

Gesture-based Interaction with a Smart Room

Jan Richarz

Intelligent Systems Group, Robotics Research Institute, TU Dortmund University, Dortmund, Germany

Introduction

- The "smartness" of a smart environment is not defined by the amount of built-in technology, but the intuitiveness and naturalness of control and access to this technology.

- **Goal:** Enable natural interaction with a smart room using 3D gestures.

⇒ **Requirements:** User-independency, real-time responsivity

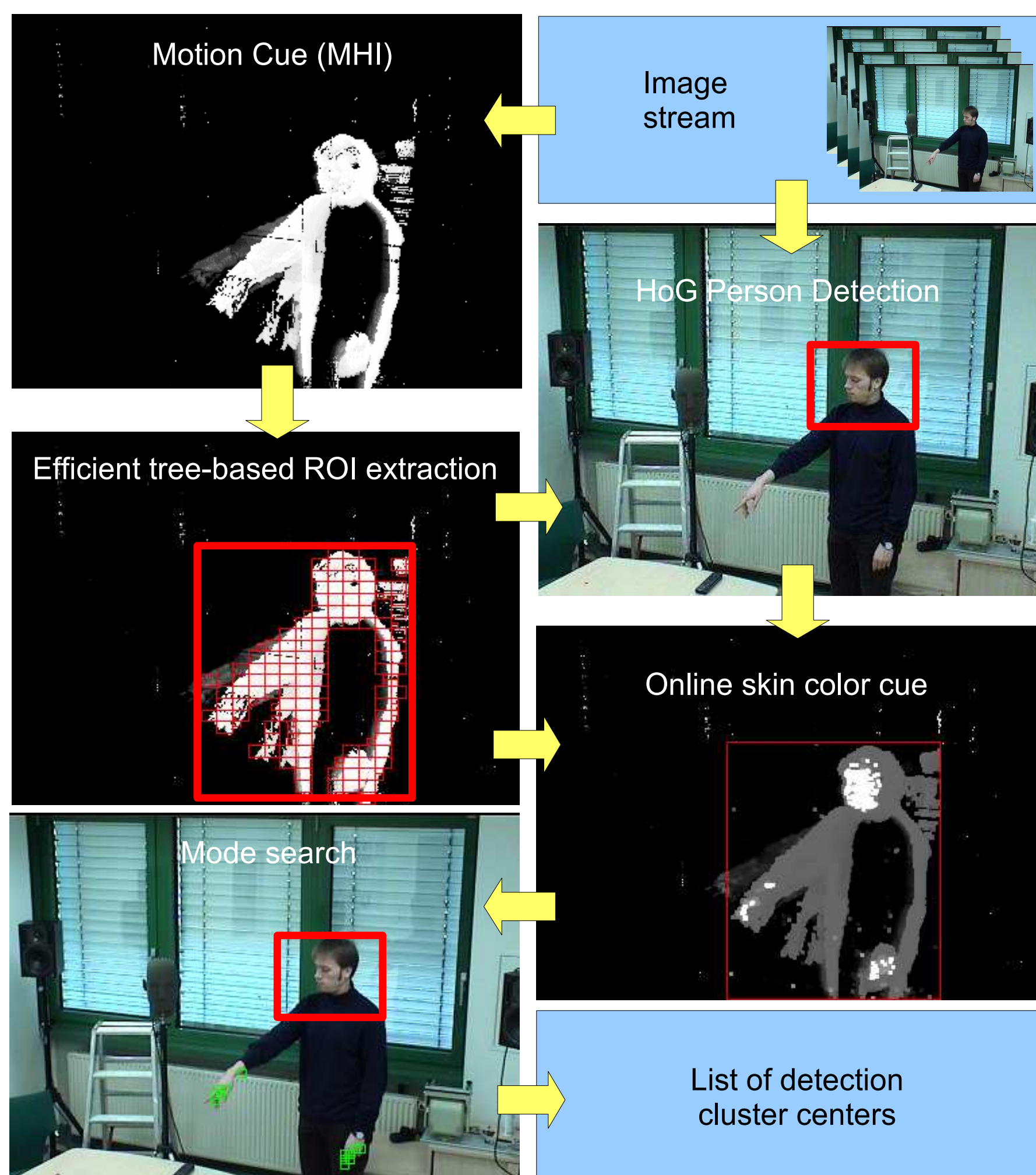
⇒ **Setup:** Smart environment FINCA, active multi-camera system, microphone arrays

Single-view Processing

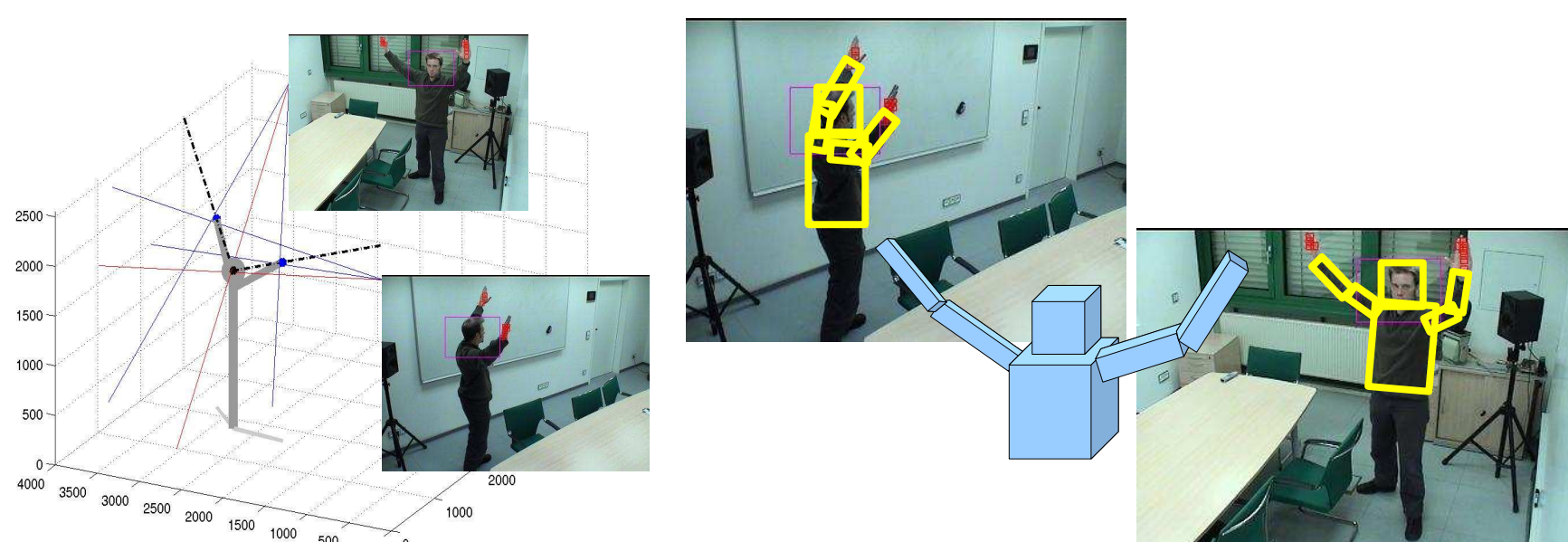
- Process image streams individually

⇒ parallelizable, scalable, reduced data amount

- Head and hand positions sufficient for pointing direction estimation.



3D Combination



- Combine individual results via ray casting → 3D pointing direction.
- For different types of gestures, analysis of spatio-temporal body pose is necessary.
- **Possible approach:** Simultaneous, view-coupled model fitting in (unsynchronized) multi-camera setup.

Spotting meaningful gestures

- Classical spotting / classification approaches: HMM, Bag of (Pose) Words, spatio-temporal interest points, ...

- Audio as additional cue

- Explicit command word recognition provides contextual information.
- Simultaneous occurrence of utterance and gesticulation can indicate interesting action.

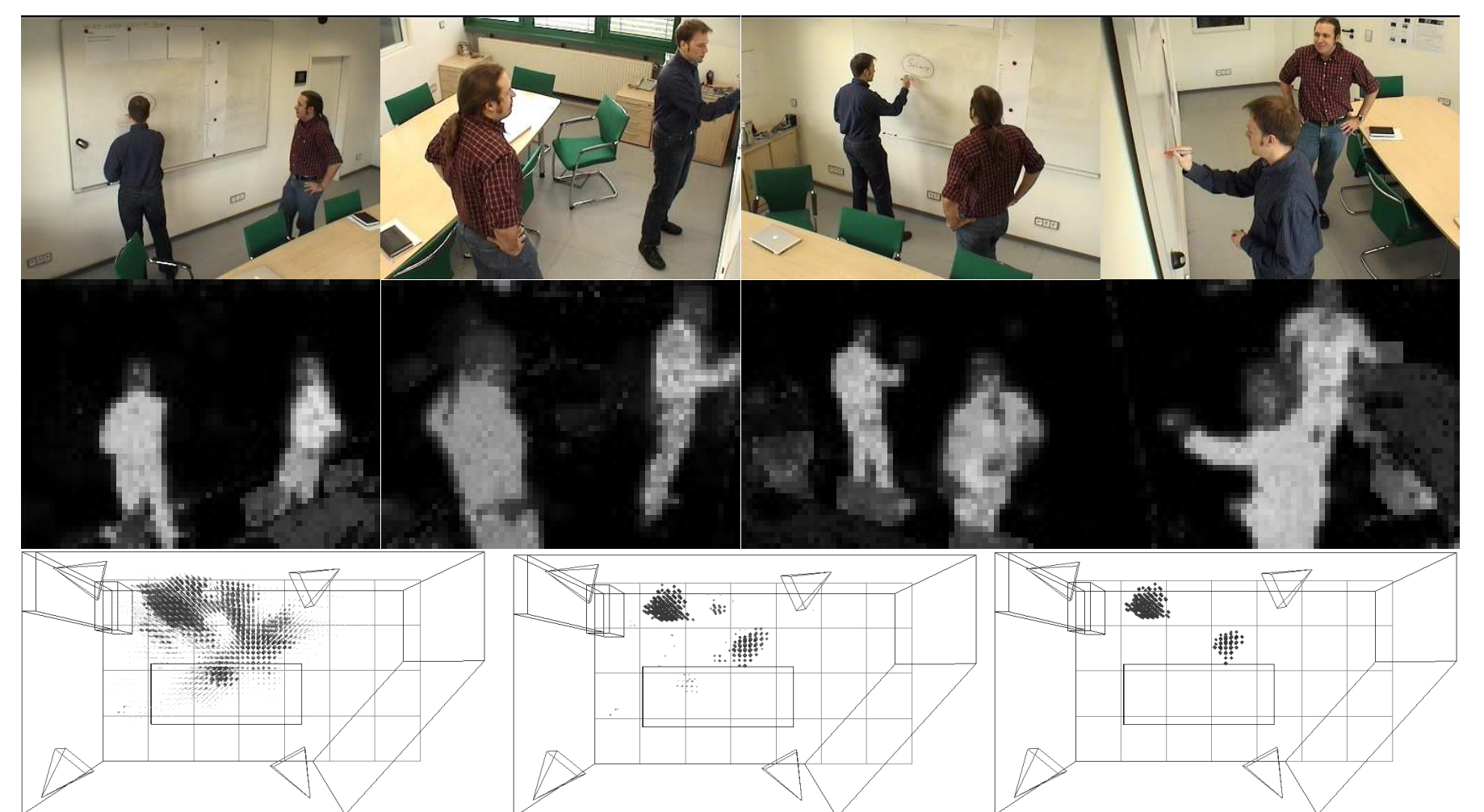
- Scene context

- Interaction with persons or objects
- Environmental / system state

⇒ Is it possible to automatically learn meaningful gesture / utterance / system state combinations?

Application

- 3D pointing gestures already used to switch lights inside the FINCA.
- Integration in attention-based active multi-camera control (with Boris Schauerte).



- Gestures (e.g. periodic actions, waving) as bottom-up saliency feature.
- Top-down attention modulation (e.g. pointing gestures to steer attention / modulate saliency).
- Active camera control → View optimization / selection for improved model and recognition quality.

Goals

- Develop methods for analysing human gestures and body poses.
- Facilitate intuitive human-computer interfaces for interaction with smart environments.
- Explicit focus on (active) multi-camera setups.
- Strong emphasis on methods feasible for real-time implementation.