RHM-HAR-SK:
A MULTIVIEW DATASET WITH SKELETON DATA FOR AMBIENT ASSISTED LIVING RESEARCH

Mohamad Reza Shahabian Alashti,
Mohammad Hossein Bamorovat Abadi,
Patrick Holthaus,
Catherine Menon,
And
Farshid Amirabdollahian

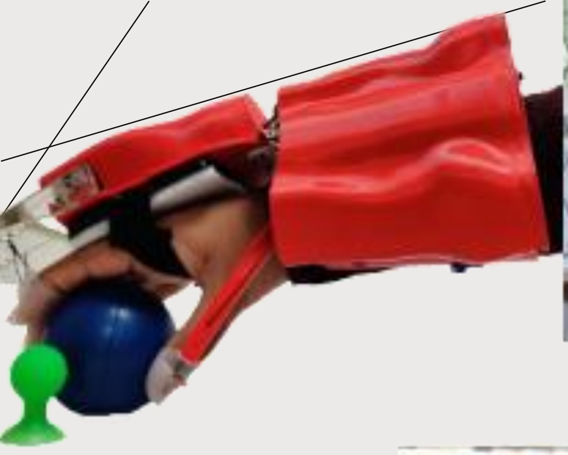
Robotics Research Group,
School of Engineering and Computer Science
University of Hertfordshire,
Hatfield, United Kingdom

Email: {m.r.shahabian , m.bamorovat, p.holthaus, c.menon, f.amirabdollahian2}@herts.ac.uk

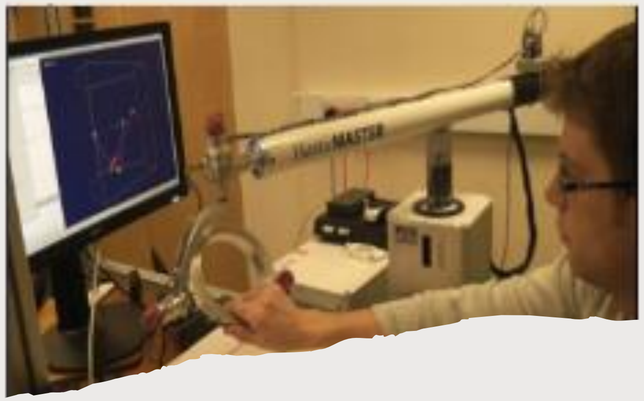
PRESENTER'S BIOGRAPHY

Mohamad Reza
Shahabian Alashti

I am a researcher with a passion for robotics, machine learning, and computer vision. I am grateful for the opportunity to pursue my PhD in Computer Science at the University of Hertfordshire, where I am currently working on skeleton-based human activity recognition using multiple cameras. Work in this area has the potential to revolutionize the way we interact with technology and could lead to new innovations in fields such as healthcare, sports, and entertainment. I am proud of my accomplishments; I remain committed to continued learning and growth in my field. I believe that my work has the potential to make a positive impact on society, and I am excited to continue exploring new applications for robotics and machine learning.



Robots in UH **Robot House** and Robotics Research Group



CONTENT

- Introduction
- Background works and challenges
- RHM-HAR-SK dataset
- Qualitative and quantitative results
- Conclusions

INTRODUCTION

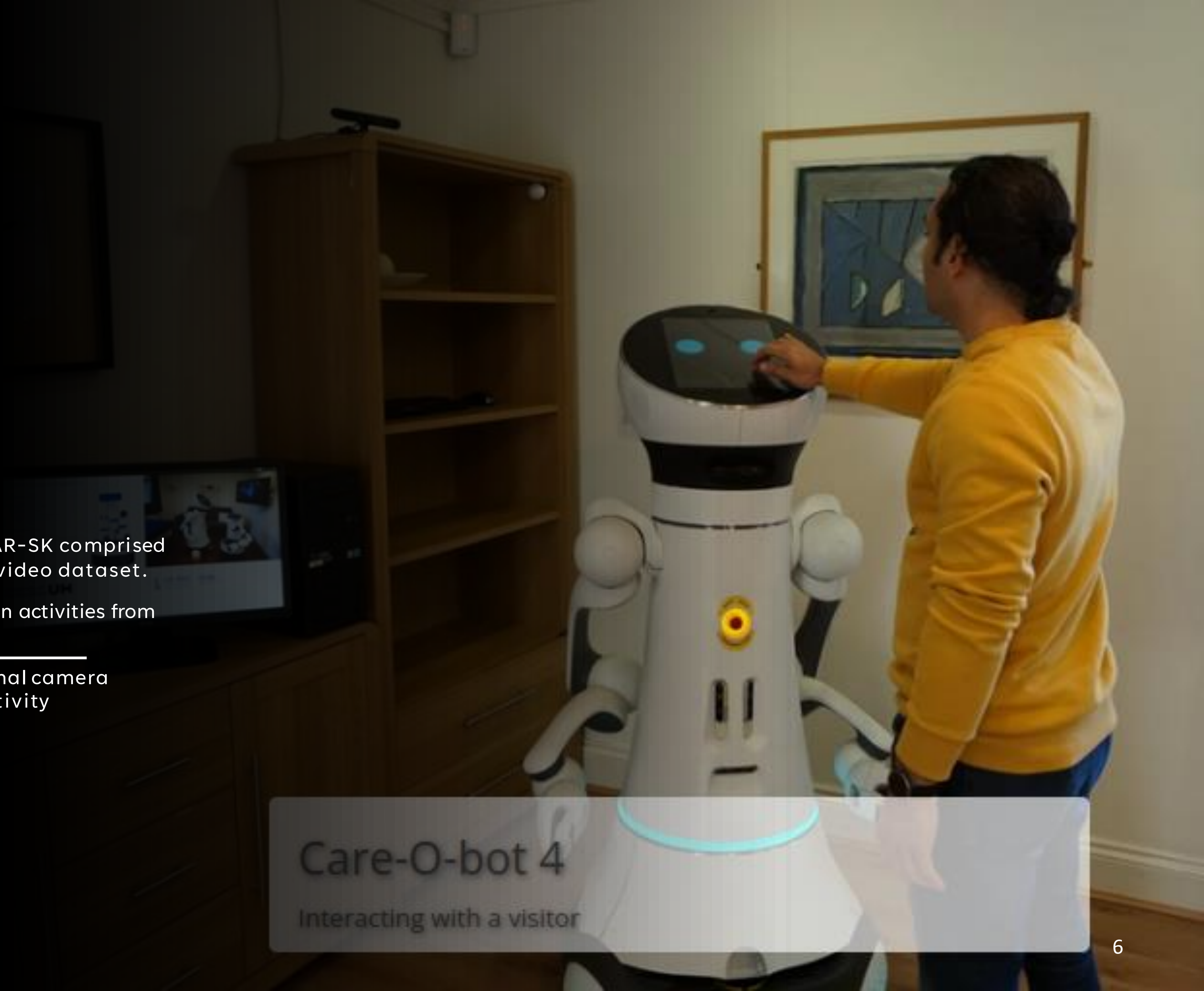
- Assistive robots are crucial for supporting older people with daily living.
- Provide effective assistance, such robots need to monitor people's activities.
- Human activity recognition (HAR) enables robots to understand and respond to human users' needs and activities.
- Skeleton-based Activity Recognition (SAR) algorithms can capture fine-grained details of human motion.
- SAR algorithms offer accurate and nuanced information about the actions performed by an individual.
- Combining the robot's view with external cameras can increase the overall robustness of activity recognition.

TWO MAIN CONTRIBUTIONS TO HUMAN ACTIVITY DETECTION IN AMBIENT ASSISTED LIVING SCENARIOS.

Firstly, we present the novel dataset RHM-HAR-SK comprised of human skeleton data on top of an existing video dataset.

The dataset contains extracted skeletons of human activities from four different perspectives.

Secondly, we demonstrate how using additional camera perspectives enhances an assistive robot's activity recognition pipeline.



BACKGROUND

HAR TECHNOLOGIES

Vision-based and sensor-based methods are commonly used for HAR.

- **Vision-based** HAR methods rely on 2D or 3D video data acquired by various devices.
- **Sensor-based** recognition relies on additional sensors like GPS, gyroscopes, accelerometers, or magnetometers.
- **Our approach** uses multiple cameras to fuse recognition results without relying on external sensory technology.
- **Our approach** relies on derived data using a pose extraction method to generate skeleton-based representations of human activities in a domestic environment.

BACKGROUND

HUMAN ACTIVITY RECOGNITION IN ASSISTIVE ROBOTICS

Few studies specifically focus on Ambient Assisted Living (AAL).

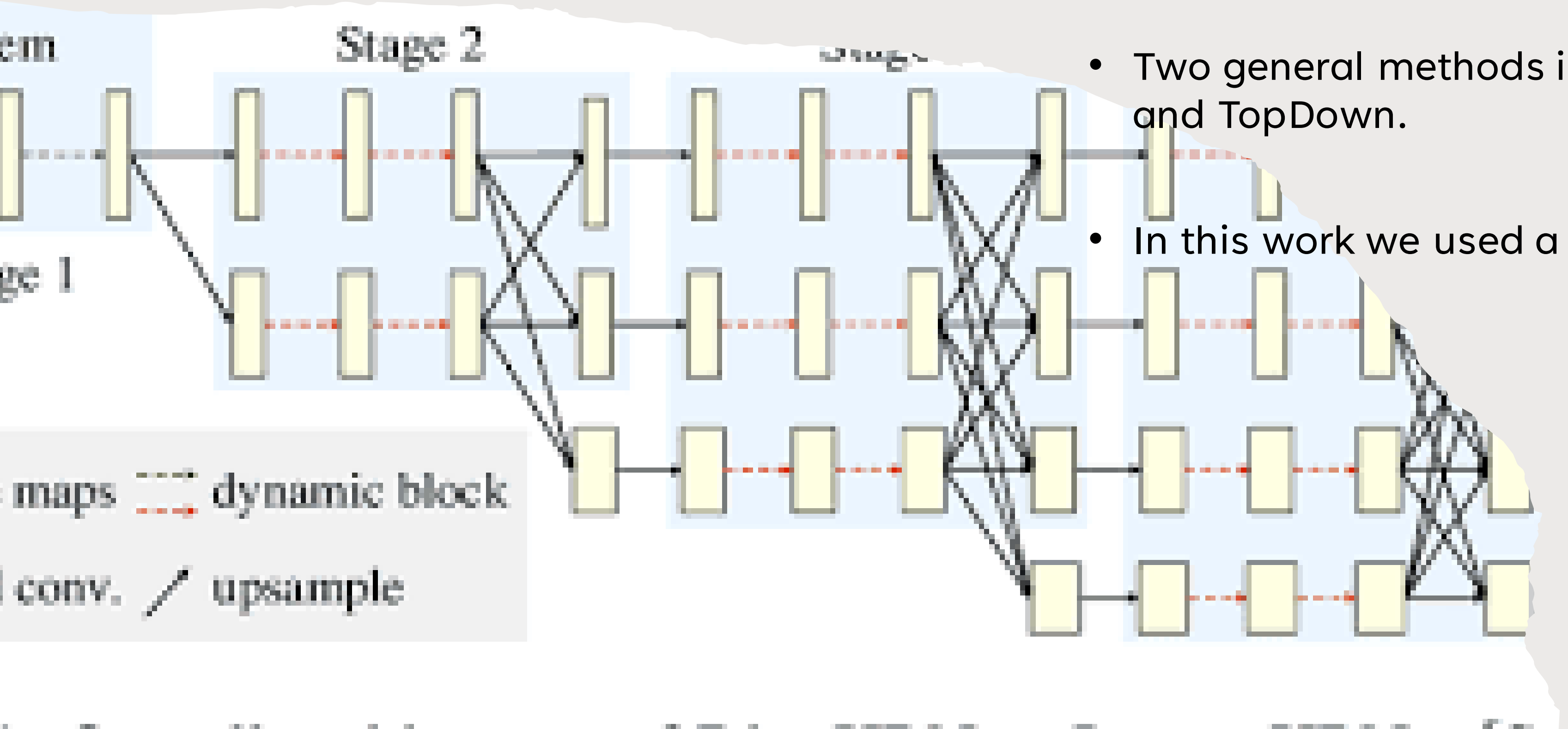
Lack of skeleton-based and multi-view HAR datasets in this field.

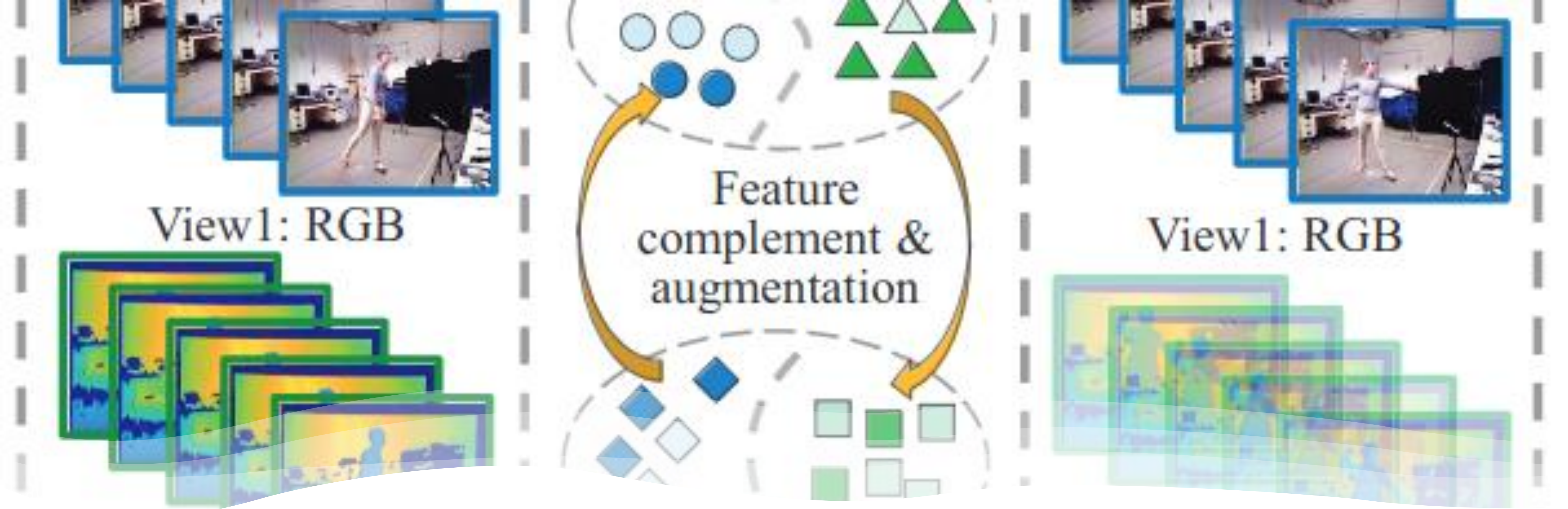
Developing a new dataset focusing on assistive robotics will open a new horizon in this field.

BACKGROUND

POSE EXTRACTION FOR ACTIVITY RECOGNITION

- Pose extraction method is applied at an early-stage task in the HAR pipeline and plays a vital role in skeleton-based HAR.
- Low or high accuracy in this section directly affects the rest of the procedure.
- A reliable HAR method is dependent on a high-accuracy pose extraction method.
- Two general methods in 2D pose estimators, BottomUp and TopDown.
- In this work we used a TopDown Mehtod (The HrNet)



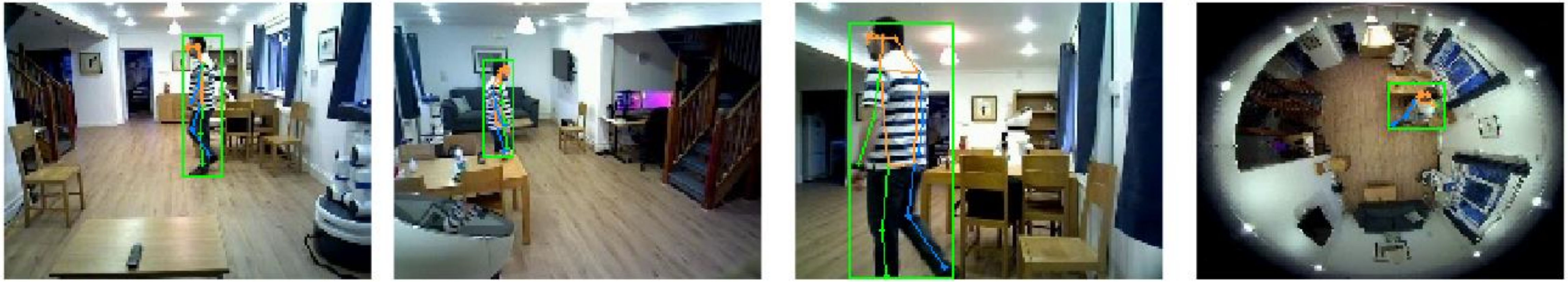


BACKGROUND

GENERATIVE AND NON-GENERATIVE DATASETS

- Generative approaches produce their input data from one or more actual views, whereas non-generative approaches acquire their data from genuine input devices.
- This work addresses the lack of non-generative skeleton-based HAR datasets including a robot view.

RHM-HAR-SK DATASET



(a) Walking Back-view. (b) Walking Front-view (c) Walking Robot-view. (d) Walking Omni-view.

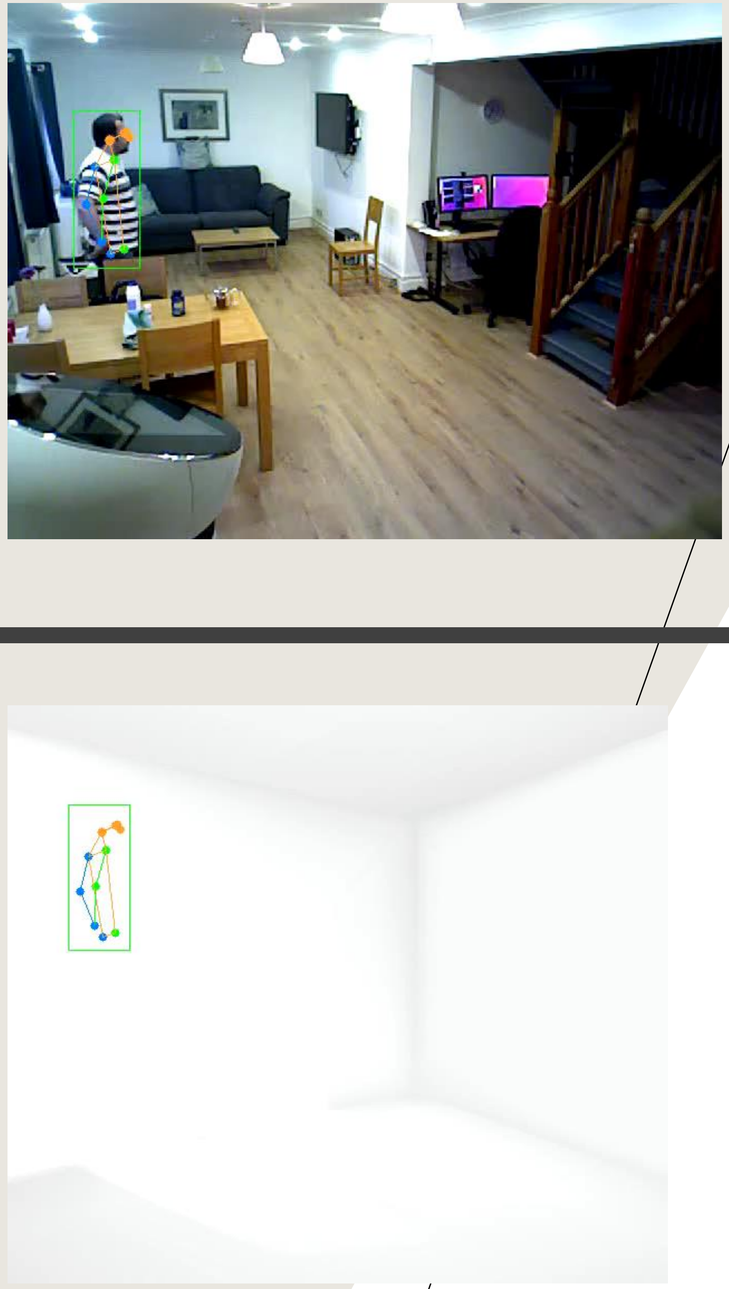
- Created on top of the extended version of RHM RGB data
- Multi-view human activity dataset
- Single person, trimmed video from four independent cameras
- Cameras used to cover a typical living room of a British home
- Captures fourteen daily indoor activities

Fourteen daily actions

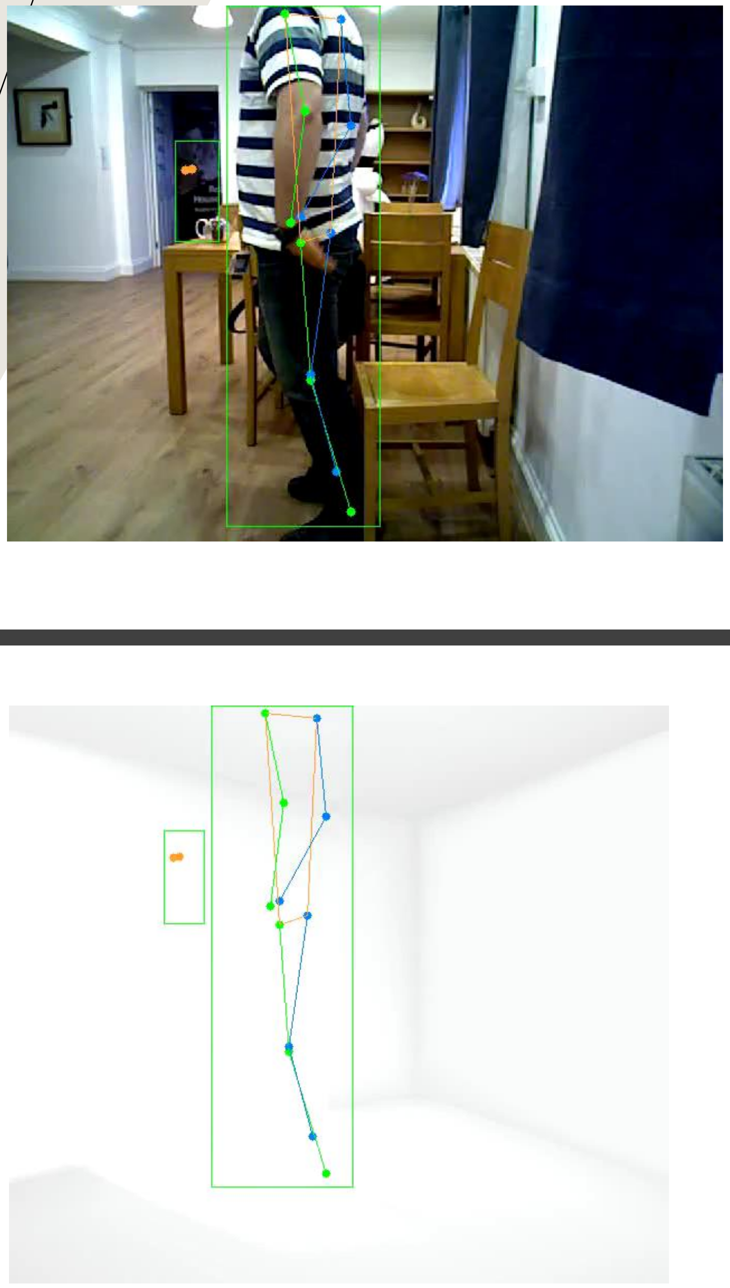
[walking, bending, sitting down, standing up, cleaning, reaching, drinking, opening can, closing can, carrying object, lifting object, putting down object, stairs climbing up, stairs climbing down]

- Two wall-mounted cameras
  (Front-view and Back-view)
- One mobile robot camera
  (Robot-view)
- One ceiling fish-eye camera
  (Omni-view)
- Videos from different views overlap and synchronized

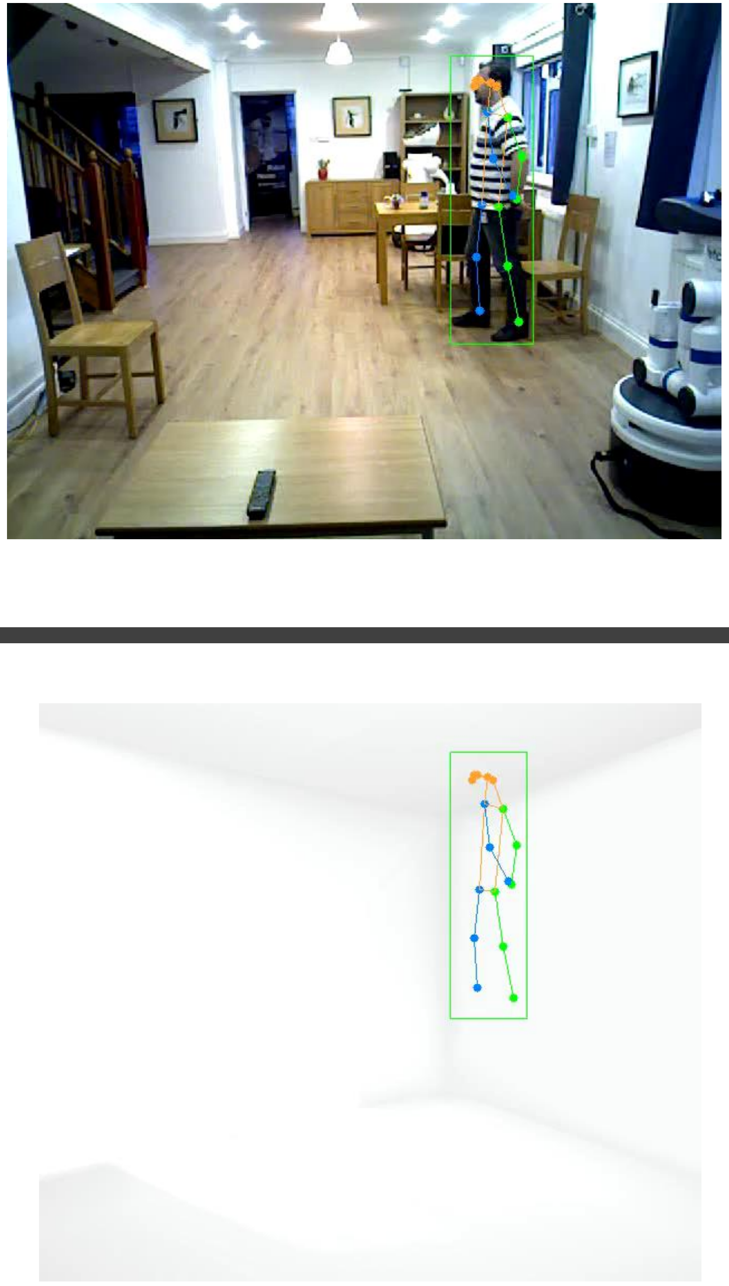
Front View



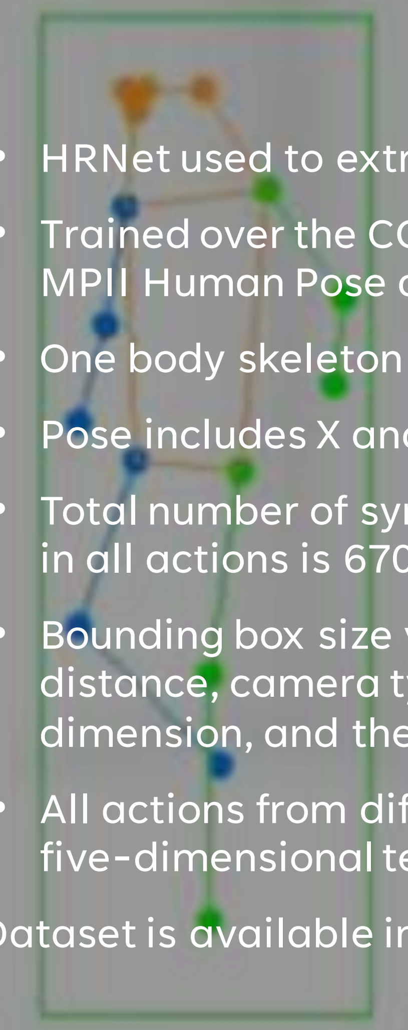
Robot View



Back View

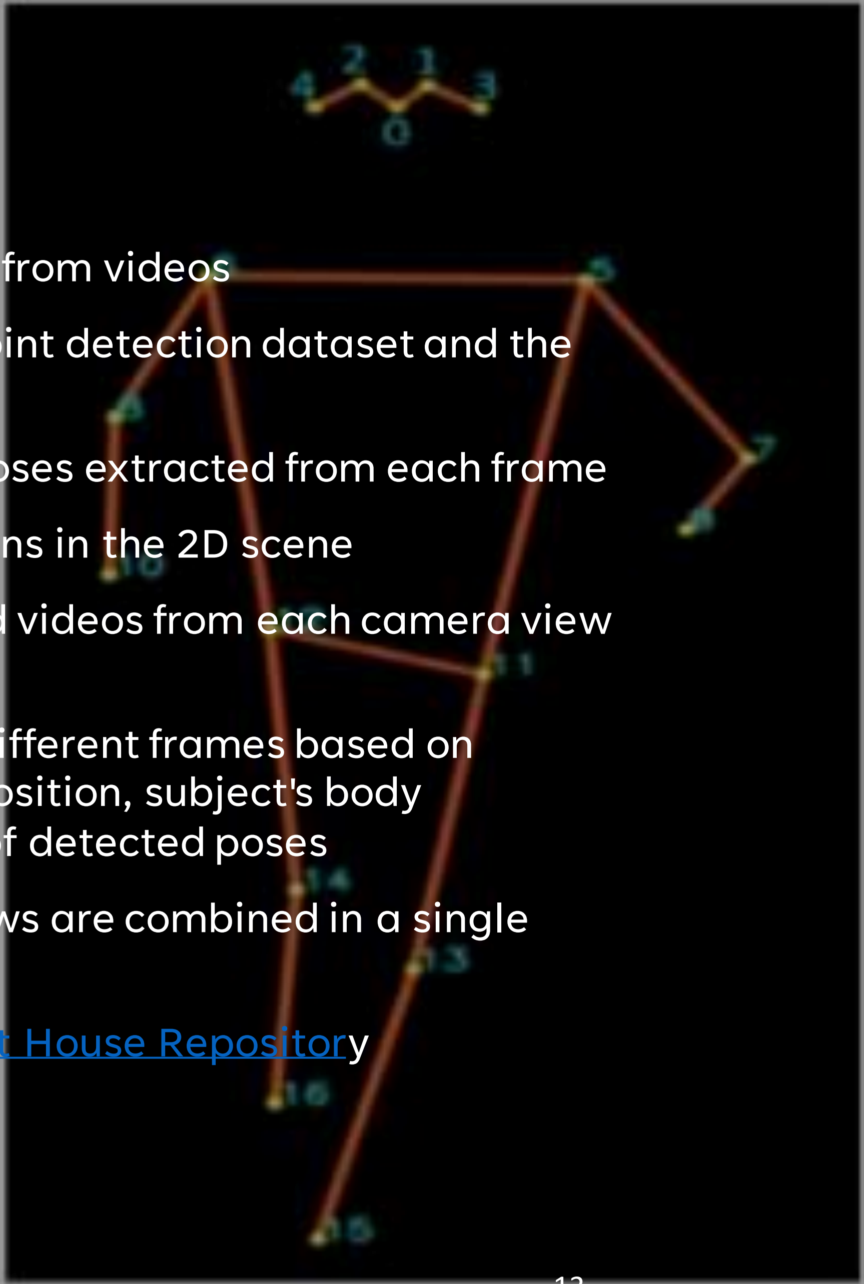


POSE EXTRACTION HRNET



- HRNet used to extract poses from videos
- Trained over the COCO keypoint detection dataset and the MPII Human Pose dataset
- One body skeleton with 17 poses extracted from each frame
- Pose includes X and Y positions in the 2D scene
- Total number of synchronized videos from each camera view in all actions is 6700
- Bounding box size varies in different frames based on distance, camera type and position, subject's body dimension, and the number of detected poses
- All actions from different views are combined in a single five-dimensional tensor

Dataset is available in the Robot House Repository

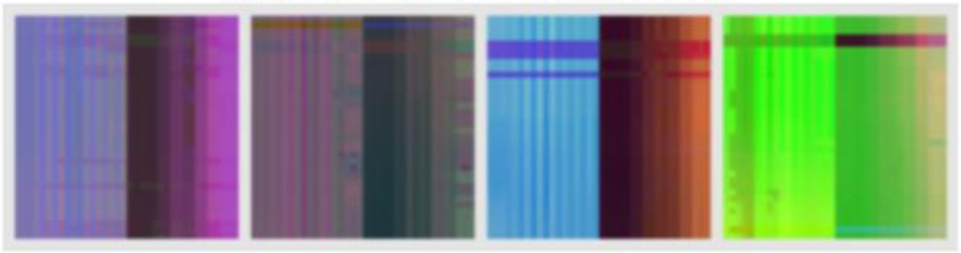


POSE EXTRACTION INPUT DATA SIZE AND SAMPLING

- Challenging part of the HAR task is video frame sampling
- ML models need to deal with dynamic input size
- Ordered random sampling method used
- A fixed number of frames like 34, 64, and 128 selected randomly from entire frames

0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19

Image examples of the RH-HAR dataset



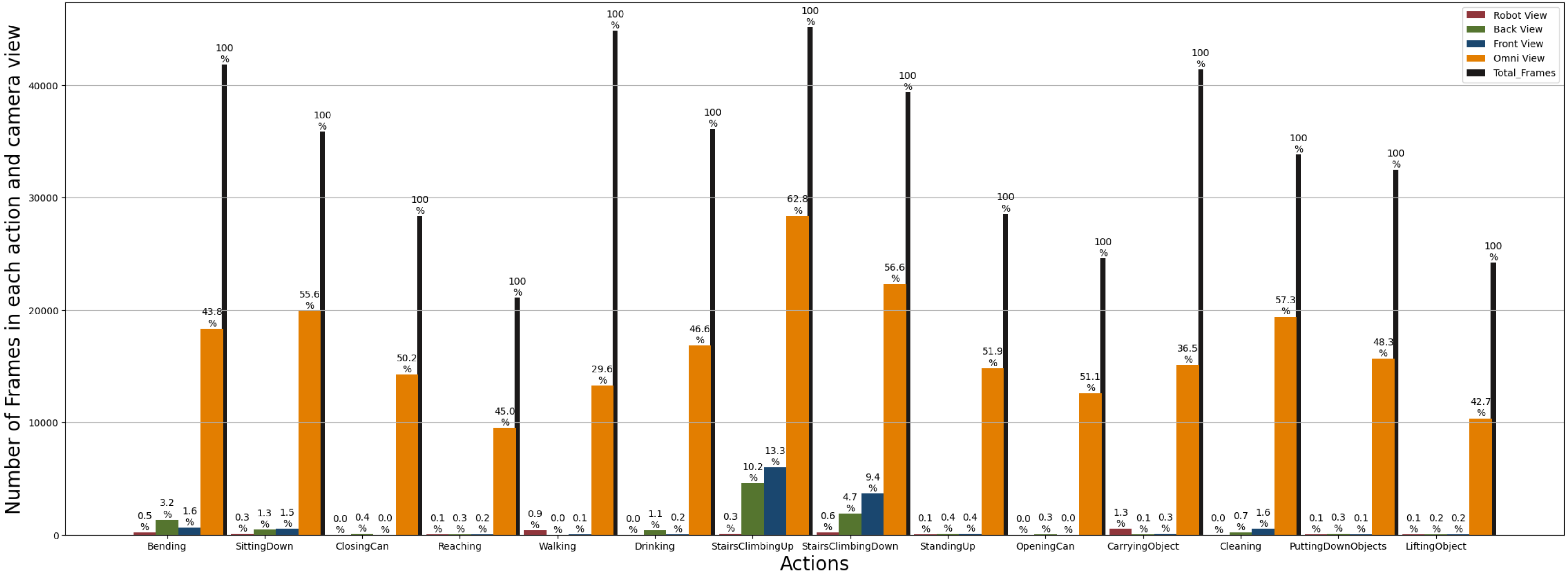
QUANTITATIVE AND QUALITATIVE ANALYSIS OF RHM-HAR-SK DATASET

Two general terms are considered to describe the quality of extracted skeleton from RGB images,

- The number of missed frames
- The number of missed poses.

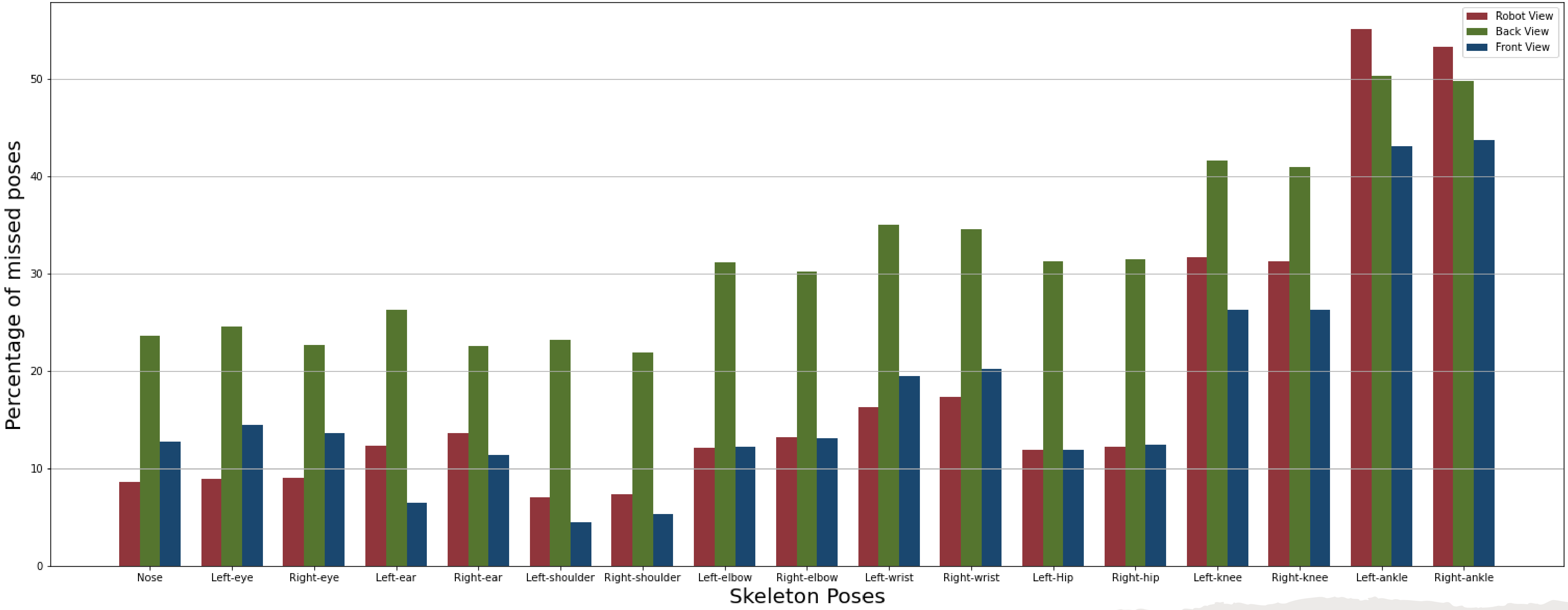
MISSED FRAME

The RGB frames on which the pose extraction methods could not find any human skeleton is considered a missed frame.



MISSED POSES

There are three parameters for each pose, X and Y values in 2D space and the confidence score. The confidence value refers to how much the extracted position is accurate. This value is between 0 to 1, and we considered the values less than 0.5 as missed poses.

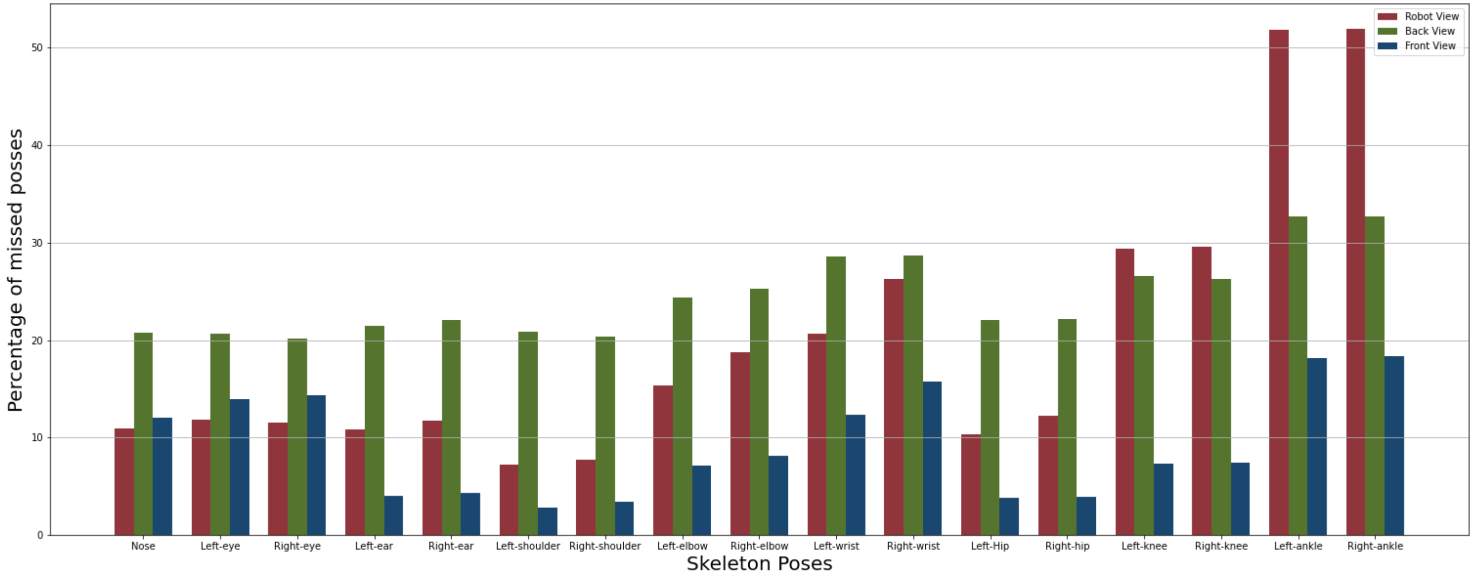


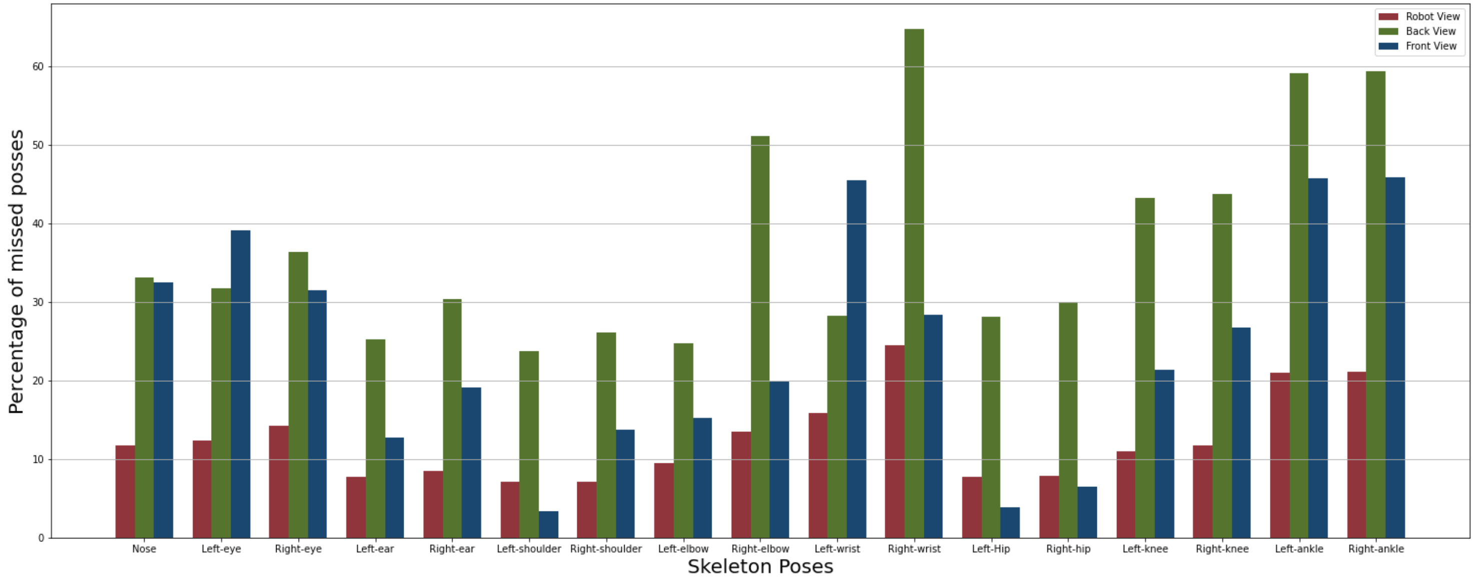
Walking

VS

stairs climbing

MISSED POSES





CONCLUSIONS

Importance of Data Acquisition Techniques for Human Activity Recognition in Ambient Assisted Living Scenarios

- Missed frames statistics reveals the reliability of different camera views frames'
- Correlation between action type and number of missed frames in fixed wall mount cameras vs. robot view
- Robot view follows human resulting in fewer errors in stairs climbing actions
- Higher attitude and broader view in wall-mounted cameras decrease missed poses
- Camera position, view, activity type, and joints are significant in quality of pose extraction
- Combining Robot-view camera and two other cameras can enhance skeleton-based human activity recognition, but incurs substantial expenses
- **Future work** should focus on developing light-weight models with multiple views to improve human activity recognition accuracy in ambient assisted living scenarios.

THANK YOU