In-bed action recognition for clinical diagnosis support: A two-stage, 3D motion capture and skeleton action recognition based approach

Tamás Karácsony[1,2,3], João Paulo Silva Cunha[1,2]
[1] Center for Biomedical Engineering Research, INESC TEC, Porto, Portugal
tamas.karacsony@inesctec.pt and joao.p.cunha@inesctec.pt

[2]Faculty of Engineering, University of Porto, Porto, Portugal

[3]Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, USA

Abstract

Many clinical scenarios involve in-bed monitoring both at the hospital or even at home giving in-bed action recognition technology a relevant position when we aim at behavior/motion capture monitoring. A wide range of potential clinical applications involve from intensive care and neuro-critical infirmary to sleep monitoring at home. One such relevant application is the epileptic seizure motion quantification for diagnosis support, in the neurology area, that is known to be a highly demanding environment to perform motion quantification for clinical support [1]. Although promising, action recognition in such clinical environments pose several challenges that need to be tackled in order to realize the full potential of this technology for routine clinical usage in the future.

Epileptic seizure semiology (the study of clinical signs during seizures to support diagnosis and therapeutic decisions) is currently performed through video monitoring data. There are other approaches utilizing techniques from classical computer vision, image or video classification, keypoint or keypoint stream classification, however only a few action recognition approaches exist [2, 3]. Most of these are end-to-end approaches. However, there is a need for quantification, and explainability of the classification to identify the movements contributing the most to the classification of the seizure, both from the clinical and technical perspective.

To address the challenges of in-bed monitoring and action recognition systems, our group is developing approaches such as BlanketGen [4], to address the challenge of blanket occlusion present in these scenarios. To evaluate the improvements of approaches utilizing BlanketGen and other systems an action recognition dataset was developed of these scenarios dubbed as BlanketSet [5].

In order to recognize actions performed in bed, a two-step process is being developed. The proposed first stage of our approach extracts 3D motion capture (MoCap) data to track the full body movements of a person in the bed. Then the second stage uses this for skeleton action recognition to identify the specific actions that are being performed by the person in bed.

During the development we have identified the need for better quality data to improve the accuracy and reliability of our approach. To address this need, we have initiated the recording of a new 4K-RGB-D dataset of epileptic seizures and a Movement of Interest (MOI) simulation dataset, building on top of our previously developed collaborative development framework with our clinical partners [6]. This framework allows us to collect data in a real-world

This work was partially funded by Fundação para a Ciência e a Tecnologia under the scope of the CMU Portugal program (Ref PRT/BD/152202/2021).
We thank members of the Epilepsy Center, Department of Neurology, University of Munich, Munich, Germany for sharing their time and expertise to improve this work, namely Nicholas Fearns, Dr. Anna Mira Loesch Biffar, PD Dr. Dr. Christian Vollmar, Prof. Dr. med. Jan Rémi, Prof. Dr. Soheyl Noachtar.
We thank members of the Neurophysiology Unit, Neurology Department, Centro Hospitalar Universitário de São João, E.P.E., Porto, Portugal, for sharing their time and expertise to improve this work, namely Dr. Ricardo Rego.

environment and to collaborate with clinicians to ensure that the data is collected in a way that is relevant to the clinical needs. This better quality data, with higher resolution is required for the quantification of full body movements, especially it is vital to track the hands and face. With this 4K resolution, the hands and face are still represented with a decent number of pixels, - on the scale of 150x150 for hands and 300x300 for face, from a viewpoint of a common clinical monitoring perspective, this is naturally 4 times the pixels compared to 1080p videos from the same viewpoint.

The expected results of using 3D MoCap for action recognition on 4K-RGB-D videos are that it will be able to quantify not just the large movements, but the smaller details of the scene, which can contribute to a better action recognition approach. It can also contribute to epileptic seizure classification with greater detail and explainability. Currently the data acquisition is ongoing, and promising initial MoCap results have been obtained on these 4K videos. Now, a temporo-spatially stable 3D MoCap system is being developed, and it is expected to have a prototype system available soon.

The future work for this project includes testing the system on a new dataset that is being acquired and to improve the accuracy of the MoCap system and making it more robust to noise. Moreover, explore the explainability of the approach, as it allows users, clinicians to understand how the system is making its decisions, where the proposed 2-stage approach opens up the opportunity to provide quantitative explanations of the action classification based on the contributing movements.

We believe that our approach has several advantages over other approaches developed in scope of epileptic seizure classification for clinical in bed action recognition. First, our approach is aiming to capture the smaller details of the scene, which can be important for accurate action recognition. Second, our approach aims to provide explainability, which is cruicial for clinical diagnosis support applications where it is important to understand how the system is making its decisions.

In conclusion, the proposed 2-stage concept is a promising approach that has the potential to improve the accuracy and explainability of action recognition systems utilized for in-bed action recognition based clinical diagnosis support, by providing the quantified and interpretable latent space of 3D MoCap for video based action classification.

References:

[1] M. do Carmo Vilas-Boas and J. P. S. Cunha, "Movement Quantification in Neurological Diseases: Methods and Applications," *IEEE Rev. Biomed. Eng.*, vol. 9, pp. 15–31, 2016

[2] T. Karacsony, A. M. Loesch-Biffar, C. Vollmar, S. Noachtar, and J. P. S. Cunha, "A Deep Learning Architecture for Epileptic Seizure Classification Based on Object and Action Recognition," in IEEE *ICASSP 2020*, Apr. 2020, pp. 4117–4121.

[3] T. Karácsony, A. M. Loesch-Biffar, C. Vollmar, J. Rémi, S. Noachtar, and J. P. S. Cunha, "Novel 3D video action recognition deep learning approach for near real time epileptic seizure classification," *Sci. Rep. 2022 121*, vol. 12, no. 1, pp. 1–13, Nov. 2022

[4] J. Carmona, T. Karácsony, and J. P. S. Cunha, "BlanketGen - A synthetic blanket occlusion augmentation pipeline for MoCap datasets," *arxiv:2210.12035*, Oct. 2022

[5] J. Carmona, T. Karácsony, and J. P. S. Cunha, "BlanketSet - A clinical real word action recognition and qualitative semi-synchronised MoCap dataset," *arXiv:2210.03600*, Oct. 2022

[6] T. Karácsony, A. M. Loesch-Biffar, C. Vollmar, S. Noachtar, and J. P. S. Cunha, "Deepepil: Towards an Epileptologist-Friendly AI Enabled Seizure Classification Cloud System based on Deep Learning Analysis of 3D videos,"IEEE *BHI 2021*