

Psychologie et Réalité virtuelle – Reconnaissance des états mentaux par les postures dynamiques pour des Avatars enrichis sémantiquement avec des postures émotionnelles

Mots clefs : Computer vision, feature extraction, dynamic posture recognition, deep learning/machine learning

Lieu(x) : Université de Reims Champagne-Ardenne, Reims (France), campus Moulin de la Housse

Encadrement : Pr. Yannick Rémission *Email :* yannick.remion@univ-reims.fr

Dr. Stéphanie Prévost *Email :* stephanie.prevost@univ-reims.fr

Contexte et état de l'art

Une communication réussie repose en partie sur l'habileté à comprendre les états mentaux d'autrui. Différents indices peuvent être utilisés pour les identifier : les expressions faciales émotionnelles (EFEs) et les postures dynamiques. Le terme *posture dynamique* fait référence à la façon dont les positions et les mouvements de tout le corps (y compris la tête et les mains, mais à l'exclusion des EFE) sont utilisés pour transmettre des informations non verbales. Si les EFEs ont fait l'objet de nombreuses recherches, principalement les plus basiques (par exemple, la tristesse, la joie...), les postures dynamiques sont encore en jachère et ont été rarement étudiées.

Les approches quantitatives basées sur la capture de mouvement ou la détection du corps semblent prometteuses pour identifier les caractéristiques essentielles des postures dynamiques reflétant les états mentaux (Solanas & al., 2020). Combinées à la réalité virtuelle (RV), ces avancées en informatique affective ou cognitive ouvrent de nouvelles perspectives pour l'étude de la communication sociale, notamment par le biais d'agents virtuels ou d'avatars présentant des postures corporelles dynamiques en adéquations avec des attitudes faciales. Cependant, la coordination des mouvements corporels et des EFE reste un défi scientifique qui n'a été relevé que récemment par les chercheurs comme dans (Pavlakos & al., 2019).

Avec l'avènement ces dernières années des réseaux de neurones profonds, la reconnaissance d'objets a fait de grandes avancées et il en est de même pour le domaine de la reconnaissance de personne ou de posture (key point) (Jiang & al., 2023). De nombreuses méthodes et réseaux ont vu le jour pour reconnaître la pose des membres d'une personne (OpenPose (Zhe et al., 2018), CPN, AlphaPose (Haoshu et al., 2022), etc.). Cependant, ce sont principalement les estimations 2D de poses humaines qui ont connu une évolution significative, tandis que les estimations 3D présentent encore plusieurs défis à surmonter : Généralisation du modèle, robustesse à l'occlusion et efficacité des calculs. L'absence de dataset 3D annoté est aussi une des problématiques. En outre, ces estimations sont essentiellement liées à des activités et non à des émotions.

Le projet ANR POSTURE¹ concerne l'intégration de la psychologie et de la réalité virtuelle pour permettre la reconnaissance et la diffusion de postures dynamiques, à savoir l'ensemble des mouvements corporels (tête, corps, main) réalisés face à une émotion. Il fédère les compétences de différents laboratoires : en psychologie (C2S) et en informatique (ICUBE, LICIIS) et s'intéresse notamment au développement d'une approche innovante pour la captation du mouvement multimodale et multi-échelles (données hétérogènes) et de détection d'émotions, à des visés d'enrichissement sémantique émotionnel des avatars et à la formation

¹ <https://posture.univ-reims.fr/>

de corpus utiles pour la recherche en psychologie et en informatique (computer vision et Virtual reality). Plus précisément, il s'intéresse à des questions non résolues en psychologie et en RV. La combinaison de nouvelles technologies et découvertes en imagerie 3D (LICIIS, ICUBE) et en psychologie (C2S) est la clé de la réussite de ce projet. L'intégration d'informations sur les mouvements fins du corps, des mains, des doigts, de la tête et du visage pour générer des attitudes significatives n'a pas encore été entreprise dans les études sur la communication non verbale humaine ou la réalité virtuelle. Le projet vise à dépasser ces limites techniques et théoriques (Wolf, 2015, Solanas & al., 2020) en induisant et en capturant des expressions, à la fois au niveau du corps entier et du visage, sur des attitudes basiques et complexes.

Objectifs de la thèse

L'objectif principal de ce projet est de spécifier le rôle respectif des postures dynamiques et des EFEs dans la communication non verbale, considérant à la fois les émotions de base (e.g. colère, joie, peur) mais également des émotions plus complexes (e.g., fierté, pouvoir). Plus spécifiquement, sur le plan psychologique, il s'agira de savoir si les postures dynamiques ne font que renforcer la signification des EFEs ou si elles fournissent une signification qui leur est propre. Pour répondre à cette question, il est nécessaire de pouvoir dissocier dans un cadre expérimental les postures dynamiques des individus de leurs EFEs, distinction difficilement réalisable avec les outils actuellement disponibles. Les nouvelles technologies de captation de mouvements et de traitement de l'image associées à l'élaboration d'agents virtuels ouvrent des perspectives de recherches nouvelles et rendent l'étude du rôle de ces attitudes posturales dynamiques dans la communication non verbale possible. Sur le plan informatique, notre objectif est de déterminer « Quel est l'équilibre entre la précision spatiale et le caractère invasif du système pour une capture corporelle réussie résultant d'une génération d'état mental ? ». En d'autres termes, la précision accrue des systèmes de capture invasifs est-elle déterminante ou pouvons-nous parvenir à l'étiquetage de l'état mental avec des données moins précises ? Pouvons-nous estimer l'état mental avec la même approche indépendamment de la manière dont les données sont capturées ? Pour y répondre, une comparaison des différents modes de capture du corps sera effectuée, ce qui nous informera sur la précision spatiale nécessaire pour capturer les mouvements corporels représentatifs de l'état mental.

Dans le cadre du projet ANR Posture, le laboratoire LICIIS (Laboratoire en Informatique Calcul Intensif et Image pour la Simulation) s'intéresse plus particulièrement au développement d'une approche innovante pour la captation du mouvement corporel multimodale et multi-échelles (données hétérogènes) et la détection des émotions, à des visées d'enrichissement sémantique émotionnel des avatars.

Les études existantes sur les postures dynamiques se basent généralement sur des captations réalisées par des acteurs, lesquels connaissent et maîtrisent, du fait de leur métier, le langage corporel et savent le surjouer afin de communiquer plus aisément. Mais ce surjeu peut provoquer un biais dans les études posturales. Afin d'éviter ce risque, et contrairement à ces études, le projet POSTURE capture et étudiera des expressions spontanées chez des profanes, induites par des stimulus visuels émotionnels. Ces captures à la fois au niveau du corps entier et du visage seront réalisées grâce à trois systèmes de capture corporelle aux qualités écologiques² et de précisions différentes afin de tester différentes hypothèses.

Les différents dispositifs/systèmes de capture, allant du plus précis spatialement pour le système 1 au moins précis pour le système 3, et inversement pour l'aspect écologique, seront combinés afin de trouver le meilleur compromis entre l'augmentation de la précision de la capture et le traitement le plus complet possible de l'humain (corps, visage, mains, doigts), tout en étant le moins invasif possible. Les systèmes de suivi habituels pour les visages ou les EFE ne seront pas utilisés afin d'éviter toute interférence avec la capture

² Plus un système nécessite que l'utilisateur porte des équipements pouvant gêner ses mouvements, moins il est écologique.

du corps, qui est notre priorité. Ils pourront être associés à des caméras RGB-D pour la capture des visages et à des gants de données pour la capture des mains et des doigts.

La thèse proposée a pour but dans un premier temps d'analyser et de modéliser les données brutes synchronisées issues des captations, afin de représenter une posture dynamique générique (indépendante des attributs physiques de la personne) annotée par l'état mental (émotion) induit. Cette étape se fera en étroite collaboration avec deux ingénieurs recrutés sur ce projet, au C2S pour la mise en place du protocole d'induction des états mentaux lors des captations et au LICIIS, pour le développement de la couche logicielle et des outils nécessaires pour le système de capture hybride dédié au projet.

Dans un second temps, le but sera de pouvoir reconnaître, via un réseau de neurone profond, des états mentaux dans une nouvelle capture de corps complet afin de reconnaître les postures des individus de la vie réelle, ainsi que pour la génération de postures corporelles dynamiques induites. L'ensemble des captures effectuées et annotées constituera un corpus, qui, en combinaison avec la modélisation des postures dynamiques, permettra d'entraîner un réseau de neurones profonds pour cette détection.

Les résultats attendus du projet POSTURE seront des datasets ouverts, des données brutes et des avatars animés avec des postures dynamiques, avec ou sans EFE, et un système robuste de reconnaissance de l'état mental à partir des mouvements du corps. Ces résultats, sont destinés à servir de référence pour de futures études en psychologie, mais pourraient également avoir un impact important dans la réalité virtuelle avec interaction sociale, comme l'éducation et l'apprentissage (Oker et al., 2020), la réadaptation clinique (Oker et al., 2015) ou l'insertion professionnelle.

Financement

Cette thèse est financée par le projet ANR Posture (ANR-PRC, ANR-22-CE38-0010-01), sur un contrat doctoral de 3 ans qui débutera à partir de septembre 2024.

Profil recherché

Le candidat ou la candidate doit être titulaire d'un M2 ou d'un diplôme d'ingénieur en informatique. Il ou elle doit avoir des connaissances en computer vision et machine learning/deep Learning. De bonnes compétences en programmation sont indispensables, notamment en C/C++ et python.

Au-delà de ces prérequis académiques et techniques, de bonnes capacités de communication et de rédaction, en particulier en anglais, sont des avantages importants.

Enfin, le doctorat nécessite autonomie et respect de l'éthique et de la confidentialité des données de recherche.

Modalité de recrutement

Lettre de motivation et CV à envoyer à Yannick Rémyon (yannick.remyon@univ-reims.fr) et à Stéphanie Prévost (stephanie.prevast@univ-reims.fr) avant le 7 juin 2024.

Bibliographie

- Haoshu Fang and Jiefeng Li and Hongyang Tang and Chaoshun Xu and Haoyi Zhu and Yuliang Xiu and Yong-Lu Li and Cewu Lu, (2022) - *AlphaPose: Whole-Body Regional Multi-Person Pose Estimation and Tracking in Real-Time*, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol 45, pp 7157-7173.
- Jiang X, Hu Z, Wang S, Zhang Y. (2023) - *A Survey on Artificial Intelligence in Posture Recognition*. Computer Modeling in Engineering & Sciences. 2023 Apr 23;137(1):35-82. Doi : 10.32604/cmes.2023.027676. PMID : 37153533; PMCID : PMC7614502.
- Oker, A., ..., Brunet-Gouet, E. (2015) - *A virtual reality study of help recognition and metacognition with an affective agent*. International Journal of Synthetic Emotions (IJSE), 6(1), 60-73.
- Oker, A., Pecune, F. & Declercq, C. (2020) - *Virtual Tutor and Pupil Interaction: A study of the empathic feedback for learning process*. Education and Information Technologies, 25, 3643–3658. <https://doi.org/10.1007/s10639-020-10123-5>
- Pavlakos, G., Choutas, V., Ghorbani, N., Bolkart, T., Osman, A., Tzionas, D., Black, M. J. (2019) - *Expressive Body Capture: 3D Hands, Face, and Body From a Single Image*, The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 10975-10985.
- Solanas, M. P., Vaessen, M. J., de Gelder, B. (2020) - *The role of computational and subjective features in emotional body expressions*. Scientific reports, 10(1), 1-13.
- Wolf, K. (2015). Measuring facial expression of emotion. *Dialogues in Clinical Neuroscience*, 17, 457-462.
- Zhe Cao and Gines Hidalgo and Tomas Simon and Shih-En Wei and Yaser Sheikh, (2018) - *OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 43, pp 172-186.

Psychology and Virtual Reality- Posture-based recognition of mental states for semantically enriched Avatars with emotional postures

Key words: Computer vision, feature extraction, dynamic posture recognition, deep learning/machine learning

Workplace: Université de Reims Champagne-Ardenne, Reims (France), campus Moulin de la Housse

Supervision: Pr. Yannick Rémiot Email : yannick.remion@univ-reims.fr

Dr. Stéphanie Prévost Email : stephanie.prevost@univ-reims.fr

Context

Successful communication depends in part on the ability and the skill to understand the mental states of others. Different cues can be used to identify these: emotional facial expressions (EFEs) and dynamic postures. The term dynamic posture refers to the way in which positions and movements of the whole body (including the head and hands but excluding EFEs) are used to convey non-verbal information. While EFEs have been the subject of much research, mainly the more basic ones (e.g. sadness, joy...), dynamic postures are still lying fallow and have rarely been investigated.

Quantitative approaches based on motion capture or body sensing seem promising for identifying the essential features of dynamic postures reflecting mental states (Solanas & al., 2020). Combined with virtual reality (VR), these advances in affective or cognitive computing offer new perspectives for the study of social communication, notably through virtual agents or avatars presenting dynamic bodily postures in line with facial attitudes. However, the coordination of body movements and EFEs remains a scientific challenge that has only recently been addressed by researchers as in (Pavlakos & al., 2019).

With the advent of deep neural networks in recent years, object recognition has made great strides, as has the field of person or posture recognition (key point) (Jiang & al., 2023). Many new methods and networks have been developed to estimate and recognize the pose of a person's limbs (OpenPose (Zhe et al., 2018), CPN, AlphaPose (Haoshu et al., 2022), etc.). However, it is mainly the 2D estimations of human poses that have undergone significant evolution, while 3D estimations still present several challenges to overcome: model generalization, robustness to occlusion and computational efficiency. The absence of annotated 3D datasets is also a problem. Furthermore, these estimates are essentially linked to activities, not to emotions.

The ANR POSTURE³ project involves the integration of psychology and virtual reality to enable the recognition and dissemination of dynamic postures, i.e. all body movements (head, body, hand) made in response to an emotion. The consortium brings together the expertise of diverse laboratories: in psychology (C2S) and computer science (ICUBE, LICIIS). Its principal focus is on developing innovative approach for multimodal and multiscale motion capture (heterogeneous data) and emotion detection, in targeting the semantic emotional enrichment of avatars and in building useful datasets for research in psychology and computer science (computer vision and virtual reality). More specifically, the project addresses unresolved issues in psychology and VR. The combination of new technologies and discoveries in 3D imaging (LICIIS, ICUBE) and psychology (C2S) is key to the success of this project. The integration of information on fine

³ <https://posture.univ-reims.fr/>

movements of the body, hands, fingers, head, and face to generate meaningful attitudes has not yet been undertaken in studies of human non-verbal communication or virtual reality. The project aims to overcome these technical and theoretical limitations (Wolf, 2015, Solanas & al., 2020) by inducing and capturing expressions, both full-body and facial, on basic and complex attitudes.

Thesis objectives

The main aim of this project is to specify the respective roles of dynamic postures and EFEs in non-verbal communication, considering both basic emotions (e.g. anger, joy, fear) and more complex emotions (e.g. pride, power). More specifically, from a psychological perspective, the question is whether dynamic postures merely reinforce the meaning of EFEs, or whether they provide a meaning of their own. To answer this question, it is necessary to be able to dissociate the dynamic postures of individuals from their EFEs in an experimental setting. Currently available tools do not permit this distinction to be made. New technologies for motion capture and image processing, coupled with the development of virtual agents, offer new research perspectives, and make it possible to study the role of these dynamic postural attitudes in non-verbal communication. On the computational level, our objective is to determine the optimal balance between spatial accuracy and system invasiveness for successful body capture and subsequent mental state induction. In other words, is the increased precision of invasive capture systems decisive, or can we achieve mental state labeling with less precise data? Can we estimate mental state with the same approach regardless of how the data are captured? To address this question, a comparison of different modes of body capture will be conducted, which will inform us about the spatial precision required to capture body movements representative of mental state.

In the ANR Posture context, the LICIIS laboratory (Laboratoire en Informatique Calcul Intensif et Image pour la Simulation) is particularly involved in developing an innovative approach to multi-modal, multi-scale body movement capture (heterogeneous data) and emotion detection, with a view to semantically enriching the emotional content of avatars.

Previous studies on dynamic postures have been conducted using recordings made by actors, who are well-versed in body language and are adept at overplaying it to facilitate communication. However, this overacting may result in a bias in postural studies. To circumvent this risk, and in contrast to the aforementioned studies, the POSTURE project will capture and study spontaneous expressions in laypersons, induced by emotional visual stimuli. These whole-body and facial captures will be carried out using three body capture systems with different ecological⁴ qualities and accuracies, in order to test different hypotheses.

The different capture devices/systems, ranging from the most spatially precise for system 1 to the least precise for system 3, and vice versa for the ecological aspect, will be combined to find the best compromise between increased capture precision and the most complete possible treatment of the human (body, face, hands, fingers), while being as non-invasive as possible. The usual tracking systems for faces or EFEs will not be used to avoid interference with body capture, which is our priority. They can be combined with RGB-D cameras for face capture and data gloves for hand and finger capture.

The objective of the proposed thesis is to analyze and model the raw synchronized data from the captures in order to represent a generic dynamic posture (independent of the person's physical attributes) annotated by the mental state (emotion) induced. This stage will be carried out in close collaboration with two engineers recruited for the project, from C2S, who will set up the protocol for inducing mental states during captures,

⁴ The more a system requires the user to wear equipment that can impede movement, the less ecological it is.

and from LICIIS, who will develop the software layer and tools required for the hybrid capture system dedicated to the project.

In a second stage, the objective will be to be able to recognize, via a deep neural network, mental states in a new full-body capture in order to recognize the postures of real-life individuals, as well as for the generation of induced dynamic body postures. The set of annotated captures will constitute a corpus, which, in combination with the modeling of dynamic postures, will enable a deep neural network to be trained for this detection.

The expected results of the POSTURE project will be open datasets, raw data and animated avatars with dynamic postures, with or without EFE, and a robust system for recognizing mental state from body movements. These results are intended to serve as a reference for future studies in psychology, but could also have a major impact in virtual reality with social interaction, such as education and learning (Oker et al., 2020), clinical rehabilitation (Oker et al., 2015) or professional integration.

Funding

This thesis is funded by the ANR Posture project (ANR-PRC, ANR-22-CE38-0010-01), on a 3-year doctoral contract starting in September 2024.

Candidate Profile required

The candidate must have a M2 or an engineering degree in computer science. He or she should have knowledge of computer vision and machine learning/deep learning. Good programming skills are essential, particularly in C/C++ and Python.

Beyond these academic and technical prerequisites, good communication and writing skills, particularly in English, are important advantages.

Finally, a doctorate requires autonomy and respect for ethics and the confidentiality of research data.

Recrutement procedure

Letter of application and CV to be sent to Yannick Rémyon (yannick.remyon@univ-reims.fr) et à Stéphanie Prévost (stephanie.prevost@univ-reims.fr) before 7 June 2024.

Bibliographie

Haoshu Fang and Jiefeng Li and Hongyang Tang and Chaoshun Xu and Haoyi Zhu and Yuliang Xiu and Yong-Lu Li and Cewu Lu, (2022) - *AlphaPose: Whole-Body Regional Multi-Person Pose Estimation and Tracking in Real-Time*, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol 45, pp 7157-7173.

Jiang X, Hu Z, Wang S, Zhang Y. (2023) - *A Survey on Artificial Intelligence in Posture Recognition*. Computer Modeling in Engineering & Sciences. 2023 Apr 23;137(1):35-82. Doi : 10.32604/cmescs.2023.027676. PMID : 37153533; PMCID : PMC7614502.

Oker, A., ..., Brunet-Gouet, E. (2015) - *A virtual reality study of help recognition and metacognition with an affective agent*. International Journal of Synthetic Emotions (IJSE), 6(1), 60-73.

- Oker, A., Pecune, F. & Declercq, C. (2020) - *Virtual Tutor and Pupil Interaction: A study of the empathic feedback for learning process*. *Education and Information Technologies*, 25, 3643–3658. <https://doi.org/10.1007/s10639-020-10123-5>
- Pavlakos, G., Choutas, V., Ghorbani, N., Bolkart, T., Osman, A., Tzionas, D., Black, M. J. (2019) - *Expressive Body Capture: 3D Hands, Face, and Body From a Single Image*, The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 10975-10985.
- Solanas, M. P., Vaessen, M. J., de Gelder, B. (2020) - *The role of computational and subjective features in emotional body expressions*. *Scientific reports*, 10(1), 1-13.
- Wolf, K. (2015). Measuring facial expression of emotion. *Dialogues in Clinical Neuroscience*, 17, 457-462.
- Zhe Cao and Gines Hidalgo and Tomas Simon and Shih-En Wei and Yaser Sheikh, (2018) - *OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields*, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, pp 172-186.