***Irene Ballester and Martin Kampel (2024): Measuring dementia behaviours through depth sensors**. In: Proceedings of the Joint visuAAL-GoodBrother Conference on trustworthy video- and audio-based assistive technologies – COST Action CA19121 - Network on Privacy-Aware Audio- and Video-Based Applications for Active and Assisted Living*

Measuring dementia behaviors through depth sensors

Irene Ballester, Martin Kampel
Computer Vision Lab, TU Wien
irene.ballester@tuwien.ac.at, martin.kampel@tuwien.ac.at

Abstract

Human Behavior Modelling aims to analyze human behaviors using visual or 3D data. In dementia, behavioral changes are strongly correlated with cognitive decline, making their monitoring essential. To address privacy concerns, especially important for vulnerable populations, this study uses depth maps instead of RGB images. We present a novel Human Pose Estimation (HPE) method using point cloud sequences derived from depth data and discuss domain gaps between benchmark and real-world datasets.

Introduction

Dementia affects millions, manifesting not just as memory loss but through behavioral changes (WHO, 2021). Monitoring these changes is crucial for tracking progression and tailoring care (Ballester et al., 2024c). To address privacy concerns associated with RGB cameras, we utilize depth sensors that capture 3D data without identifiable features. Our aim is to develop a robust framework for measuring dementia behaviors. In this work, we present a novel HPE method that processes point cloud sequences, and discuss domain adaptation techniques to improve performance in real-world settings.

Methodology

HPE from Point Cloud Sequences (SPiKE)

SPiKE takes point cloud sequences and predicts the 3D coordinates of the joints (Ballester et al., 2024a). Unlike previous works that process timestamps independently (Zhou et al., 2020), SPiKE leverages temporal information by using a Transformer to encode the spatio-temporal structure of the sequence.

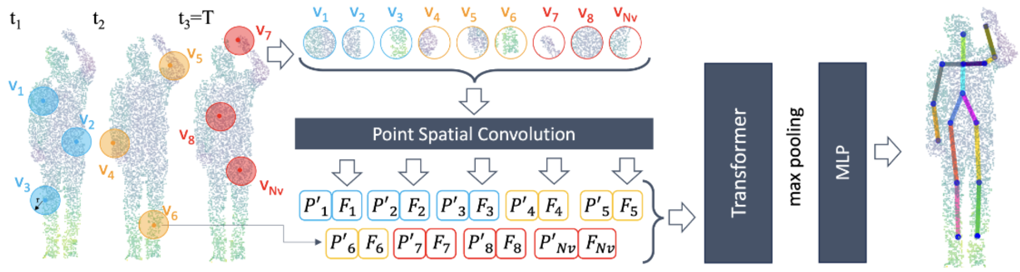

Figure 1. SPiKE pipeline: each point cloud is divided into local volumes for feature extraction using point spatial convolution, ensuring efficient processing while preserving spatial integrity.

In Ballester et al. (2024a), we show that SPiKE outperforms existing state-of-the-art methods on the ITOP benchmark with an mAP of 89.19%, Ablation studies highlight the benefits of using sequence information and maintaining spatial structure. Future work will assess SPiKE on real-world data.

BAD Dataset and domain gaps

The Bathroom Activities Dataset (BAD) consists of depth sequences from 19 dementia patients performing various activities and presents real-world challenges like occlusions, variable sensor angles, and unbalanced class distributions, making it ideal for testing robust models. Benchmarks often fail to reflect these complexities, resulting in significant performance gaps when models trained on controlled data are applied to real-world environments (Ballester and Kampel, 2024b; Fan et al., 2021).

Domain adaptation techniques

To bridge this gap, we explore Test-Time Training (TTT) (Sun et al., 2020) for point cloud sequences. The model's encoder is trained in the source domain with two heads: one for the primary task (e.g., HAR or HPE) and another for a self-supervised task, optimizing a combined loss. In the target domain, the encoder is fine-tuned while both heads remain frozen, allowing adaptation to target data using self-supervised learning. The challenge is finding an self-supervised task that effectively adapts the feature extractor to the target domain.

Conclusions

Robust and effective measurement of dementia behavior from depth maps requires not only the development of HAR and HPE models, but also domain adaptation strategies for real-world applicability. This study presents a new HPE model and discusses the application of TTT to point cloud sequences, with the ultimate goal of achieving robust performance in practical settings.

Acknowledgments

This publication is based upon work from COST Action GoodBrother - Network on Privacy-Aware Audio- and Video-Based Applications for Active and Assisted Living (CA19121), supported by COST (European Cooperation in Science and Technology). This publication is part of the visuAAL project on Privacy-Aware and Acceptable Video-Based Technologies and Services for Active and Assisted Living. This project has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 861091.

References

Ballester, I., Peterka, O., & Kampel, M. (2024a). SPiKE: 3D Human Pose from Point Cloud Sequences. arXiv preprint arXiv:2409.01879

Ballester, I., & Kampel, M. (2024b). Action recognition from 4D point clouds for privacy-sensitive scenarios in assistive contexts. In International Conference on Computers Helping People with Special Needs (pp. 359-364). Cham: Springer Nature Switzerland.

Ballester, I., Gall, M., Münzer, T., & Kampel, M. (2024c). Vision-Based Toilet Assistant for People with Dementia in Real-Life Situations. In 2024 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops) (pp. 141-147). IEEE.

Fan, H., Yang, Y., & Kankanhalli, M. (2021). Point 4D Transformer Networks for Spatio-Temporal Modeling in Point Cloud Videos. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 14204-14213).

Sun, Y., Wang, X., Liu, Z., Miller, J., Efros, A., & Hardt, M. (2020). Test-time training with self-supervision for generalization under distribution shifts. In International Conference on Machine Learning (pp. 9229-9248). PMLR.

WHO (2021) Global status report on the public health response to dementia.

Zhou, Y., Dong, H., & El Saddik, A. (2020). Learning to estimate 3D human pose from point cloud. IEEE Sensors Journal, 20(20), 12334-12342.