car ce type de posture est rarement représenté dans les datasets d’entraînement. Cette limitation illustre les défis que peuvent encore rencontrer les modèles HPE lorsqu’ils sont confrontés à des contextes atypiques ou peu représentés dans les données d’apprentissage.

2.4 Domaines d’applications

2.4.1 Applications générales des HPE

Aujourd’hui, la HPE joue un rôle clé dans de nombreux secteurs, car elle permet d’analyser, d’interpréter et d’interagir avec les mouvements humains de manière automatisée. Elle est utilisée dans l’interaction homme-machine (Heindl et al., 2019) pour rendre les interfaces plus intuitives et accessibles, notamment via la reconnaissance gestuelle.

Dans la réalité augmentée (Eric Marchand et al., 2016) et la réalité virtuelle, elle permet des expériences plus immersives en capturant précisément les gestes et déplacements des utilisateurs. Les jeux vidéo (Liu et al., 2024) exploitent cette technologie pour offrir un gameplay interactif basé sur le mouvement du joueur.

En surveillance (Ando et al., 2007) (Xu et al. 2021), la HPE est utilisée pour détecter des mouvements suspects, et donc assurer la sécurité dans les espaces publics ou surveiller des environnements à risque.

L’utilisation des HPE dans le secteur médical représente une avancée prometteuse pour le suivi automatisé des patients et le développement de solutions cliniques innovantes et intelligentes comme « ExerciseCheck » (Gu et al.2019) (Chen et al.2018).

Enfin, dans les sciences du sport (Xi et al., 2024), une HPE joue un rôle clé dans l'analyse biomécanique des athlètes en permettant une évaluation des gestes techniques, de la posture et des types de mouvement. Grâce à cette technologie, les entraîneurs et chercheurs peuvent identifier des inefficacités dans l'exécution d'un mouvement, ajuster les programmes d'entraînement et améliorer la performance des sportifs. De plus, la détection en temps réel des asymétries ou des mouvements inhabituels peut aider à prévenir les blessures en identifiant des charges excessives ou des compensations susceptibles d'engendrer des traumatismes (Difini et al.2021).

Grâce à sa capacité à modéliser précisément les mouvements humains, la HPE transforme ces domaines en offrant des solutions interactives et adaptées aux besoins spécifiques de chaque secteur.

2.4.2 Intégration des HPE dans l'analyse et l'optimisation de la performance sportive

L'estimation de la pose humaine est de plus en plus utilisée dans l'entraînement sportif pour exploiter ces technologies afin d'améliorer l'analyse de la performance. Cette technique suscite un intérêt croissant en recherche sportive en raison de son efficacité et de sa précision. Elle constitue une alternative aux systèmes traditionnels de capture de mouvement, souvent inaccessibles car nécessitant un équipement coûteux et un environnement de laboratoire. Cependant, sa validité dans les contextes sportifs et athlétiques reste encore peu étudiée (Fukushima, T et al, 2024; X. Xi et al, 2024).

L'intégration HPE dans le domaine sportif est relativement récente. Les premières applications datent d'il y a seulement quelques années, et aujourd'hui, cette technologie se retrouve dans divers sports, illustrant son adoption croissante. Que ce soit dans le football, où André Dübbell (2013) a montré les prémices de l'estimation de pose dans le sport, le tennis (Wang et al., 2024), le baseball, sport pionnier dans l'analyse (Li et al., 2021), ou encore les sports artistiques comme la gymnastique ou le patinage où l'analyse détaillée des postures est cruciale (Ballabio & Giorgia. 2023)(Wen et al.2021), les HPE sont de plus en plus utilisées pour analyser les mouvements, améliorer la performance, prévenir les blessures ou encore affiner l'évaluation technique.

En athlétisme, les HPE connaissent un essor particulier, avec plusieurs travaux récents portant sur l'annotation automatique des phases techniques dans des disciplines telles que le saut en longueur (M. Einfalt et al., 2019) (Gan et al.2024) ou le triple saut (K. Ludwig et al., 2021). Ces études visent à détecter automatiquement les événements clés comme les impulsions et les poses au cours des sauts, bien que certaines erreurs aient été relevées lors de la première phase d'analyse (cf. Figure 6). Plus récemment, dans la discipline du saut en longueur, X. Gong et al. (2023) ont exploré l'apport de modèles de pointe tels que YOLO et Openpose pour affiner l'analyse des mouvements, proposant des solutions mieux adaptées à la complexité et à la dynamique du geste sportif.

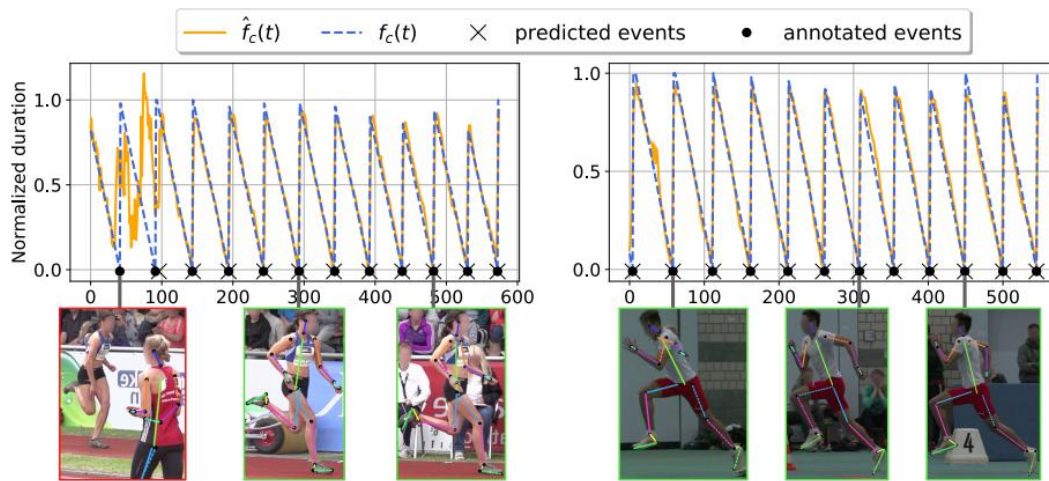


Figure 6 : Annotation automatique de la réception et de l'impulsion au saut en longueur (M. Einfalt et al., 2019)

Dans le cas spécifique du 110 mètres haies, qui constitue le cœur de notre étude, deux recherches ont été menées. L'une, réalisée par Jafarzadeh, P et al. (2021), met en évidence de bons résultats avec OpenPose tandis que l'autre (Jafarzadeh, P et al. 2025), plus récente conclut que YoloPose est particulièrement adapté à ce contexte.

Ces études montrent que l'utilisation des HPE dans le sport s'est fortement développée ces dernières années. Grâce aux progrès de l'intelligence artificielle et des modèles de vision par ordinateur, les HPE sont désormais un outil utile pour l'analyse et l'optimisation de la performance sportive.

2.5 Limites

Tous les modèles présentés jusqu'ici présentent encore de nombreuses limites. Bien que leurs performances s'améliorent continuellement, ils rencontrent encore des difficultés face à certaines occultations, ce qui peut altérer la cohérence de la pose estimée (Huang et al., 2019). Certaines postures ne sont pas correctement reconnues(cf.Datasets), car les jeux de données utilisés pour l'entraînement ne couvrent pas toutes les spécificités des mouvements propres à chaque sport.

De plus, la répartition inégale des caractéristiques (sexe, âge, couleur de peau) dans les ensembles de données d'entraînement, comme COCO, introduit des biais. Une étude récente (LaChance et al., 2023) souligne ainsi que certains modèles d'HPE affichent des écarts de performance en fonction du genre, de la couleur de peau ou de l'âge des individus, en raison de la nature des données sur lesquelles ils ont été entraînés.

En conséquence, il n'existe pas de modèle parfait. Il est donc essentiel de choisir celui qui répond le mieux aux exigences spécifiques de l'application visée, chaque modèle présentant des forces et des faiblesses selon le contexte d'utilisation.

3. Problématique

L'analyse de la performance sportive a connu une transformation significative grâce à l'essor de l'intelligence artificielle et plus particulièrement par l'intégration des HPE dans l'analyse vidéo. Plusieurs études ont démontré l'efficacité de ces technologies dans des disciplines telles que le tennis ou le saut en longueur, permettant d'extraire des informations biomécaniques, de segmenter des phases de mouvement ou encore d'automatiser certaines tâches jusqu'alors manuelles.

En revanche, les épreuves techniques comme le 110 mètres haies restent peu explorées dans la littérature, alors même que la nature rapide et répétitive des franchissements, associée à des variations posturales importantes, complique considérablement l'analyse. Ce déficit de recherches souligne la nécessité d'investiguer spécifiquement le potentiel des HPE dans ce contexte.

En France, l'usage de la vidéo dans l'analyse de performances en athlétisme demeure encore peu développé et repose essentiellement sur des outils manuels. C'est le cas notamment à la FFA, qui utilise actuellement des logiciels comme Méta-vidéo ou Kinovea, fondés sur une annotation image par image. Cette approche présente deux limites majeures : elle est à la fois chronophage et sujette à une variabilité inter-observateur. L'introduction de systèmes automatisés fondés sur les HPE permettrait d'accélérer le processus tout en standardisant les mesures, apportant ainsi un gain d'objectivité et d'efficacité.

Dans cette optique, la FFA envisage de recourir à des technologies d'IA, notamment les HPE 2D, pour automatiser l'annotation des courses de 110 m haies. Ces solutions permettraient de détecter des événements clés tels que les franchissements, les phases d'impulsion ou de réception, tout en allégeant la charge de travail des analystes.

Dès lors, une question centrale émerge : les technologies d'estimation de la pose humaine sont-elles adaptées à l'automatisation de l'analyse des courses de 110 m haies ?

L'objectif d'une part, d'évaluer la pertinence et les limites des approches fondées sur la HPE dans un cadre aussi exigeant, et d'autre part, concevoir un outil capable d'identifier automatiquement des instants clés à partir de vidéos de compétition. Une attention particulière sera portée à la comparaison entre les résultats issus de cette détection automatique et ceux obtenus par annotation manuelle, afin de juger de la fiabilité du système. Cette application devra répondre aux attentes du terrain, en fournissant aux entraîneurs des données exploitables rapidement, voire en temps réel.

4. Méthodologie

Notre objectif méthodologique est donc de développer, à partir d'une HPE, une solution automatique capable d'identifier, à partir de vidéos fournies par la FFA, les moments d'impulsion et de réception à chaque haie, afin de calculer :

- Le temps de vol.
- Le temps de course.
- L'intervalle entre les haies.

4.1 Sélection de l'HPE candidate

Après avoir étudié la littérature sur les HPE, ses différentes technologies, ses évolutions récentes ainsi que ses applications dans l'annotation d'événements sportifs, la première étape de notre méthodologie a consisté à tester plusieurs modèles sur les vidéos fournies par la FFA. L'objectif était donc de répondre à la première problématique de notre projet. Quel est le modèle le plus pertinent pour notre étude, centrée sur l'analyse des courses de 110 mètres haies ?

Ces vidéos sur lesquelles les HPE ont été testées, peuvent être regroupées en deux grandes catégories :

- **Les vidéos single-person**, généralement issues des entraînements, où un seul athlète est filmé de près, dans un environnement peu perturbé. (Figure 7)



Figure 7 : Vidéo single-person

- **Les vidéos multi-person**, enregistrées en situation de compétition, avec plusieurs athlètes, un environnement chargé (spectateurs, juges, etc.) et une prise de vue souvent éloignée (Figure 8)

Pour les vidéos single-person, nous avons testé cinq modèles : OpenPose, YoloPose, Mediapipe Pose et RTMpose. Dans le cas des vidéos multi-person, les mêmes modèles ont été évalués, à l'exception de Mediapipe Pose, qui, comme l'indique la littérature, est limité à la détection d'une seule personne à la fois.

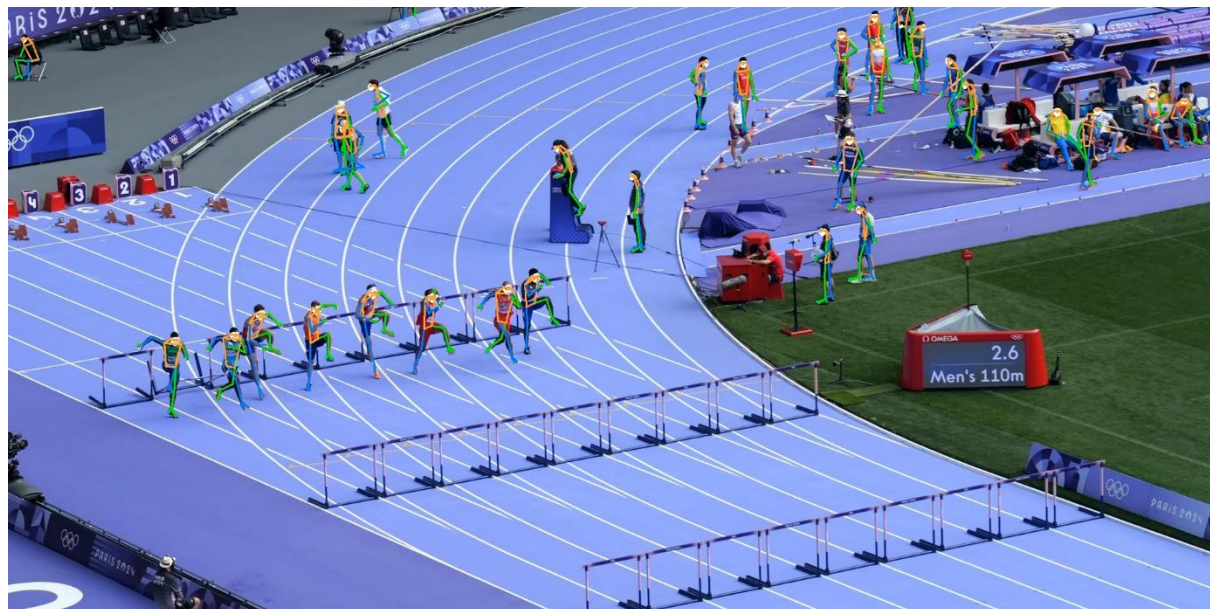


Figure 8 : Vidéo multi-person

L'objectif de ces tests était de comparer des modèles issus de différentes générations, afin d'évaluer l'évolution des performances en fonction de divers contextes. Cette phase exploratoire a rapidement mis en évidence plusieurs problématiques spécifiques liées à l'analyse des courses de haies :

- Difficultés de détection lors des franchissements de haies.
- Superpositions entre athlètes.
- Bruits parasites (spectateurs, juges, faux squelettes).
- Distance camera-athlètes.
- Positions accroupies dans les starting-blocks mal détectées.
- Robustesse face à la qualité de la vidéo.
- Temps de traitement.

Une fois ces enjeux identifiés, nous avons évalué la manière dont chaque modèle réagissait à ces contraintes. L'objectif était de déterminer les modèles les plus sensibles à ces problèmes, mais aussi ceux qui s'en affranchissaient le mieux.

Enfin, un tableau récapitulatif a été réalisé afin de synthétiser l'ensemble des observations. Il offre une vue d'ensemble des performances des différents modèles sur nos vidéos et facilite l'identification du modèle le plus adapté pour la suite de l'analyse.

| | Distance caméra/athlète | Franchissement de haies | Multi person | Single person | Resistance au flou | Starting blocks | Occlusions | Temps de traitement |
|-----------|----------------------------|----------------------------|-----------------|------------------|-----------------------|--------------------|------------|------------------------|
| OpenPose | - | = | - | + | - | -- | = | - |
| YoloPose | -- | = | - | + | -- | -- | = | + |
| RTMpose | + | = | + | + | + | = | = | + |
| Mediapipe | -- | + | ? | + | - | - | ? | ++ |

Tableau 1 : Analyse comparative des modèles d'estimation de pose humaine dans le cadre du 110 m haies

On constate sur le Tableau 1 que RTMPose se distingue par les meilleurs résultats globaux parmi les modèles testés. Il montre une bonne robustesse face à la distance, là où des modèles comme YoloPose cessent de détecter les athlètes au-delà de 10 mètres. RTMPose gère relativement bien les occultations entre les athlètes, même si certaines limitations subsistent. Il résiste mieux que les autres modèles au flou présent dans certaines vidéos et parvient à détecter plus précisément les squelettes dans les starting-blocks. Enfin, il dispose d'une interface de programmation d'application (API) et présente également un temps de traitement satisfaisant, ce qui en fait un candidat solide pour la suite de notre étude. Nous avons donc décidé de sélectionner RTMPose comme solution HPE candidate.

4.2 Nettoyage des données

Pour initier le développement de l'application d'annotation automatique à partir de RTMPose, nous avons travaillé à partir d'une vidéo simple, dans laquelle le modèle d'estimation de pose fonctionne correctement : un seul athlète est présent, les perturbations visuelles sont minimales, et le contexte est globalement propre. L'objectif final est de construire une solution robuste, capable de s'adapter progressivement à des situations plus complexes.

La première étape consiste à identifier uniquement les hurdlers, c'est-à-dire les athlètes en action, tout en éliminant les éléments parasites présents dans l'environnement (spectateurs, juges, objets non-humains, etc.).

Un premier problème rencontré concerne la détection de « faux positifs » : le modèle génère parfois des données parasites, correspondant à des formes ou structures ne représentant pas de véritables individus (des objets ayant vaguement une silhouette humaine) (Figure 9). Lorsqu'il renvoie les coordonnées des points clés, le modèle y associe un score de confiance. Nous avons donc mis en place un filtrage basé sur ce score : seules les détections qui fournissent une confiance supérieure à 0.45 sont conservées, car les données parasites ont généralement un score plus faible. Néanmoins, cette méthode présente une limite : il arrive que le squelette de l'athlète lui-même soit détecté avec un score relativement bas, ce qui peut entraîner l'exclusion de données pourtant pertinentes.

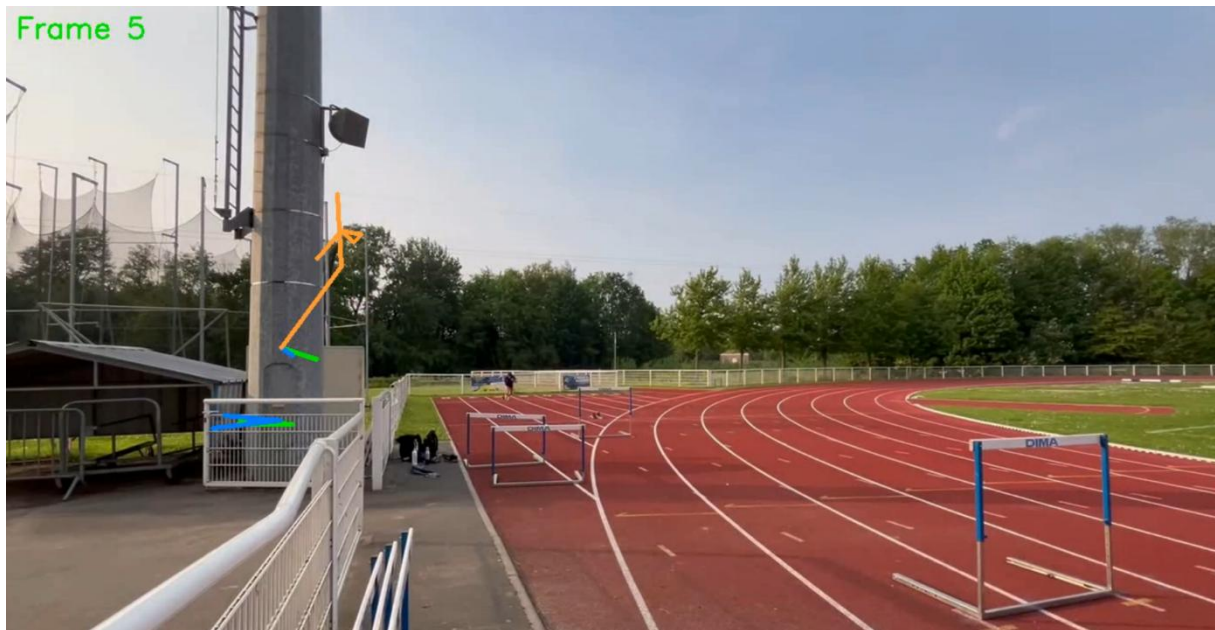


Figure 9 : Exemple de détection de faux positif

Un second défi concerne la détection des personnes non impliquées dans la course (spectateurs, juges, etc.). L'API attribue un identifiant unique à chaque individu détecté, mais celui-ci peut varier d'une image à l'autre. L'apparition ou disparition ponctuelle de personnes dans le champ perturbe la continuité du suivi, entraînant une réattribution des identifiants et compliquant ainsi le suivi cohérent d'un même athlète.

Pour contourner cette difficulté, nous avons conçu un algorithme qui sélectionne uniquement les individus détectés de manière continue du début à la fin de la vidéo. Cette approche repose sur l'hypothèse qu'un athlète engagé dans une course reste présent tout au long de la séquence. L'algorithme analyse d'abord la vidéo et exclut les personnes qui n'apparaissent pas de façon continue. Il traite ensuite la vidéo en sens inverse selon le même principe pour éliminer ceux qui apparaissent. Cette double analyse permet d'éliminer efficacement les éléments indésirables, comme les spectateurs, qui ne sont visibles que partiellement.

4.3 Collecter les trajectoires des points d'intérêts pertinents

Une fois les athlètes correctement identifiés, l'étape suivante a consisté à sélectionner les KP les plus pertinents afin d'analyser leurs trajectoires. Dans le cadre du 110 mètres haies, l'enjeu est de repérer les points dont l'évolution dans le temps permet d'identifier les phases clés du franchissement, en particulier l'impulsion et la réception à chaque haie.

Les coordonnées récupérées pour chaque point pour chaque frame sont exprimées dans un repère U, V, où (0, 0) correspond au coin supérieur gauche de l'image. Ainsi, l'axe U représente la dimension horizontale (de gauche à droite), tandis que l'axe V correspond à la dimension verticale (de haut en bas). Pour une lecture intuitive des trajectoires verticales, les coordonnées sur l'axe V ont été inversées, de

manière à ce qu'un mouvement vers le haut dans l'image soit représenté par une élévation dans la courbe.

Par ailleurs, nous avons constaté que, pour cette détection, seules les coordonnées verticales (axe V) étaient véritablement pertinentes, car l'objectif n'était pas de suivre les valeurs absolues dans le temps, mais bien d'analyser leurs variations.

À ce stade, nous avons commencé à analyser les courbes de trajectoire des KP au fil du temps, dans le but d'observer leur comportement pendant les phases de course et de franchissement. Après plusieurs tests, il est rapidement apparu que les points situés au niveau des pieds, et plus précisément les orteils, étaient les plus informatifs pour détecter automatiquement ces phases.

Ci-dessous (Figure 10), nous présentons la courbe brute d'un exemple de trajectoire verticale des orteils, extraite d'une vidéo d'entraînement. On y observe des variations marquées dans certaines portions, correspondant aux franchissements. Cela est cohérent avec la dynamique du mouvement, puisque les pieds doivent s'élever plus haut au moment du passage des haies que lors des foulées intermédiaires.

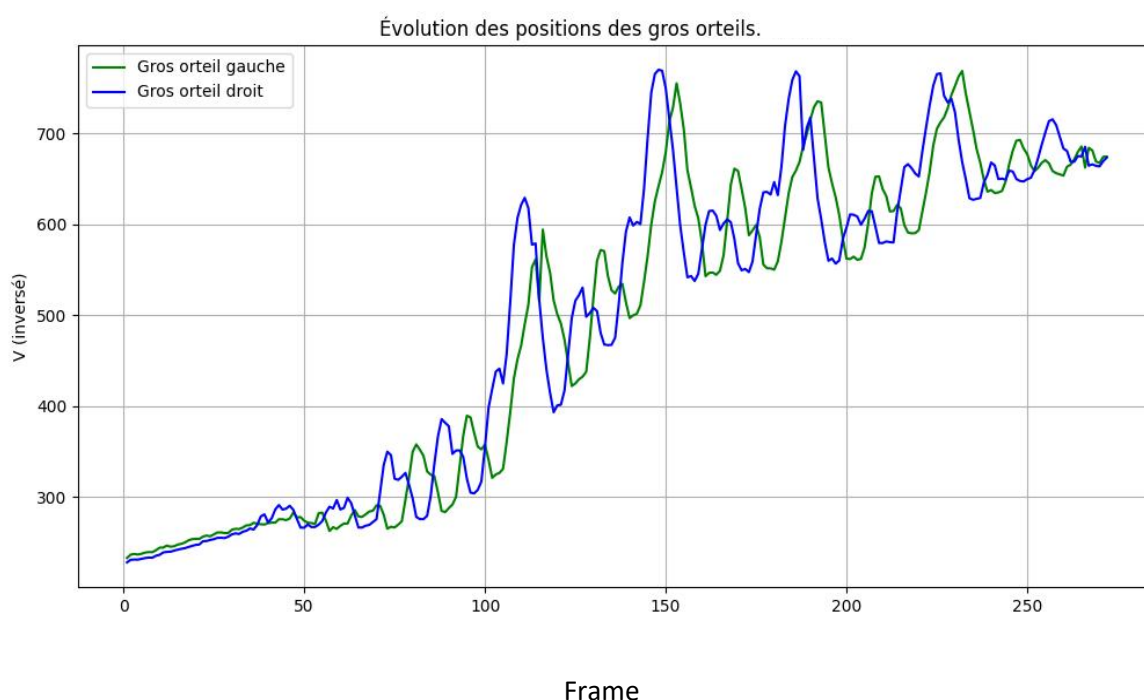


Figure 10 : Courbe brute de l'évolution des positions des gros orteils sur l'axe V au fil des images

4.4 Traiter les données collectées

La problématique suivante consiste à déterminer automatiquement, à partir des trajectoires des orteils, les événements tels que les franchissements de haies, ainsi que les phases d'impulsion et de réception associées.

Pour faciliter l'analyse des variations de trajectoires, nous avons d'abord appliqué un lissage des courbes à l'aide du filtre de Savitzky-Golay (Gallagher & Neal ,2020), permettant d'atténuer le bruit et de rendre les profils plus réguliers. (Figure 11)

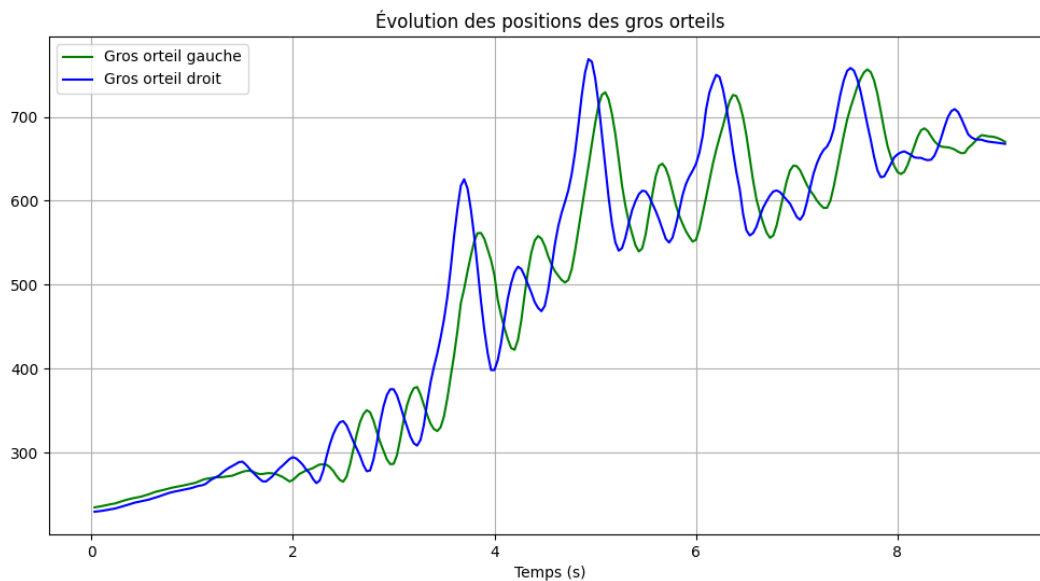


Figure 11 : Courbe lisse de l'évolution des positions des gros orteils sur l'axe V en fonction du temps.

Une fois les courbes lissées, un programme a été développé pour détecter automatiquement les événements clés selon les critères présentés ci-après.

Franchissement de haies :

La première étape de ce processus consiste à détecter tous les maxima locaux sur les courbes des orteils gauche et droit. Ensuite, un critère de sélection permet d'identifier parmi ces maxima ceux qui correspondent à de véritables franchissements de haies. Un maximum est retenu comme franchissement s'il est encadré par deux maxima plus petits que lui (un avant et un après), et si son augmentation par rapport au précédent maximum dépasse un certain seuil, fixé ici à 15 %.

Une fois les franchissements détectés sur les courbes des deux pieds, il est nécessaire d'identifier, pour chaque franchissement de haie, lequel des deux pieds correspond au pied d'attaque.

Pied d'attaque :

Pour identifier le pied d'attaque, l'algorithme compare les maxima identifiés comme franchissements sur les courbes des orteils gauche et droit, de manière paire par paire : pour chaque franchissement, on observe quel pied atteint son maximum en premier. Le pied dont le maximum apparaît en premier est considéré comme le pied d'attaque.

Une fois le pied d'attaque identifié pour chaque haie, il devient possible de déterminer sur quel pied se déroulent l'impulsion et la réception. En effet, la jambe d'attaque est celle qui passe la haie en premier, elle est donc aussi celle qui retombe en premier au sol : elle correspond à la jambe de réception. À l'inverse, la jambe opposée décolle en dernier avant le franchissement : elle correspond alors à la jambe d'impulsion.

L'algorithme permet ainsi d'identifier, pour chaque franchissement de haie, la jambe d'attaque ainsi que la jambe correspondante à la réception et à l'impulsion. Une fois cette information identifiée, il est alors possible de déterminer avec précision les instants d'impulsion et de réception pour chaque franchissement de haie, comme détaillé dans les sections suivantes.

Réception :

La réception est relativement simple à détecter. Elle correspond au minima suivant le franchissement sur la courbe du pied d'attaque. Ce point marque l'instant du premier contact entre le pied et le sol après le franchissement.

Impulsion :

La détection de l'impulsion est une tâche délicate. Le minima de l'orteil précédent le franchissement de haie, ne correspond pas nécessairement au moment où le pied quitte le sol comme le montre la Figure 12.

Pour affiner cette détection, on utilise la position de la cheville. Parmi les 10 images qui suivent ce minima, on cherche celle où l'orteil et la cheville sont le plus alignés verticalement. Autrement dit, on identifie l'image où la différence horizontale (coordonnées U) entre ces deux points est la plus faible. Ce moment précis correspond alors à l'impulsion (Figure 13).

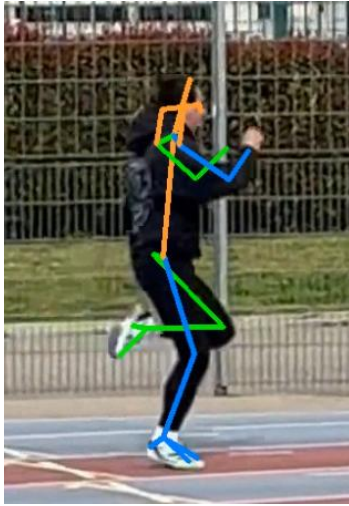


Figure 12 : Détection simple de l'impulsion (minima orteil)



Figure 13 : Détection affinée de l'impulsion (orteil + cheville)

L'ensemble de ces étapes permet ainsi de détecter automatiquement les différents événements clés, comme illustré sur la figure suivante.

Événements associés à l'évolution des positions des gros orteils

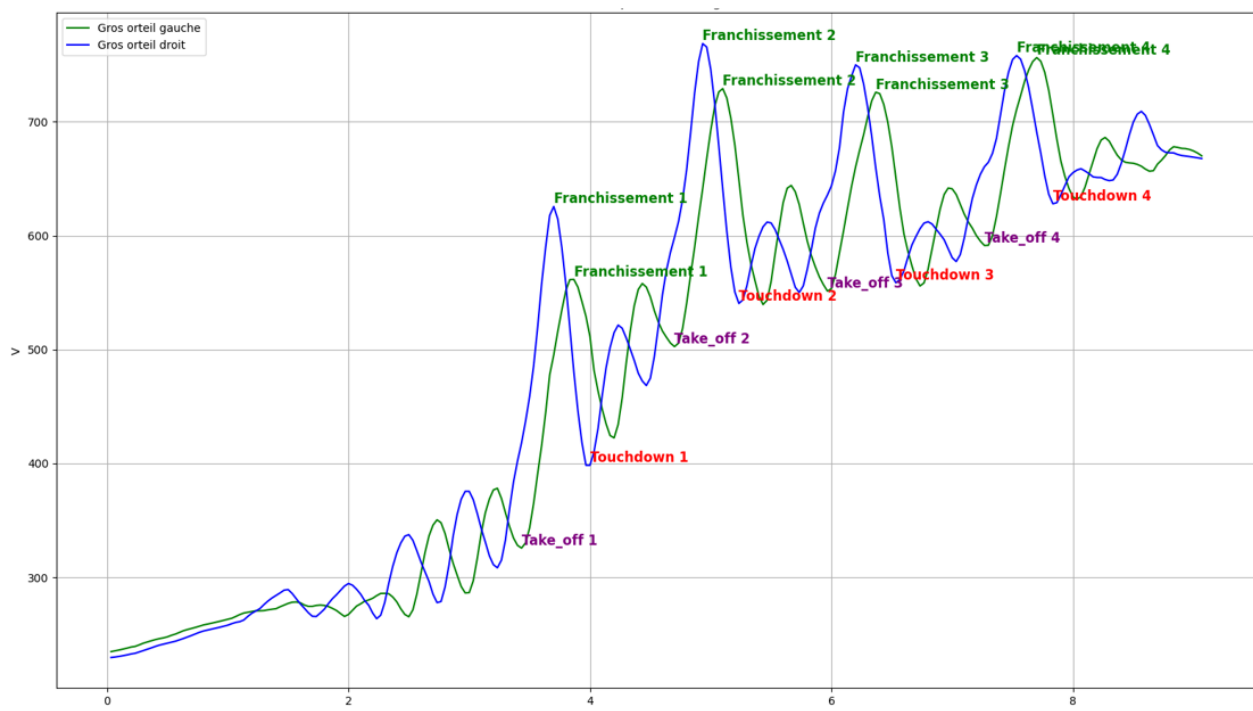


Figure 14 : Événements clés identifiés automatiquement à partir des trajectoires temporelles des orteils

4.5 Calcul/Sauvegarde des métriques

L'application permet, à partir d'une simple vidéo en entrée, d'extraire automatiquement les métriques suivantes :

- Le pied d'attaque pour chaque franchissement de haie.
- Le temps (en secondes) de chaque impulsion.
- Le temps (en secondes) de chaque réception.

À partir de ces données de base, il devient alors possible de calculer automatiquement :

- Le temps de vol pour chaque haie (Réception n – Impulsion n).
- Le temps de course entre chaque haie (Impulsion n – Réception $n-1$).
- L'intervalle entre chaque haie (Réception n – réception $n-1$).

Ces métriques sont ensuite automatiquement enregistrées dans un fichier Excel, selon le même format que celui utilisé par les fichiers générés avec Méta-vidéo.

4.6 Protocole expérimental

Intégration de l'outil et protocole de validation :

L'ensemble du traitement, du nettoyage des données à l'export des métriques a été développé dans un programme nommé CITRUS (*automatiC annotatIon Tool foR hUrdle raceS*) dont les étapes sont présentées sur la Figure 15.

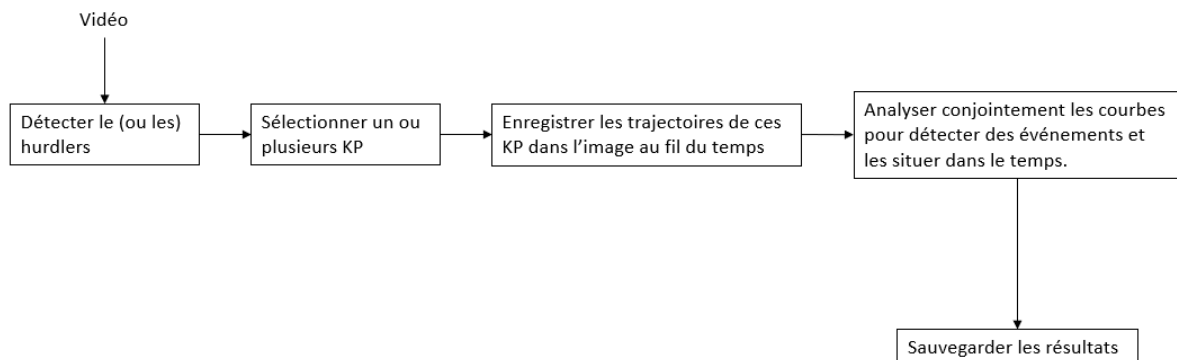


Figure 15 : Schéma représentant les fonctionnalités de CITRUS

L'objectif est d'évaluer la précision de CITRUS en comparant les résultats générés automatiquement par notre solution à ceux obtenus par la méthode manuelle de référence, à savoir l'annotation réalisée sur Méta-vidéo. Cette comparaison permettra d'évaluer :

- La fiabilité des métriques issues de l'outil automatique.
- Le gain de temps de traitement.
- La possibilité d'intégrer CITRUS dans les outils utilisés par la FFA.

L'annotation automatique présente en effet plusieurs avantages :

- Un temps de traitement réduit.
- Une suppression de la variabilité inter-observateur.
- Et une meilleure standardisation des mesures.

Données et matériel :

Jeu de données : vidéos d'entraînement de 110 m haies (configuration *single-person*, 30 Hz), pour la plupart fournies par la FFA, d'autre filmées avec nos moyens. Bien que leur nombre soit limité, elles suffisent pour une première évaluation de performances :

| | Vidéo 1 | Vidéo 2 | Vidéo 3 | Vidéo 4 | Vidéo 5 | Vidéo 6 |
|-------------------------------|---------|---------|---------|---------|---------|---------|
| Nombre de haies dans la vidéo | 4 | 3 | 3 | 3 | 3 | 1 |
| Nombre de spectateurs | 3 | 2 | 2 | 0 | 1 | 7 |
| Nombre de Frame | 272 | 149 | 149 | 280 | 238 | 201 |
| Fréquence de la vidéo | 30 Hz | 30 Hz | 30 Hz | 30 Hz | 30 Hz | 30 Hz |
| Fournie par | FFA | FFA | FFA | Auteur | Auteur | FFA |

Tableau 2 : Caractéristique des vidéos du jeu de données



Tableau 3 : Illustration des vidéos du jeu de données

Matériel : les traitements ont été effectués sur un poste doté d'un Intel Xeon W3-2423, 32 Go de RAM, et d'une Nvidia RTX 2000 16 Go ada génération. Ces spécifications restent modestes pour ce type d'analyse.

Protocole expérimental :

Chaque vidéo est analysée de deux manières :

1. Automatiquement, via l'application CITRUS
2. Manuellement, avec l'outil Méta-vidéo

Cette double annotation permet une comparaison directe sur les mêmes supports vidéo, garantissant une évaluation équitable.

Indicateurs de performance :

Deux critères principaux sont retenus :

- L'écart entre les métriques générées par CITRUS et celles de la vérité terrain (Méta-vidéo) : pour évaluer la précision de l'outil automatique.
- Temps de traitement de chaque vidéo : pour quantifier le gain d'efficacité.

5. Résultats

5.1 Comparaison des métriques

Nous avons ainsi extrait les résultats des métriques pour chaque vidéo, afin de les comparer entre les deux méthodes utilisées. L'objectif est que les résultats obtenus avec Citrus se rapprochent le plus possible de ceux générés par Méta-vidéo, dans la perspective d'une adoption future par la FFA. Nous présenterons ici les résultats pour les deux premières vidéos du jeu de données. Les résultats des autres vidéos seront présentés dans l'annexe.

5.1.1 Résultat pour la vidéo 1

| | Impulsion H1 | Réception H1 | Impulsion H2 | Réception H2 | Impulsion H3 | Réception H3 | Impulsion H4 | Réception H4 |
|------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| Méta-vidéo | 3.52 | 3.952 | 4.764 | 5.197 | 6.006 | 6.472 | 7.334 | 7.769 |
| CITRUS | 3.533 | 4 | 4.767 | 5.233 | 6.033 | 6.563 | 7.333 | 7.833 |
| Différence | 0.013 | 0.048 | 0.003 | 0.036 | 0.027 | 0.091 | 0.001 | 0.064 |

Différence Moyenne avec Méta-vidéo : 0.035375 secondes

Tableau 4 : Temps en secondes des Impulsions/Réceptions de la vidéo 1

| | Haie 1 | Haie 2 | Haie 3 | Haie 4 |
|------------|--------|--------|--------|--------|
| Méta-vidéo | 0.45 | 0.431 | 0.529 | 0.51 |
| Citrus | 0.467 | 0.467 | 0.5 | 0.5 |
| Différence | 0.017 | 0.036 | 0.029 | 0.01 |

Différence Moyenne avec Méta-vidéo : 0.023 secondes

Tableau 5 : Temps de vol en secondes de la vidéo 1

| | Départ-H1 | H1-H2 | H2-H3 | H3-H4 |
|------------|-----------|-------|-------|-------|
| Méta-vidéo | 3.488 | 0.804 | 0.793 | 0.823 |
| Citrus | 3.533 | 0.767 | 0.8 | 0.8 |
| Différence | 0.045 | 0.037 | 0.007 | 0.023 |

Différence Moyenne avec Méta-vidéo : 0.028 secondes

Tableau 6 : Temps de course en secondes de la vidéo 1

| | Départ-H1 | H1-H2 | H2-H3 | H3-H4 |
|------------|-----------|-------|-------|-------|
| Méta-vidéo | 3.938 | 1.235 | 1.322 | 1.333 |
| Citrus | 4 | 1.233 | 1.267 | 1.30 |
| Différence | 0.062 | 0.002 | 0.055 | 0.033 |

Différence Moyenne avec Méta-vidéo : 0.038 secondes

Tableau 7 : Intervalles en secondes de la vidéo 1

5.1.2 Résultat pour la vidéo 2

| | Impulsion_H1 | Réception_H1 | Impulsion_H2 | Réception_H2 | Impulsion_H3 | Réception_H3 |
|------------|--------------|--------------|--------------|--------------|--------------|--------------|
| Méta-vidéo | 0.952 | 1.343 | 2.16 | 2.533 | 3.334 | 3.744 |
| Citrus | 0.933 | 1.36 | 2.167 | 2.56 | 3.33 | 3.76 |
| Différence | 0.019 | 0.017 | 0.007 | 0.027 | 0.004 | 0.022 |

Différence Moyenne avec Méta-vidéo : 0.016 secondes

Tableau 8 : Temps en secondes des Impulsions/Réceptions de la vidéo 2

| | Haie 1 | Haie 2 | Haie 3 |
|------------|--------|--------|--------|
| Méta-vidéo | 0.391 | 0.373 | 0.41 |
| Citrus | 0.4 | 0.333 | 0.433 |
| Différence | 0.008 | 0.040 | 0.023 |

Différence Moyenne avec Méta-vidéo : 0.023 secondes

Tableau 9 : Temps de vol en seconde de la vidéo 2

| | Départ -> H1 | H1-H2 | H2-H3 |
|------------|--------------|-------|-------|
| Méta-vidéo | 0.952 | 0.817 | 0.801 |
| Citrus | 0.933 | 0.807 | 0.77 |
| Différence | 0.019 | 0.01 | 0.031 |

Différence Moyenne avec Méta-vidéo : 0.020 secondes

Tableau 10 : Temps de course en seconde de la vidéo 2

| | Départ -> H1 | H1-H2 | H2-H3 |
|------------|--------------|-------|-------|
| Méta-vidéo | 1.343 | 1.19 | 1.211 |
| Citrus | 1.36 | 1.2 | 1.206 |
| Différence | 0.03 | 0.001 | 0.004 |

Différence Moyenne avec Méta-vidéo : 0.011 secondes

Tableau 11 : Temps des intervalles en secondes de la vidéo 2

Les résultats les plus significatifs sont ceux figurant dans les tableaux relatifs aux impulsions et réceptions, car les autres métriques (temps de vol, temps de course et intervalles) sont simplement dérivées de ces données principales.

Sur la première vidéo, on observe sur le Tableau 4 une différence moyenne des métriques de 0,035 secondes, allant de 0.001 à 0.091. On constate que les écarts les plus importants apparaissent lors des phases de réception, aucune valeur ne descendant en dessous de 0,038. Sur la seconde vidéo (Tableau 8), la différence moyenne est de 0,016, ce qui constitue un bon résultat (cf. Discussion). Là encore, les écarts les plus marqués concernent les réceptions.

En ce qui concerne les métriques dérivées (temps de vol, temps de course, intervalles), les écarts moyens observés sur les deux vidéos tournent autour de 0,020 secondes, avec des valeurs comprises entre 0,011

et 0,038. Il est important de rappeler que ces résultats ont été obtenus à partir de vidéos enregistrées à 30 Hz (cf. Discussion).

5.2 Temps de traitement

Nous présentons à présent le temps de traitement de chaque méthode pour chacune des vidéos.

| | Vidéo 1 | Vidéo 2 | Vidéo 3 | Vidéo 4 | Vidéo 5 | Vidéo 6 |
|------------------------------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Nombre de haie dans la vidéo | 4 | 3 | 3 | 3 | 3 | 1 |
| Nombre de Frame | 272 | 149 | 149 | 280 | 238 | 201 |
| Fréquence de la vidéo | 30 Hz | 30 Hz | 30 Hz | 30 Hz | 30 Hz | 30 Hz |
| Temp de traitement avec Méta-vidéo | 6 min 10 | 5 min 33 | 5 min 03 | 5 min 0 | 4 min 59 | 5 min |
| Temps de traitement avec Citrus | 1 min 33 | 0 min 47 | 0 min 47 | 1 min 40 | 1 min 12 | 1 min 09 |
| Gain de temps | 4 min 37 | 4 min 46 | 4 min 16 | 3 min 20 | 3 min 47 | 3 min 51 |
| Gain de temps (%) | 74.86% | 85.89% | 84.49% | 66.67% | 75.92% | 77.00% |

Tableau 12 : Temps de traitement de chaque méthode sur chacune des vidéos

On observe, sur le Tableau 12, pour chaque vidéo, un gain de temps de traitement allant de 3 minutes 20 à 4 minutes 46, ce qui est loin d'être négligeable puisque cela correspond en moyenne à environ 77,5% de gain de temps en passant de l'annotation manuelle à l'annotation automatique. Dans le cadre d'un traitement sur plusieurs vidéos, cela représenterait un gain de temps conséquent, permettant aux entraîneurs de fournir un retour à leurs athlètes de manière beaucoup plus rapide. Il est également important de souligner que, lorsque l'application pourra traiter des vidéos plus longues, le gain de temps comparé à la solution manuelle sera encore plus significatif.

6. Discussion

Notre travail vise à évaluer la pertinence de l'utilisation des techniques d'HPE pour l'annotation automatique dans le cadre des courses de 110 mètres haies. L'objectif principal est de proposer une alternative à l'annotation manuelle, en développant une solution automatisée capable de réduire significativement le temps de traitement et d'éliminer la variabilité inter-observateurs, tout en conservant un niveau de précision équivalent.

Les HPE connaissent un développement rapide dans de nombreuses disciplines de l'athlétisme, notamment pour prédire certains événements (M. Einfalt et al., 2019 ; Gan et al., 2024), notamment en course de haies (Jafarzadeh, P. et al., 2025). Toutefois, leur utilisation pour l'annotation automatique d'événements précis reste encore très limitée. À notre connaissance, le modèle développé dans le cadre de cette étude est le premier à permettre l'annotation automatique des phases clés d'une course de 110 mètres haies. La méthode proposée représenterait ainsi une avancée significative pour l'analyse de la performance en course de haies en France, tout en ouvrant la voie à une application plus large de la vision par ordinateur dans d'autres disciplines de l'athlétisme.

Les métriques obtenues par CITRUS sont très encourageantes pour une première version. Nous avons régulièrement échangé avec la FFA, qui s'est montrée très intéressée des résultats. Selon leurs indications, l'objectif est d'atteindre une erreur inférieure à 0,01 pour envisager l'intégration de CITRUS dans leurs outils. Il est également important de préciser que les tests ont été réalisés à partir de vidéos enregistrées à 30 Hz, ce qui limite la précision des mesures. L'utilisation de vidéos à 100 Hz ou plus permettrait de réduire significativement la différence de résultats entre les deux outils, aujourd'hui évaluer entre 0.016 et 0.035 secondes et ainsi d'atteindre l'objectif fixé. À ce jour, nous n'avons pas encore reçu de vidéos à 100 Hz. Nous prévoyons donc de les réaliser nous-mêmes dans les semaines à venir afin de procéder aux tests.

Enfin, bien que nous utilisions Méta-vidéo comme référence, l'outil actuel de la FFA qui avait fait l'objet de mon stage de Master 1, il ne s'agit pas d'une vérité terrain totalement fiable. Ses annotations, réalisées manuellement, sont sujettes à une part de subjectivité et peuvent varier légèrement d'une prise à l'autre. Nous avons vérifié plusieurs détections générées par CITRUS en observant les frames des impulsions et réceptions, et celles-ci correspondaient globalement aux moments réels dans la vidéo. Toutefois, nous avons remarqué que la détection des réceptions semblait parfois décalée d'une à deux frames par rapport à la vérité terrain selon la fréquence d'images. Cela indique qu'il serait pertinent de disposer de vidéos à fréquence élevée et renforcer spécifiquement l'algorithme pour améliorer la précision sur ces phases clés, qui jouent un rôle crucial dans l'analyse technique.

En résumé, les métriques générées par CITRUS devraient rester très proches sur des vidéos à 100 Hz, à condition qu'elles soient capturées dans de bonnes conditions, en situation réelle avec un seul athlète.

Le temps de traitement est également bien plus court que celui requis pour une annotation manuelle, tout en éliminant la variabilité inter-observateur. L'outil pourrait ainsi permettre à l'entraîneur d'accéder aux métriques de la course seulement 2 à 3 minutes après celle-ci, ce qui représente un gain considérable par rapport aux méthodes actuelles. Nous avons pu vérifier que l'application fonctionne à différents niveaux, allant des athlètes de haut niveau grâce aux vidéos fournies par la FFA, jusqu'au niveau débutant avec les vidéos où je suis le coureur de 110m haie.

Par ailleurs, si l'outil actuel permet déjà de détecter les métriques attendues par la FFA, l'objectif est désormais de le faire évoluer en intégrant des indicateurs encore inaccessibles via l'annotation manuelle. L'idée est d'apporter une réelle plus-value en fournissant des métriques imperceptibles à l'œil nu. C'est dans cette optique que nous avons échangé avec Benjamin Crouzet, entraîneur de Pascal Martinot-Lagarde, afin qu'il puisse nous orienter sur les indicateurs qui l'intéresseraient particulièrement. Il en est ressorti qu'une analyse fine des 13,72 premiers mètres, du départ jusqu'à la première haie, pourrait être très pertinente. Nous avons donc commencé à développer un programme visant à détecter le moment où l'athlète remonte la tête et commence à regarder droit devant lui. Nous sommes actuellement dans l'attente des retours de Benjamin Crouzet sur ces premiers résultats.

Les limites actuelles de l'outil résident dans le fait qu'il n'est, pour l'instant, pleinement opérationnel que dans un contexte d'entraînement, avec un seul athlète présent sur la piste. Sur des vidéos de compétition, les HPE rencontrent encore des difficultés, notamment en présence d'occultations ou lorsque les athlètes sont éloignés de la caméra. L'objectif à moyen terme est que CITRUS soit progressivement capable de fonctionner dans des contextes de plus en plus complexes, jusqu'à pouvoir analyser des vidéos de compétition, avec un athlète par couloir.

Une piste d'amélioration consisterait à combiner plusieurs modèles HPE afin de compenser les faiblesses spécifiques à chacun. De plus, ces algorithmes ont encore du mal à détecter correctement les poses dans des situations particulières comme le départ en starting-blocks. Une solution envisagée serait de réentraîner le modèle HPE sélectionné à partir d'un jeu de données contenant des images spécifiques au 110 mètres haies, afin d'améliorer la détection des squelettes dans les postures typiques du départ et du franchissement.

Enfin, pour enrichir encore davantage l'analyse, une évolution future de CITRUS pourrait viser à détecter de nouvelles métriques jusqu'ici non analysées. Cela permettrait d'approfondir la compréhension technique de la discipline et d'identifier des paramètres de performance encore inexplorés.

7. Conclusion

En conclusion, cette étude met en lumière le potentiel et la pertinence des techniques d'estimation de la pose humaine pour améliorer l'analyse et la compréhension des performances en 110 mètres haies. Malgré les défis rencontrés, notamment liés aux faux positifs et à la détection précise des événements à partir des coordonnées, notre approche a permis d'extraire des données fiables sur les moments clés identifiés comme essentiels par la FFA au cours de la course. Ces métriques, auparavant extraites manuellement, peuvent désormais être obtenues automatiquement, ce qui permet un gain de temps considérable tout en éliminant la variabilité inter-observateurs.

L'application de notre méthode ouvre la voie à une analyse plus fine et approfondie des performances sur 110 mètres haies. L'objectif est désormais de poursuivre le développement de l'outil afin qu'il soit capable de fonctionner dans des contextes de plus en plus complexes, ce qui constituera l'axe principal de la fin de mon stage. Par ailleurs, ce modèle pourrait offrir la possibilité, avec un peu plus de temps, d'extraire de nouvelles métriques jusqu'ici inaccessibles, représentant une avancée significative pour la Fédération Française d'Athlétisme. Ces informations supplémentaires pourraient fournir aux entraîneurs des indicateurs inédits, contribuant ainsi de manière concrète à l'optimisation de la performance.

Il convient toutefois de souligner que ce projet est encore en phase de développement, et que des améliorations restent nécessaires pour atteindre un niveau de précision pleinement satisfaisant. Par ailleurs, l'utilisation de cet outil automatisé implique de considérer les bénéfices et les risques pour les différentes parties prenantes. Les athlètes peuvent bénéficier d'une analyse plus fine et non intrusive (sans marqueurs) mais une attention particulière doit être portée au respect de leur vie privée et à l'interprétation des données. En ce qui concerne les entraîneurs, la mise en place de l'expérimentation reste relativement simple, puisqu'une seule caméra suffit. Toutefois, certaines recommandations (vue sagittale, distance caméra-athlète faible, single-person) de captation doivent être respectées afin de garantir la précision des résultats obtenus.

En résumé, notre recherche apporte une contribution significative à la compréhension et à l'analyse de la performance au 110 mètres haie, ouvrant de nouvelles perspectives pour l'entraînement, la recherche et le développement technologique dans ce domaine.

8. Bibliographie

Ando, Shingo & Wu, Xiaojun & Suzuki, Akira & Wakabayashi, Kaoru & Koike, Hideki. (2007). Human Pose Estimation for Image Monitoring. Vol. 5, No. 11, pp. 25–32, NTT Technical Review. 5. 10.53829/ntr200711sf4.

Ballabio, G., Perego, P., and Wang, Q. (2023) Designing an interactive system based on pose-estimation to support rhythmic gymnastics basic coaches in enhancing their learning, in De Sainz Molestina, D., Galluzzo, L., Rizzo, F., Spallazzo, D. (eds.), *IASDR 2023: Life-Changing Design*, 9-13 October, Milan, Italy. <https://doi.org/10.21606/iasdr.2023.571>

Z. Cao, T. Simon, S. -E. Wei and Y. Sheikh, "Realtime Multi-person 2D Pose Estimation Using Part Affinity Fields," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, 2017, pp. 1302-1310, doi: 10.1109/CVPR.2017.143.

Chen K, Gabriel P, Alasfour A, Gong C, Doyle WK, Devinsky O, Friedman D, Dugan P, Melloni L, Thesen T, Gonda D, Sattar S, Wang S, Gilja V. Patient-Specific Pose Estimation in Clinical Environments. *IEEE J Transl Eng Health Med*. 2018 Oct 10;6:2101111. doi: 10.1109/JTEHM.2018.2875464

T. F. Cootes, G. J. Edwards and C. J. Taylor, "Active appearance models," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 681-685, June 2001, doi: 10.1109/34.927467.

I. Culjak, D. Abram, T. Pribanic, H. Dzapo and M. Cifrek, "A brief introduction to OpenCV," *2012 Proceedings of the 35th International Convention MIPRO*, Opatija, Croatia, 2012, pp. 1725-1730

Difini, Marcio Garcia Martins, and Jorge Luis Victória Barbosa. 2021. Human Pose Estimation for Training Assistance: a Systematic Literature Review. In *Proceedings of the Brazilian Symposium on Multimedia and the Web (WebMedia '21)*. Association for Computing Machinery, New York, NY, USA, 189–196. <https://doi.org/10.1145/3470482.3479633>

Dubbel , Real-time Head Pose Estimation in Low-resolution Football Footage Using Random Forests (2013)

Eichner, M., Marin-Jimenez, M., Zisserman, A. *et al.* 2D Articulated Human Pose Estimation and Retrieval in (Almost) Unconstrained Still Images. *Int J Comput Vis* **99**, 190–214 (2012). <https://doi.org/10.1007/s11263-012-0524-9>

Moritz Einfalt, Charles Dampérou, Dan Zecha, and Rainer Lienhart. 2019. Frame-Level Event Detection in Athletics Videos with Pose-Based Convolutional Sequence Networks. In *Proceedings of the 2nd International Workshop on Multimedia Content Analysis in Sports (MMSports '19)*. Association for Computing Machinery, New York, NY, USA, 42–50. <https://doi.org/10.1145/3347318.3355525>

Fukushima, T., Blauburger, P., Guedes Russomanno, T. *et al.* The potential of human pose estimation for motion capture in sports: a validation study. *Sports Eng* **27**, 19 (2024). <https://doi.org/10.1007/s12283-024-00460-w>

Gallagher, Neal. (2020). Savitzky-Golay Smoothing and Differentiation Filter. 10.13140/RG.2.2.20339.50725.

Q. Gan, S. M. Nguyen, M. A. El-Yacoubi, E. Fenaux and S. Cléménçon, "Human Pose Estimation Based Biomechanical Feature Extraction for Long Jumps," *2024 16th International Conference on Human System Interaction (HSI)*, Paris, France, 2024, pp. 1-6, doi: 10.1109/HSI61632.2024.10613530.

- Y. Gu, S. Pandit, E. Saraee, T. Nordahl, T. Ellis and M. Betke, "Home-Based Physical Therapy with an Interactive Computer Vision System," *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, Seoul, Korea (South), 2019, pp. 2619-2628, doi: 10.1109/ICCVW.2019.00320.
- X. Gong, X. Geng, G. Nie, T. Wang, J. Zhang and J. You, "Normative Evaluation Method of Long Jump Action Based on Human Pose Estimation," in *IEEE Access*, vol. 11, pp. 125452-125459, 2023, doi: 10.1109/ACCESS.2023.3330757.
- Hartley, R., & Zisserman, A. (2004). *Multiple view geometry in computer vision* (2nd ed.). Cambridge University Press.
- Heindl, Christoph & Ikeda, Markus & Stübl, Gernot & Pichler, Andreas & Scharinger, Josef. (2019). Enhanced Human-Machine Interaction by Combining Proximity Sensing with Global Perception. 10.48550/arXiv.1910.02445.
- Huang, Ying & Sun, Bin & Kan, Haipeng & Zhuang, Jiankai & Qin, Zengchang. (2019). FollowMeUp Sports: New Benchmark for 2D Human Keypoint Recognition. 10.1007/978-3-030-31726-3_10.
- Insafutdinov, E., Pishchulin, L., Andres, B., Andriluka, M., Schiele, B. (2016). DeeperCut: A Deeper, Stronger, and Faster Multi-person Pose Estimation Model. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds) Computer Vision – ECCV 2016. ECCV 2016. Lecture Notes in Computer Science(), vol 9910. Springer, Cham. https://doi.org/10.1007/978-3-319-46466-4_3
- P. Jafarzadeh, P. Virjonen, P. Nevalainen, F. Farahnakian and J. Heikkonen, "Pose Estimation of Hurdles Athletes using OpenPose," *2021 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME)*, Mauritius, Mauritius, 2021, pp. 1-6, doi: 10.1109/ICECCME52200.2021.9591066.
- Jafarzadeh, P., Zelioli, L., Virjonen, P. *et al.* Enhancing hurdles athletes' performance analysis: A comparative study of cnn-based pose estimation frameworks. *Multimed Tools Appl* (2025). <https://doi.org/10.1007/s11042-024-20587-z>
- Jiang, Tao & Lu, Peng & Zhang, Li & Ma, Ningsheng & Han, Rui & Lyu, Chengqi & Li, Yining & Chen, Kai. (2023). RTMPose: Real-Time Multi-Person Pose Estimation based on MMPose. 10.48550/arXiv.2303.07399.
- Kresović, Milan & Nguyen, Thong. (2021). Bottom-up approaches for multi-person pose estimation and its applications: A brief review. 10.48550/arXiv.2112.11834.
- LaChance, J., Thong, W., Nagpal, S., & Xiang, A. (2023). A case study in fairness evaluation: Current limitations and challenges for human pose estimation. *arXiv*.
- Y. -C. Li, C. -T. Chang, C. -C. Cheng and Y. -L. Huang, "Baseball Swing Pose Estimation Using OpenPose," *2021 IEEE International Conference on Robotics, Automation and Artificial Intelligence (RAAI)*, Hong Kong, Hong Kong, 2021, pp. 6-9, doi: 10.1109/RAAI52226.2021.9507807.
- Li, Yanjie & Yang, Sen & Liu, Peidong & Zhang, Shoukui & Wang, Yunxiao & Wang, Zhicheng & Yang, Wankou & Xia, Shu-Tao. (2022). SimCC: A Simple Coordinate Classification Perspective for Human Pose Estimation. 10.1007/978-3-031-20068-7_6.
- Liu, Chang & Gao, Zichun & Liao, Zhenyu & Sun, Yue & Feng, Xianglong. (2024). An Exploration of Human Pose Estimation Based Cheating Tools for FPS Video Game and its Defense Solution. 1-6. 10.1109/MMSP61759.2024.10743575.
- K. Ludwig, S. Scherer, M. Einfalt and R. Lienhart, "Self-Supervised Learning for Human Pose Estimation in Sports," *2021 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, Shenzhen, China, 2021, pp. 1-6, doi: 10.1109/ICMEW53276.2021.9456000.

D. Maji, S. Nagori, M. Mathew and D. Poddar, "YOLO-Pose: Enhancing YOLO for Multi Person Pose Estimation Using Object Keypoint Similarity Loss," *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, New Orleans, LA, USA, 2022, pp. 2636-2645, doi: 10.1109/CVPRW56347.2022.00297.

Morshidi, Malik & Gunawan, Teddy & Olanrewaju, Rashidah & Arifin, Fatchul. (2022). Pose estimation algorithm for mobile augmented reality based on inertial sensor fusion. *International Journal of Electrical and Computer Engineering (IJECE)*. 12. 3620. 10.11591/ijece.v12i4.pp3620-3631.

Newell, A., Yang, K., Deng, J. (2016). Stacked Hourglass Networks for Human Pose Estimation. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds) *Computer Vision – ECCV 2016*. ECCV 2016. Lecture Notes in Computer Science(), vol 9912. Springer, Cham. https://doi.org/10.1007/978-3-319-46484-8_29

Nguyen, T.D., & Kresovic, M. (2022). A survey of top-down approaches for human pose estimation. *ArXiv*, abs/2202.02656.

Qu, Haoxuan & Xu, Li & Cai, Yujun & Foo, Lin Geng & Liu, Jun. (2022). Heatmap Distribution Matching for Human Pose Estimation. 10.48550/arXiv.2210.00740.

Sun, X., Shang, J., Liang, S., & Wei, Y. (2017). Compositional human pose regression. In *Proceedings of the IEEE international conference on computer vision* (pp. 2602-2611).

Sun, Bin Xiao, Fangyin Wei, Shuang Liang, and Yichen Wei. 2018. Integral Human Pose Regression. In *Computer Vision – ECCV 2018: 15th European Conference, Munich, Germany, September 8–14, 2018, Proceedings, Part VI*. Springer-Verlag, Berlin, Heidelberg, 536–553. https://doi.org/10.1007/978-3-030-01231-1_33

ke, Sun & Xiao, Bin & Liu, Dong & Wang, Jingdong. (2019). Deep High-Resolution Representation Learning for Human Pose Estimation. 10.48550/arXiv.1902.09212.

A. Toshev and C. Szegedy, "DeepPose: Human Pose Estimation via Deep Neural Networks," *2014 IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, USA, 2014, pp. 1653-1660, doi: 10.1109/CVPR.2014.214.

Wang Dong, Wenjun Xie, Youcheng Cai, Xinjie Li, Xiaoping Liu, Transformer-based rapid human pose estimation network, *Computers & Graphics*, Volume 116, 2023, Pages 317-326, ISSN 0097-8493, <https://doi.org/10.1016/j.cag.2023.09.001>.

Wang Chun-Yi, Kalin Guanlun Lai, Hsu-Chun Huang, Wei-Ting Lin, Tennis player actions dataset for human pose estimation, *Data in Brief*, Volume 55, 2024, 110665, ISSN 2352-3409, <https://doi.org/10.1016/j.dib.2024.110665>.

W. Wen, Y. Yang, J. Du, L. Liu and J. Wang, "Gymnastic Posture Detection Based on Deep Learning," *2019 IEEE 3rd Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC)*, Chongqing, China, 2019, pp. 1277-1281, doi: 10.1109/IMCEC46724.2019.8983972.

Xi, X., Zhang, C., Jia, W., & Jiang, R. (2024). Enhancing human pose estimation in sports training: Integrating spatiotemporal transformer for improved accuracy and real-time performance. *Alexandria Engineering Journal*, 109, 144-156. <https://doi.org/10.1016/j.aej.2024.08.072>

Xu, Dongpo & Liu, Yunqing & Wang, Qian. (2021). Random forest-based human pose detection system for through-the-wall radar. *Journal of Physics: Conference Series*. 1966. 012040. 10.1088/1742-6596/1966/1/012040.

Xu, Y., Zhang, J., Zhang, Q., & Tao, D. (2022). ViTPose: Simple Vision Transformer Baselines for Human Pose Estimation. *arXiv*. <https://doi.org/10.48550/arXiv.2204.12484>.

Zheng Ce, Wenhan Wu, Chen Chen, Taojiannan Yang, Sijie Zhu, Ju Shen, Nasser Kehtarnavaz, and Mubarak Shah. 2023. Deep Learning-based Human Pose Estimation: A Survey. *ACM Comput. Surv.* 56, 1, Article 11 (January 2024), 37 pages. <https://doi.org/10.1145/3603618>

Zhou, Mu & Stofl, Lucas & Mathis, Mackenzie & Mathis, Alexander. (2023). Rethinking pose estimation in crowds: overcoming the detection information bottleneck and ambiguity. 14643-14653. 10.1109/ICCV51070.2023.01350.

9. Annexes

Résultats de la vidéo 3 :

| | Impulsion_H1 | Réception_H1 | Impulsion_H2 | Réception_H2 | Impulsion_H3 | Réception_H3 |
|------------|--------------|--------------|--------------|--------------|--------------|--------------|
| Méta-vidéo | 0.952 | 1.343 | 2.16 | 2.533 | 3.334 | 3.744 |
| Citrus | 0.966 | 1.36 | 2.23 | 2.56 | 3.33 | 3.76 |
| Différence | 0.014 | 0.017 | 0.07 | 0.027 | 0.004 | 0.022 |

Différence Moyenne avec Méta-vidéo : 0.025 secondes

Annexe 1 : Temps en secondes des Impulsions/Réceptions de la vidéo 3

| | Haie 1 | Haie 2 | Haie 3 |
|------------|--------|--------|--------|
| Méta-vidéo | 0.391 | 0.373 | 0.41 |
| Citrus | 0.467 | 0.467 | 0.5 |
| Différence | 0.017 | 0.036 | 0.029 |

Différence Moyenne avec Méta-vidéo : 0.024 secondes

Annexe 2 : Temps en secondes de vol de la vidéo 3

| | Départ -> H1 | H1-H2 | H2-H3 |
|------------|--------------|-------|-------|
| Méta-vidéo | 0.952 | 0.817 | 0.801 |
| Citrus | 0.966 | 0.87 | 0.77 |
| Différence | 0.014 | 0.053 | 0.031 |

Différence Moyenne avec Méta-vidéo : 0.0326 secondes

Annexe 3 : Temps en secondes de course sur la vidéo 3

| | Départ -> H1 | H1-H2 | H2-H3 |
|------------|--------------|-------|-------|
| Méta-vidéo | 1.343 | 1.19 | 1.184 |
| Citrus | 1.36 | 1.2 | 1.206 |
| Différence | 0.017 | 0.016 | 0.022 |

Différence moyenne avec Méta-vidéo : 0.018 secondes

Annexe 4 : Intervalles en secondes sur la vidéo 3

Résultat avec la vidéo 4 :

| | Impulsion_H1 | Réception_H1 | Impulsion_H2 | Réception_H2 | Impulsion_H3 | Réception_H3 |
|------------|--------------|--------------|--------------|--------------|--------------|--------------|
| Méta-vidéo | 3.899 | 4.368 | 5.631 | 6.073 | 7.368 | 7.796 |
| Citrus | 3.867 | 4.33 | 5.6 | 6.033 | 7.3 | 7.767 |
| Différence | 0.032 | 0.038 | 0.031 | 0.04 | 0.068 | 0.029 |

Différence Moyenne avec Méta-vidéo : 0.0396

Annexe 5 : Temps des Impulsions/Réceptions de la vidéo 4

| | Haie 1 | Haie 2 | Haie 3 |
|------------|--------|--------|--------|
| Méta-vidéo | 0.469 | 0.442 | 0.428 |
| Citrus | 0.463 | 0.433 | 0.467 |
| Différence | 0.006 | 0.009 | 0.039 |

Différence Moyenne avec Méta-vidéo : 0.018 secondes

Annexe 6 : Temps de vol de la vidéo 4

| | Départ -> H1 | H1-H2 | H2-H3 |
|------------|--------------|-------|-------|
| Méta-vidéo | 3.899 | 1.263 | 1.295 |
| Citrus | 3.867 | 1.27 | 1.267 |
| Différence | 0.032 | 0.007 | 0.028 |

Différence Moyenne avec Méta-vidéo : 0.022 secondes

Annexe 7 : Temps de course en secondes de la vidéo 4

| | Départ -> H1 | H1-H2 | H2-H3 |
|------------|--------------|-------|-------|
| Méta-vidéo | 4.368 | 1.705 | 1.723 |
| Citrus | 4.33 | 1.703 | 1.734 |
| Différence | 0.038 | 0.002 | 0.011 |

Différence Moyenne avec Méta-vidéo : 0.017

Annexe 8 : Intervalles en secondes de la vidéo 4

Résultat avec la vidéo 5

| | Impulsion_H1 | Réception_H1 | Impulsion_H2 | Réception_H2 | Impulsion_H3 | Réception_H3 |
|------------|--------------|--------------|--------------|--------------|--------------|--------------|
| Méta-vidéo | 2.455 | 2.919 | 4.326 | 4.791 | 6.378 | 6.852 |
| Citrus | 2.367 | 2.967 | 4.267 | 4.833 | 6.333 | 6.9 |
| Différence | 0.088 | 0.048 | 0.059 | 0.042 | 0.045 | 0.048 |

Différence Moyenne avec Méta-vidéo : 0.055 secondes

Annexe 9 : Temps en secondes des Impulsions/Réceptions de la vidéo 5

| | Haie 1 | Haie 2 | Haie 3 |
|------------|--------|--------|--------|
| Méta-vidéo | 0.464 | 0.465 | 0.474 |
| Citrus | 0.6 | 0.566 | 0.567 |
| Différence | 0.136 | 0.101 | 0.093 |

Différence Moyenne avec Méta-vidéo : 0.11 secondes

Annexe 10 : Temps de vol en secondes de la vidéo 5

| | Départ -> H1 | H1-H2 | H2-H3 |
|------------|--------------|-------|-------|
| Méta-vidéo | 2.455 | 1.407 | 1.587 |
| Citrus | 2.367 | 1.3 | 1.5 |
| Différence | 0.088 | 0.107 | 0.087 |

Différence Moyenne avec Méta-vidéo : 0.094 secondes

Annexe 11 : Temps de course en secondes de la vidéo 5

| | Départ -> H1 | H1-H2 | H2-H3 |
|------------|--------------|-------|-------|
| Méta-vidéo | 2.919 | 1.872 | 2.061 |
| Citrus | 2.967 | 1.866 | 2.067 |
| Différence | 0.048 | 0.006 | 0.006 |

Différence Moyenne avec Méta-vidéo : 0.02 secondes

Annexe 12 : Intervalles en secondes de la vidéo 5

Résultat avec la vidéo 6

| | Impulsion_H1 | Réception_H1 |
|------------|--------------|--------------|
| Méta-vidéo | 5.631 | 6.043 |
| Citrus | 5.828 | 6.241 |
| Différence | 0.197 | 0.198 |

Différence Moyenne : 0.1975

Annexe 13 : Temps en secondes des Impulsions/Réceptions de la vidéo 6

Master 2 – Sciences du numérique et du sport, EUR Digisport

Institut de Recherche en Informatique et Systèmes Aléatoires

Annotation Automatique Markerless de la Performance au 110 Mètres Haies: Vers Une Approche fondée sur l'Estimation de Pose Humaine.

2024-2025

Présenté par Raphaël Peus

Encadré par Laurent Guillo

Problématique : L'analyse vidéo en athlétisme repose encore majoritairement sur des méthodes manuelles, longues et sujettes à des variations entre observateurs, notamment dans des disciplines techniques comme le 110 mètres haies. Les avancées récentes en estimation de la pose humaine (HPE) ouvrent des perspectives d'automatisation de ces analyses. Ce mémoire explore dans quelle mesure ces approches permettent d'identifier automatiquement les événements clés d'une course de haies, dans le but de proposer un outil fiable et utilisable sur le terrain.

Méthodologie : L'outil CITRUS, conçu durant ce projet, a été testé à partir de vidéos d'entraînement enregistrées à 30 Hz, dans un contexte mono-athlète. Les résultats produits automatiquement ont été comparés à ceux issus d'annotations manuelles, réalisées via l'outil Méta-vidéo actuellement utilisé. L'évaluation repose sur deux axes : la similitude des métriques extraites et le temps de traitement requis par chaque méthode.

Résultats : Les résultats montrent une différence moyenne comprise entre 0,02 et 0,03 secondes entre les deux approches. Atteindre une précision de 0,01 semble envisageable avec des vidéos à plus haute fréquence (100 Hz). Par ailleurs, l'outil CITRUS permet un gain de temps moyen de 77,5 %, ce qui représente un avantage opérationnel important.

Discussion/Conclusion : L'HPE se révèle être une voie prometteuse pour automatiser l'analyse de la performance au 110 m haies. Cependant, des améliorations restent nécessaires pour garantir sa fiabilité dans des contextes plus complexes (compétition, multiples athlètes...).

Perspectives : Le développement de CITRUS constitue une première étape. L'objectif est désormais de l'adapter à des situations plus variées, avec la perspective de générer de nouveaux indicateurs de performance, utiles pour les entraîneurs et la FFA.

Mots-clés : Estimation de la pose humaine, annotation automatique, 110 mètres haie.

Master's Degree (Year 2) – Digital Science and Sport, EUR Digisport

Research Institute of Computer Science and Random Systems (IRISA)

**Markerless Automatic Annotation of 110-Meter Hurdles Performance: Toward a
Human Pose Estimation-Based Approach**

2024-2025

Presented by Raphaël Peus

Supervised by Laurent Guillo

Problematic : Video analysis in athletics still relies heavily on manual procedures that are time-consuming and prone to inter-observer variability—particularly in technical events such as the 110-meter hurdles. Recent developments in Human Pose Estimation (HPE) offer promising avenues to automate this type of analysis. This project examines how HPE-based methods can be used to automatically detect key events in a hurdle race, with the aim of creating a tool that is both reliable and relevant for field use.

Methodology : The CITRUS tool, developed during this project, was tested on training videos recorded at 30 Hz in a single-athlete setup. The automatically generated results were compared to manual annotations made using the Méta-vidéo software currently in use. The evaluation focused on two key aspects: the similarity between the metrics extracted by both methods, and the processing time required.

Results : The analysis revealed an average difference ranging from 0.02 to 0.03 seconds between the two approaches. Achieving a 0.01-second difference appears feasible with higher frame rate recordings (e.g., 100 Hz). In terms of processing time, CITRUS demonstrated an average gain of 77.5%, which is a notable advantage in real-world applications.

Discussion/Conclusion : HPE technologies appear to be a promising solution for automating performance analysis in hurdling. However, further refinements are still needed to ensure the tool's robustness in more complex scenarios such as competitions or multi-athlete environments.

Perspectives : This project represents a first step toward more automated and detailed analysis. Future developments will aim to generalize the tool to broader contexts and extract new performance indicators to support coaches and federations.

Keywords : Human pose estimation, automatic annotation, 110-meter hurdles.