

Efficient Refinement of Human Pose Estimation for Industry 5.0

Enrico Martini^{1,2}, Nicola Bombieri¹

¹ Dept. of Engineering for Innovation Medicine, University of Verona, Italy

² Dept. of Computer Science, University of Verona, Italy

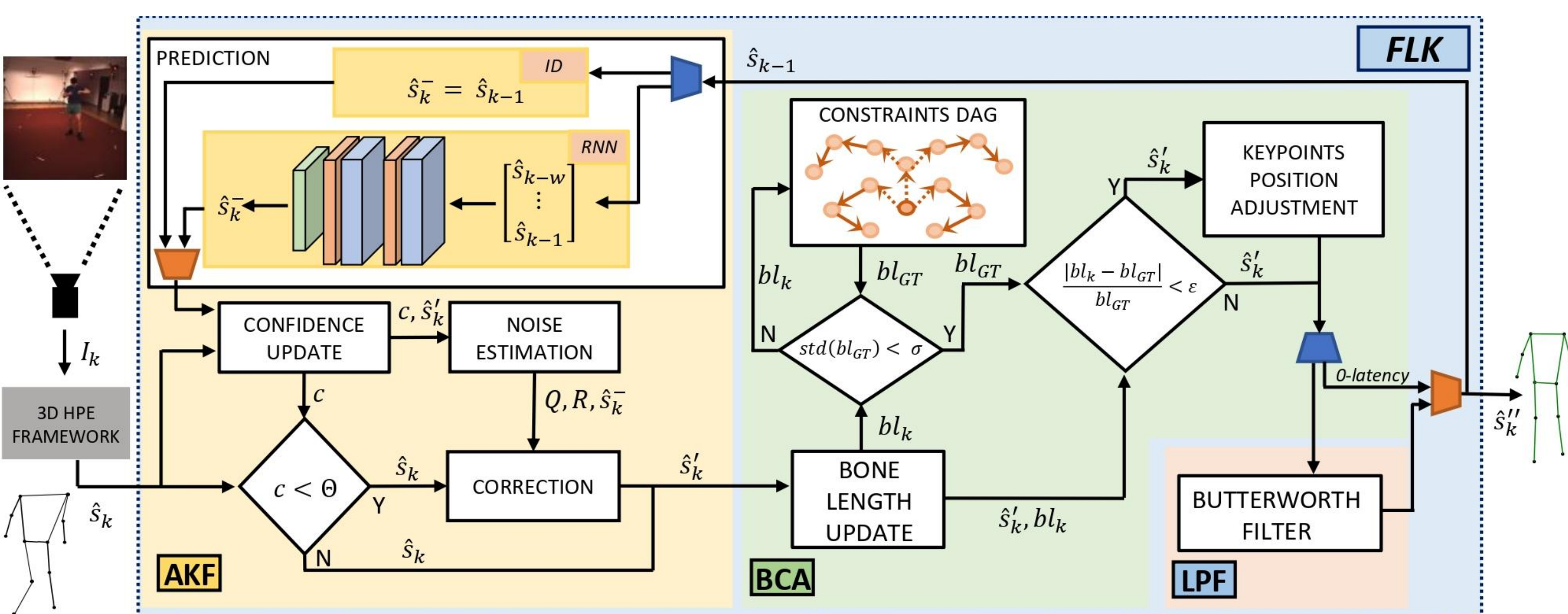
enrico.martini@univr.it - nicola.bombieri@univr.it

Introduction

This thesis tackles challenges in **Human Pose Estimation (HPE)**, such as noise, occlusions, and computational limitations, by developing real-time **filtering techniques** that combine learned models with traditional methods. These methods are applied to a distributed 3D HPE system for industrial telemonitoring that captures multi-view poses using **edge devices** with advanced filtering and clustering algorithms. In human-robot interaction, a filtering pipeline restores incomplete 3D poses from RGB-D cameras, mitigating occlusions and enabling **collision prediction**. This work advances markerless motion capture with practical, real-world applications.

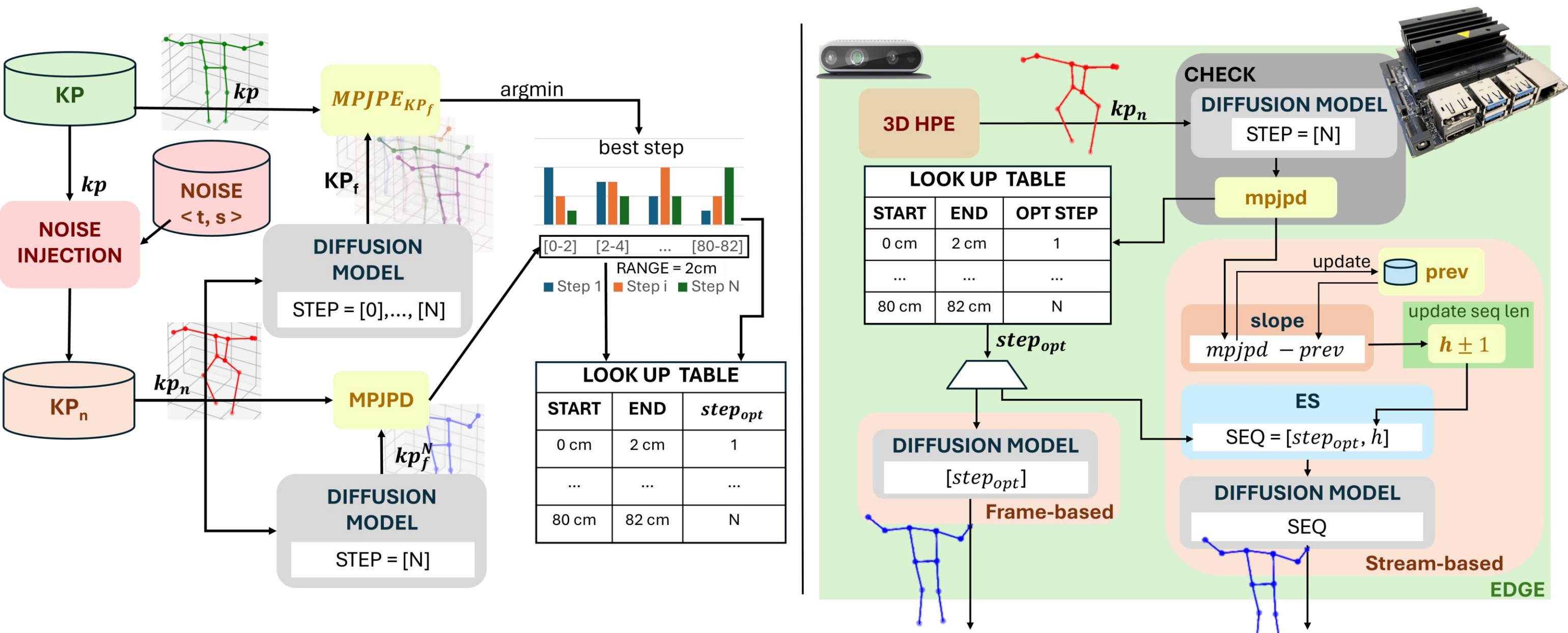
FLK: a real-time filter for HPE with learned kinematics

There is increased interest in using 3D HPE for safety-critical systems. However, these neural networks often struggle with inaccuracies and must meet strict **real-time performance standards**. To tackle this, we propose FLK (Filter with Learned Kinematics), which refines 3D human motion data in real-time with low latency. FLK combines a **Kalman filter** with a **low-pass filter**, utilizing a **recurrent neural network** to learn the motion model and incorporating biomechanical constraints for spatial coherence. This approach effectively addresses noise issues like jittering and dropped frames. Testing on multiple datasets and different 3D pose estimation backbones showed accuracy improvements of up to 140 mm with non-Gaussian noise and 53 mm with missing data [1].



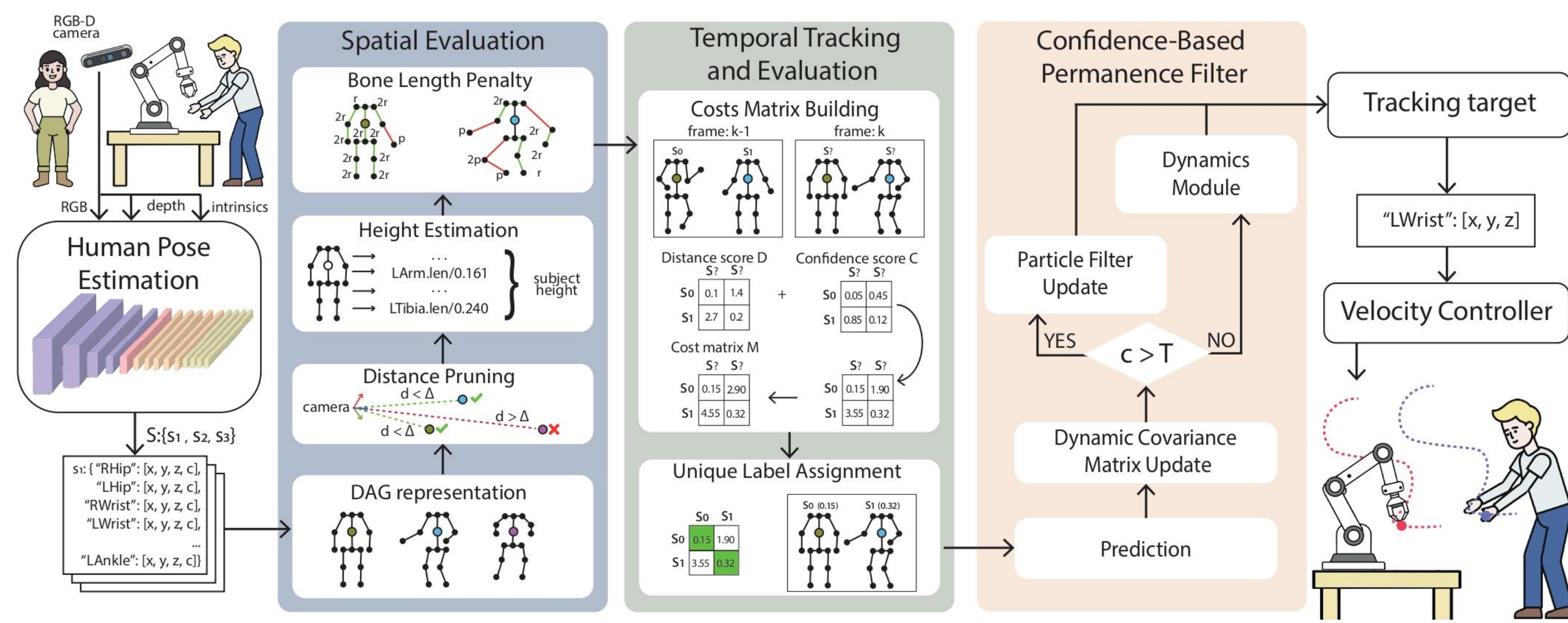
Motion Filtering with DDPM at the edge

Current filtering models demand high computational resources, limiting their use on edge devices. We propose a real-time filtering approach using optimized **denoising diffusion probabilistic models (DDPM)**. We analyze DDPM's adaptability to noise variations through micro-benchmarking and determine optimal configurations. Our solution employs two DDPM configurations to address different use cases, evaluated on edge devices with both synthetic and real noise from state-of-the-art HPE systems. Results show significant filtering improvements with **minimal computational overhead** [2].



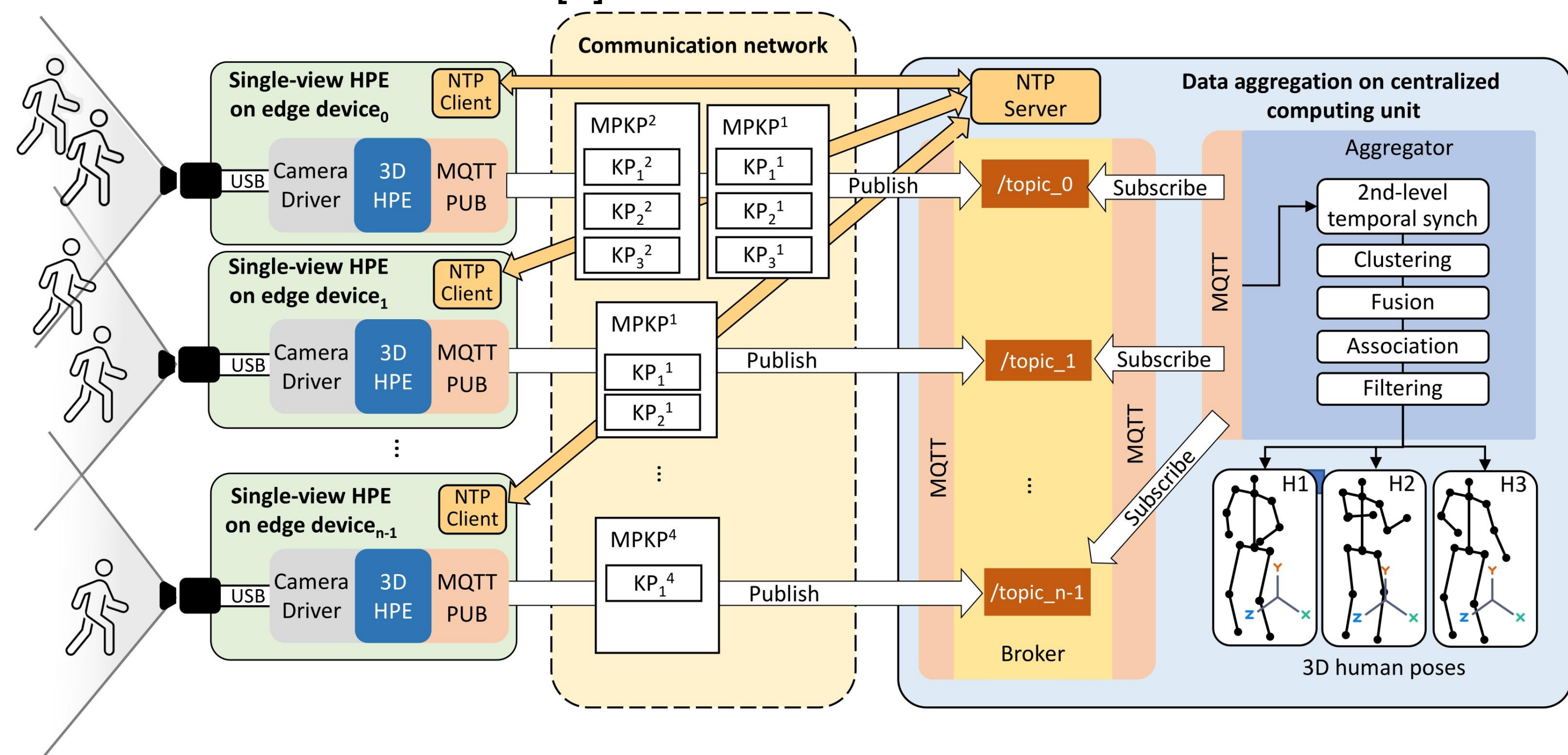
Efficient Multi-person Tracking for Human-Robot Interaction

Pursuing natural and marker-less **human-robot interaction** has been a long-standing focus of robotics research. Markerless approaches promise an improved user experience, but state-of-the-art struggles with the challenges posed by intrinsic errors in HPE and depth cameras. These errors can lead to issues such as robot jittering, which can significantly impact the trust users have in **collaborative systems**. We propose a filtering pipeline that refines incomplete 3D human poses from an HPE backbone and a single RGB-D camera to address these challenges, solving for **occlusions** that can degrade the interaction. Experimental results show that using the proposed filter leads to more consistent and noise-free motion representation, reducing unexpected robot movements and **enabling smoother interactions** [3].



Real-time Multi-camera 3D HPE for Industrial Applications

Implementing a markerless mocap system in **industrial settings** is challenging due to occlusion issues that require **multiple cameras** and synchronized HPE software nodes on resource-limited edge devices. We introduce a distributed 3D HPE platform, which uses a set of HPE software nodes, one for each camera, to collect pose data from different viewpoints. A centralized aggregator merges this information in real-time through **filtering, clustering, and association algorithms**, addressing communication issues like delays and bandwidth variability with a two-level synchronization process. We provide evaluation results from a case study that compares its accuracy and scalability to state-of-the-art methods [4].



References

1. E. Martini, M. Boldo, and N. Bombieri. "FLK: A filter with learned kinematics for real-time 3D human pose estimation". In: Signal Processing 224 (2024).
2. C. Bozzini, M. Boldo, E. Martini, et al. "A Real-time Filter for Human Pose Estimation based on Denoising Diffusion Models for Edge Devices". In: IEEE/RSJ IROS 24 (2024).
3. E. Martini, H. Parekh, S. Peng, et al. "A Robust Filter for Marker-less Multi-person Tracking in Human-Robot Interaction Scenarios". In: IEEE ROMAN 24 (2024)
4. M. Boldo, M. De Marchi, E. Martini, et al. "Real-time multi-camera 3D human pose estimation at the edge for industrial applications". In: Expert Systems with Applications 252 (2024).