UNIVERSITÀ DI PAVIA
Department of Electrical, Computer and Biomedical Engineering

Azure Kinect data Fusion for Enhanced Skeleton Tracking

Presented by:

PhD student **Leo Russo**

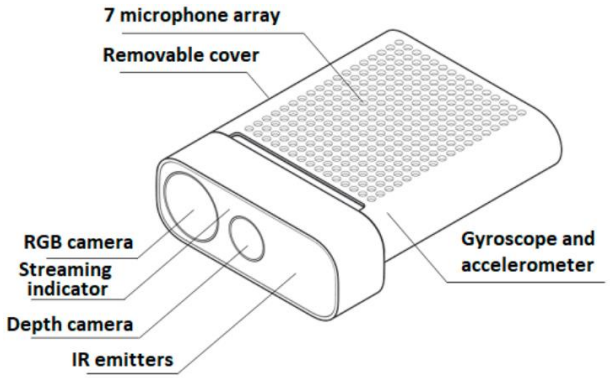
ZüMüTü 2025

Index

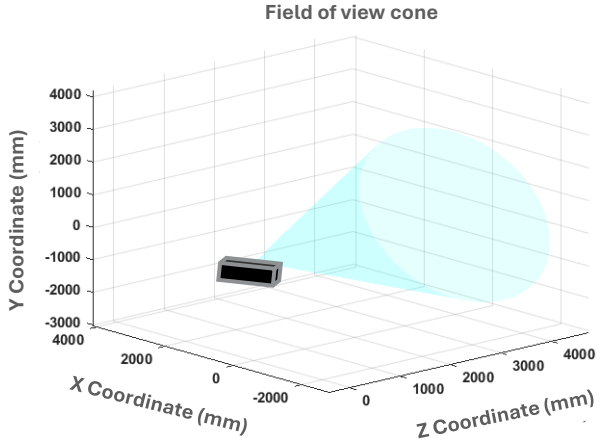
- Kinect overview
- Kinect field of application: HAR & AAL
- Proposed skeletal fusion algorithm
- Experimental validation of the algorithm
- Results and discussion

Bioengineering Laboratory – University of Pavia
leo.russo01@universitadipavia.it

Microsoft Kinect Azure

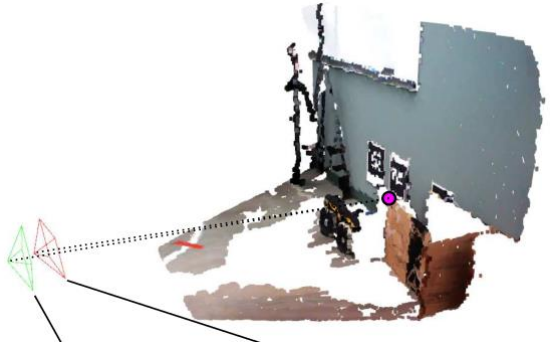


Mode	Resolution	FoI	Operating range
NFOV unbinned	640x576	75°x65°	0.5 – 3.86 m
NFOV 2x2 binned (SW)	320x288	75°x65°	0.5 – 5.46 m
WFOV 2x2 binned	512x512	120°x120°	0.25 – 2.88 m
WFOV unbinned	1024x1024	120°x120°	0.25 – 2.21 m



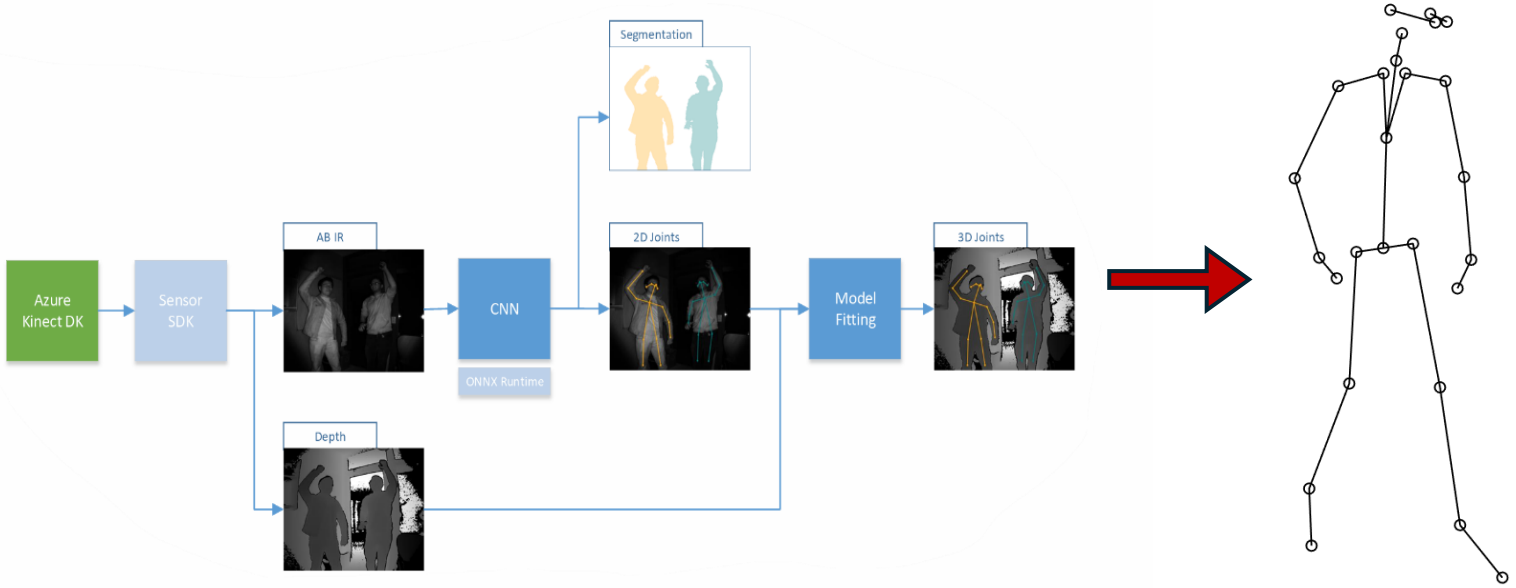
2D RGB image + Depth Map → 3D scene reconstruction

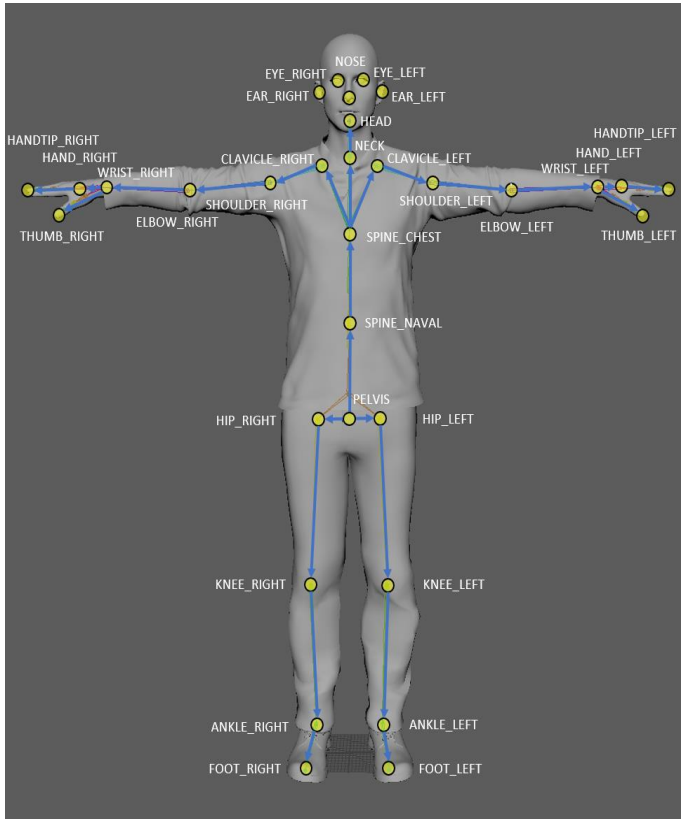
2D IR image



SDK Body Tracking

- Recognition of multiple subjects in the scene
- Skeleton subject reconstruction by detecting **32 anatomical joints**





Azure Kinect for HAR and AAL

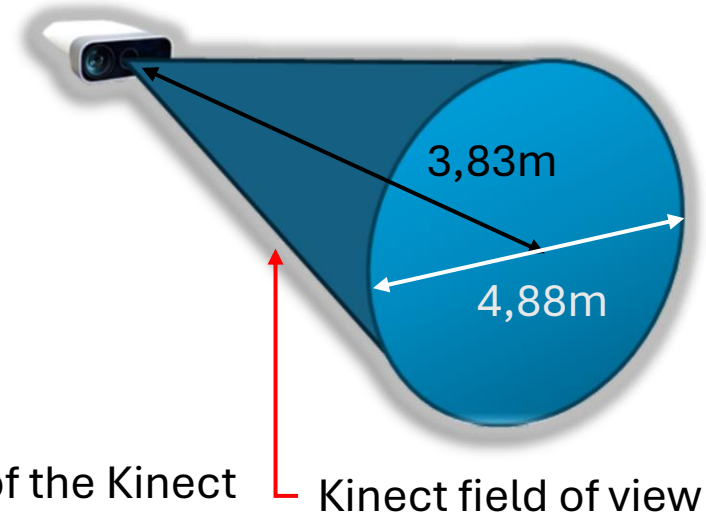
Human Activity Recognition uses technologies to detect and classify a person's action (e.g., walking, sitting, falling), by analyzing the movement of the joints and using deep learning models.

Ambient Assisted Living refers to the use of technologies to support the daily life of frail individuals, by monitoring the environment and behavior to improve safety and independence.

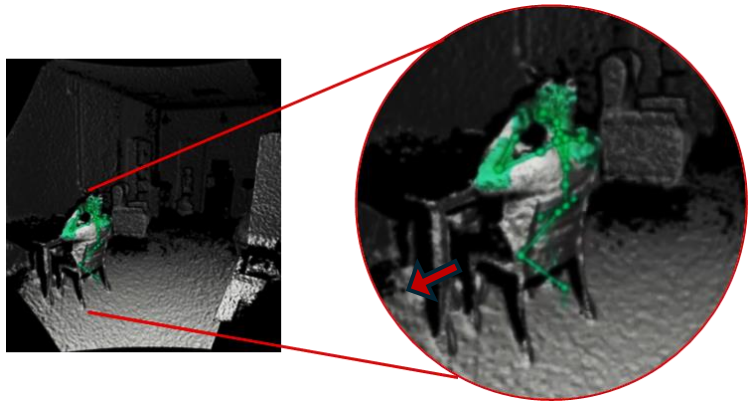
Use the Kinect Azure for **HAR** in **AAL**

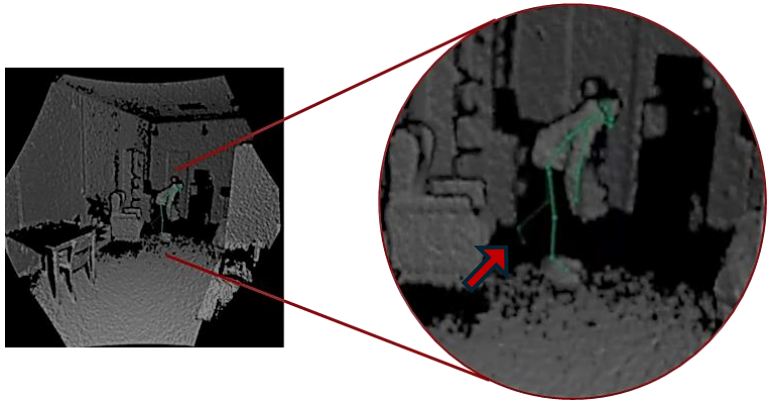
Limitations of the Kinect Azure in AAL

- Limited field of view for covering an entire room

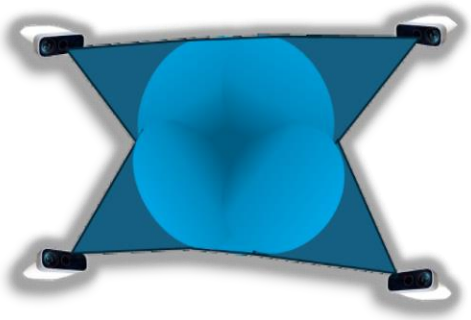


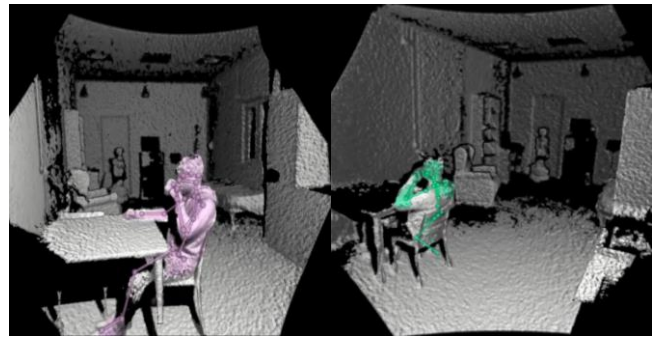
- Occlusions and self-occlusions of the subject in the field of view of the Kinect





Proposed solution: Network of Kinect

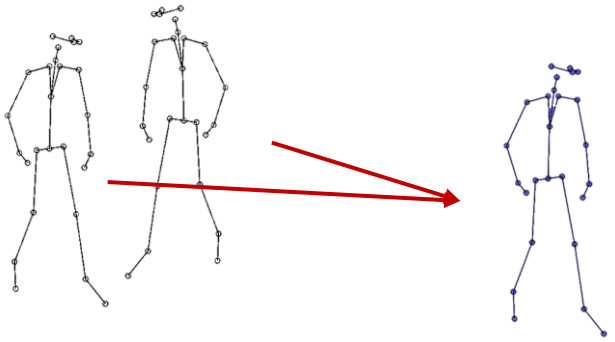


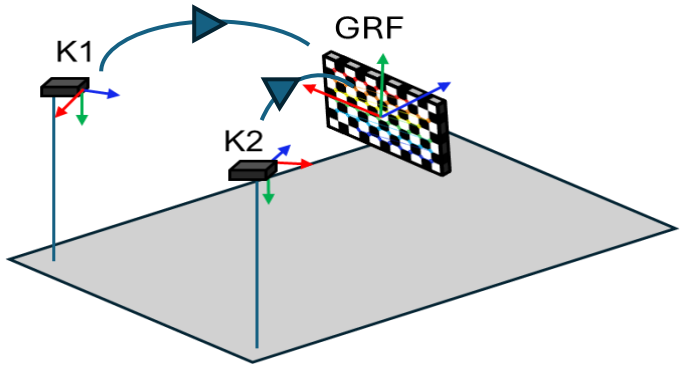


✓ Expand the field of view

✓ Address occlusion or self-occlusion issues

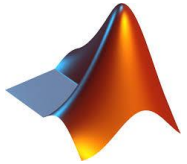
- Algorithm that combines the different outputs of the devices
- Calibration to obtain each Kinect output in a **G**lobal **R**eference **S**ystem





Algorithm

- MATLAB, MathWorks (MATLAB R2024b)

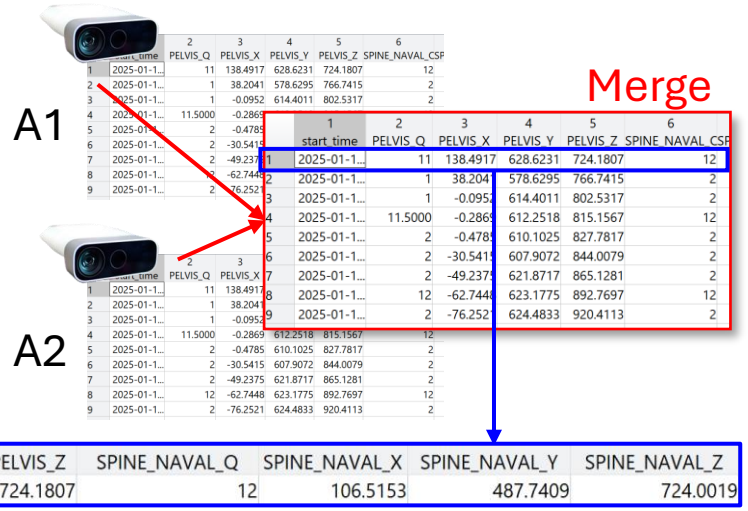


MATLAB

- Allows to obtain a single capture file (Merge) from two different Kinect acquisitions (A1 and A2)

- Each frame contains the timestamp in which the image is captured, and for each joint the 3D position and the **quality value** of detected joint

- In Merge file the data stored are obtained by selecting or combining the 3D joints coordinates of frames contained in A1 and A2



Quality	Description
0	No reliability information
1	Estimated
2	Tracked

Algorithm

Time control

✓ ΔTimestamp (A1,A2) >16 ms

✗ ΔTimestamp (A1,A2) ≤ 16 ms

The frame shotted first into Merge

Combine or Select

➢ Subject's orientation with respect to each kinect (α_{Ki})

α_{K1} and α_{K2} ≤ 45° ✓ Combine

α_{K1} or α_{K2} ≤ 45° ✓ Select the frame that respects condition

α_{K1} and α_{K2} > 45° ✗ Score S calculation

Algorithm

Orientation-based score (PA):

$$P_a = \frac{(180° - |\alpha|)}{180°}$$

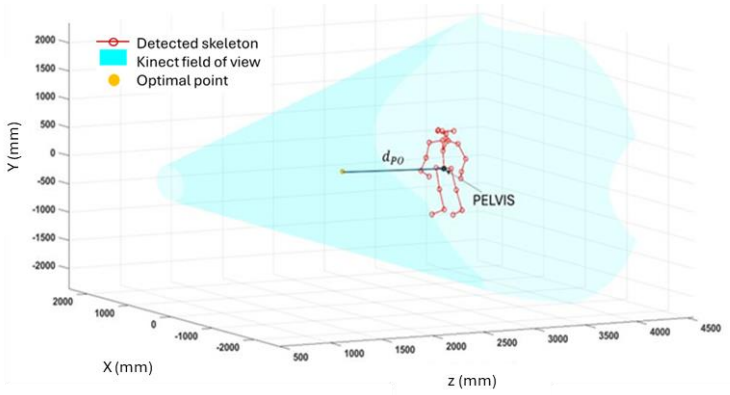
Joint quality-based score (P_Q):

$$P_q = \frac{N_{jointQmax}}{Tot_{joint}}$$

Position based score (P_{PO}):

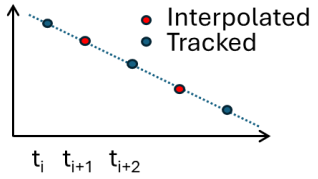
$$P_{PO} = \begin{cases} 1 \text{ if dPO } \leq 150 \text{ mm} \\ \frac{1}{\text{dPO}} \text{ if dPO } > 150 \text{ mm} \end{cases}$$

➢ **Score S = P_A•P_Q•P_{PO}**



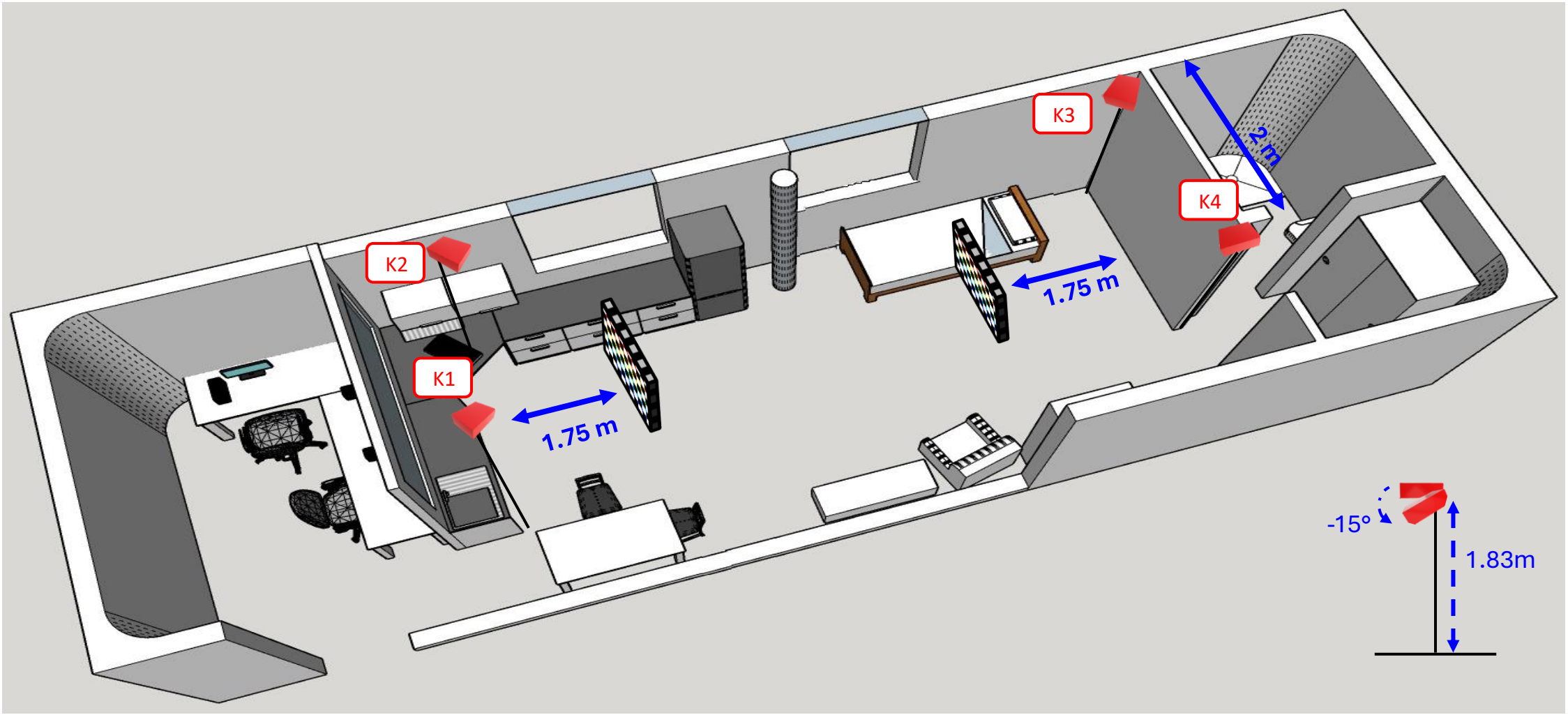
Optimal point
defined as the center of the Kinect's field of view cone [0;0;1,75 m]

✓ Select the frame that **maximize** S score

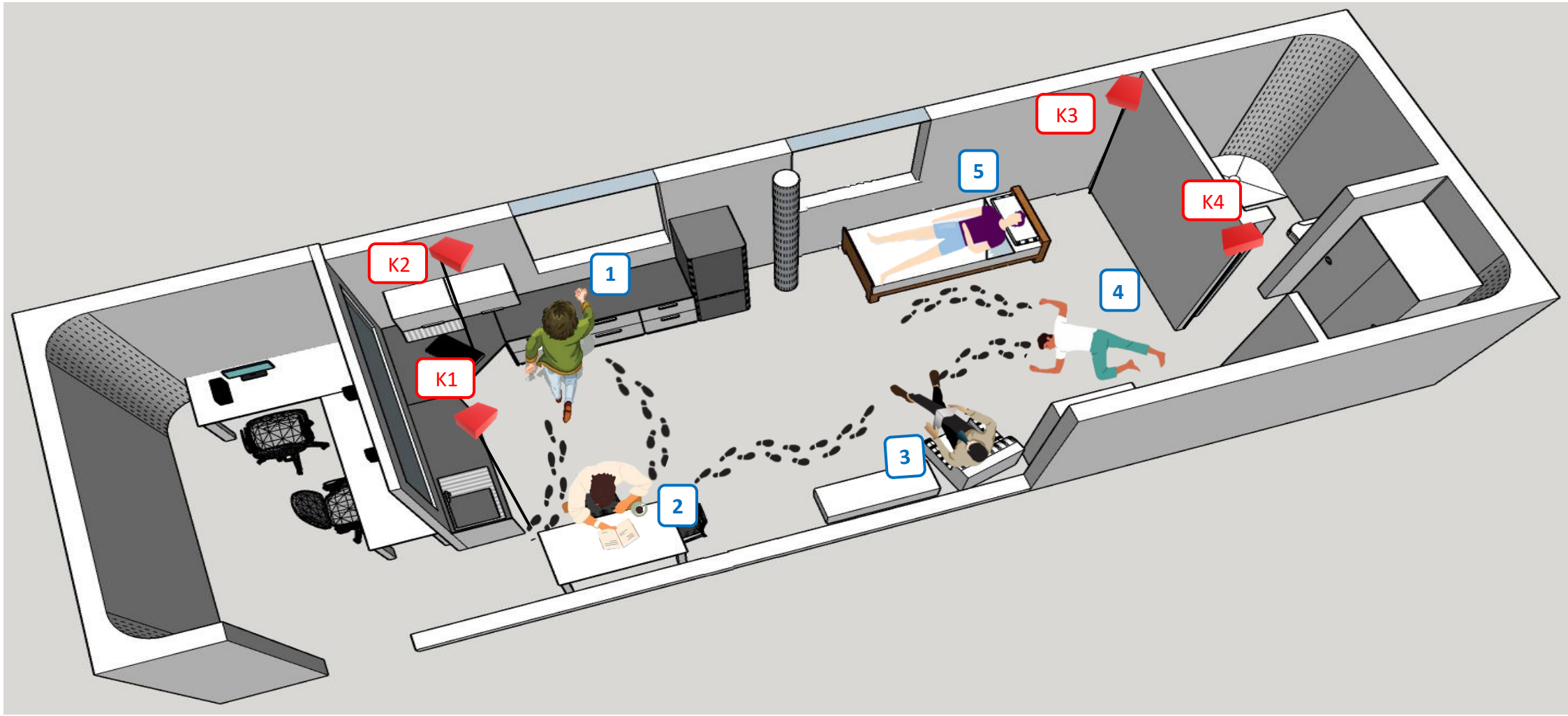


Data linear interpolation of the Merged file (1 sample window)

Experimental Setup



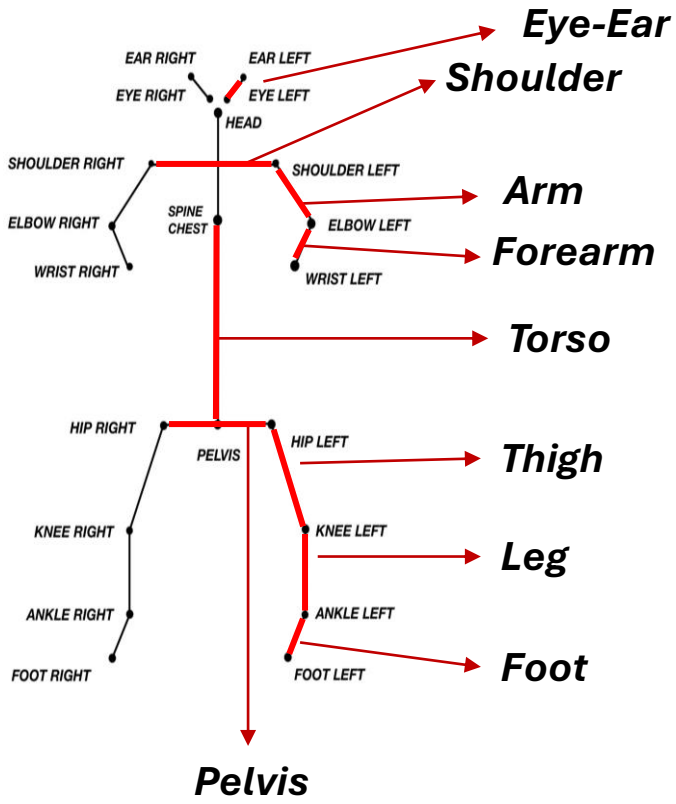
Experimental Setup



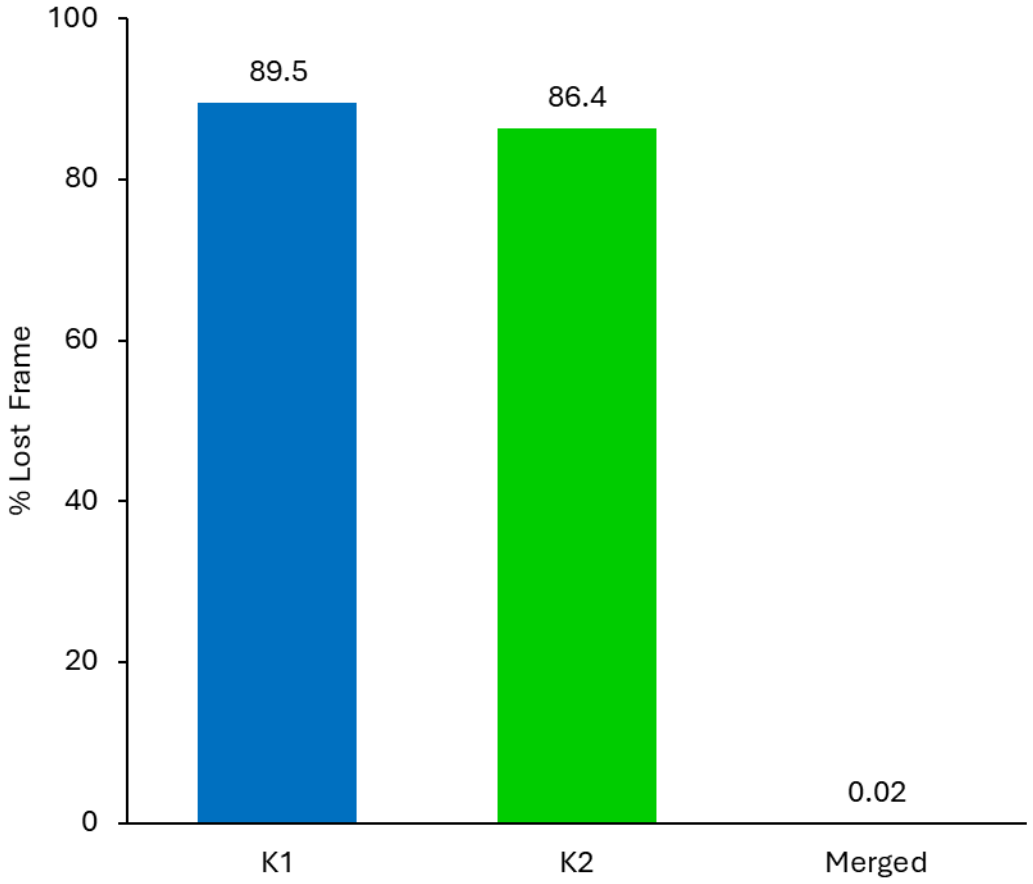
Evaluation Metrics

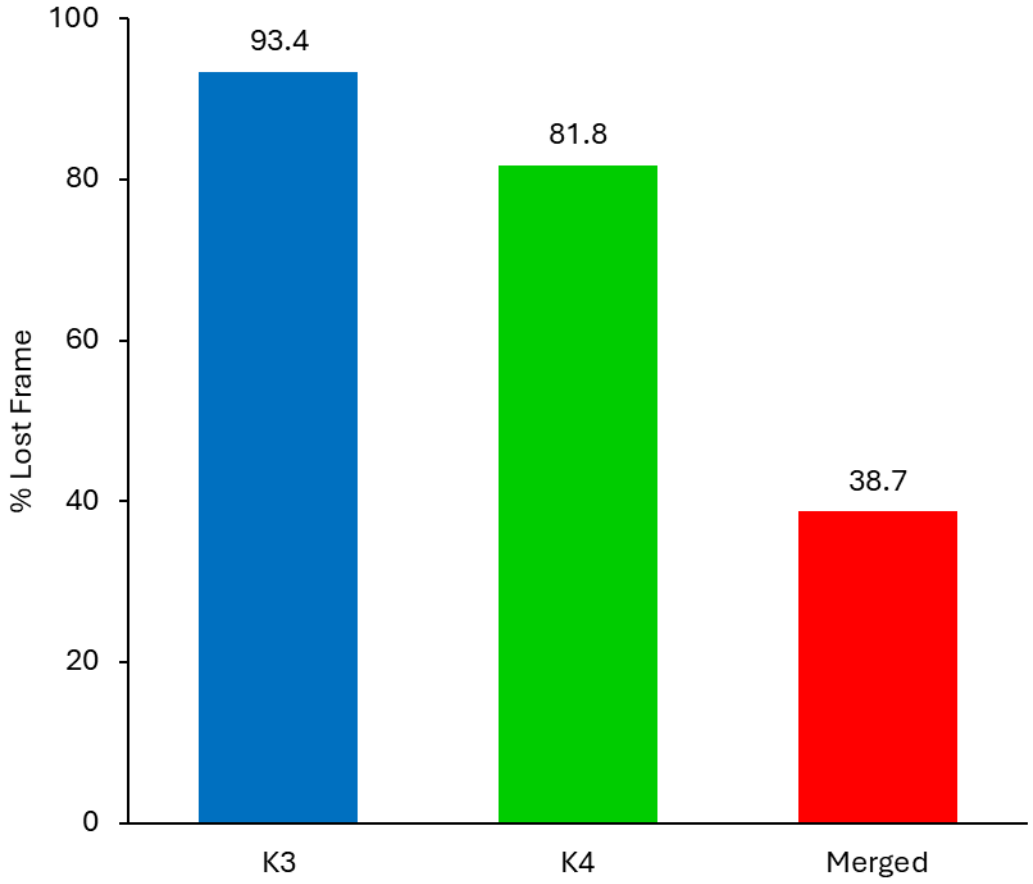
- **% Lost frames**= $\frac{\text{Number of lost frames}}{\text{Total duration of acquisition}}$

- **S**tandard **D**eviation **(SD)** of body segment lengths

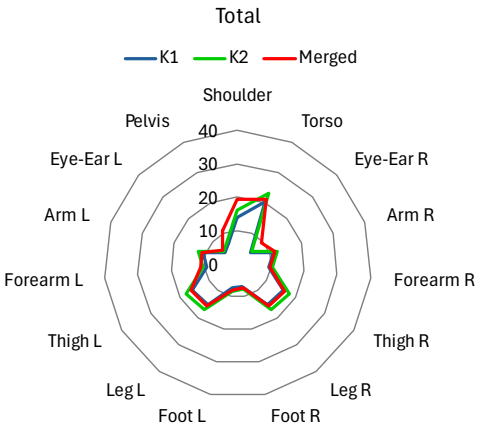


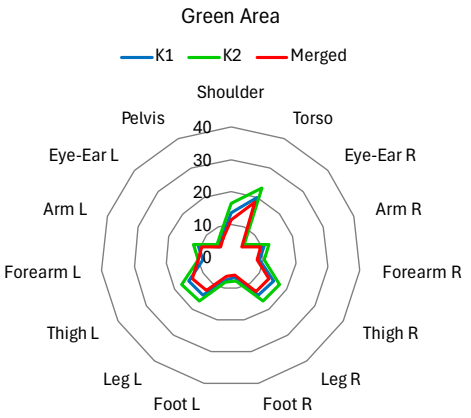
Results – Lost Frames

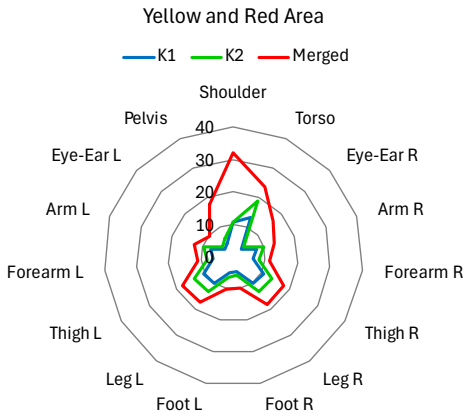


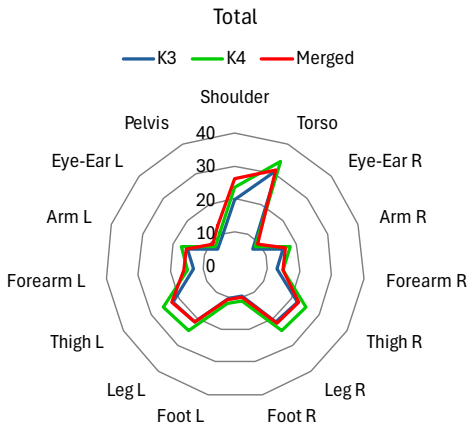


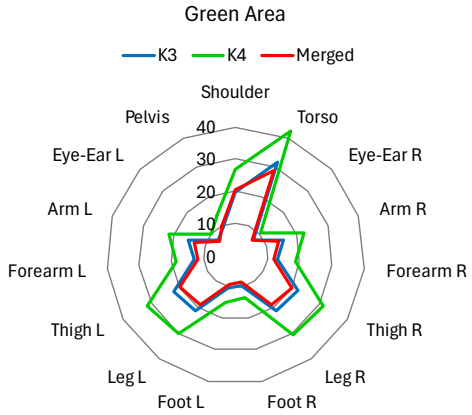
Results – SD segment lengths

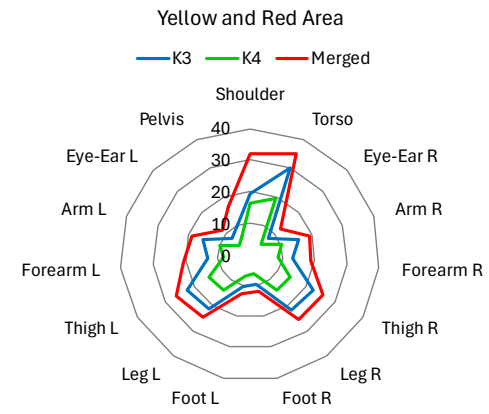




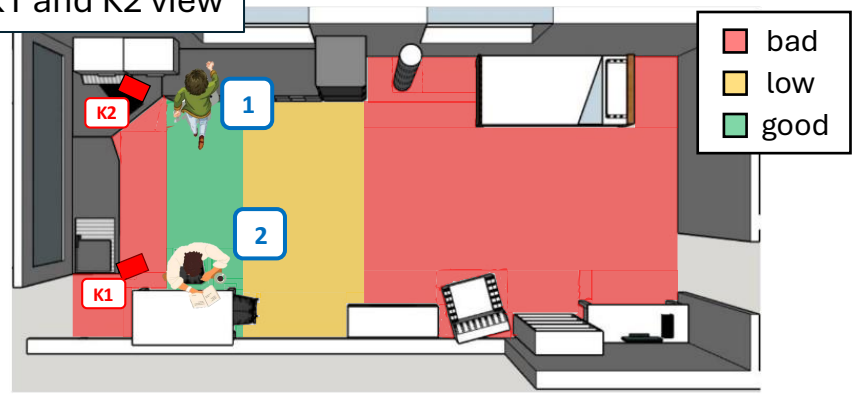




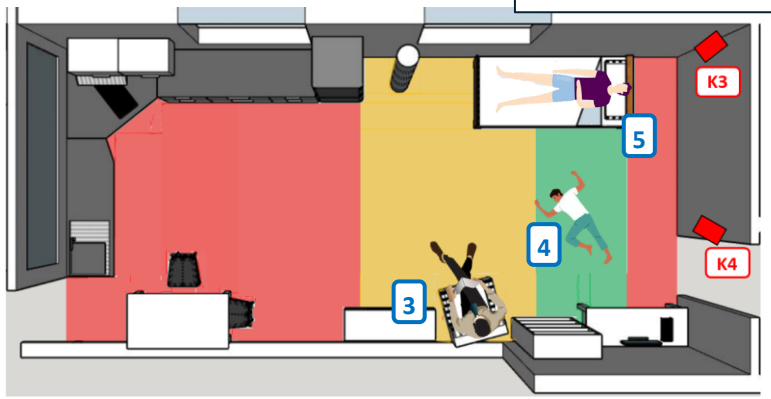




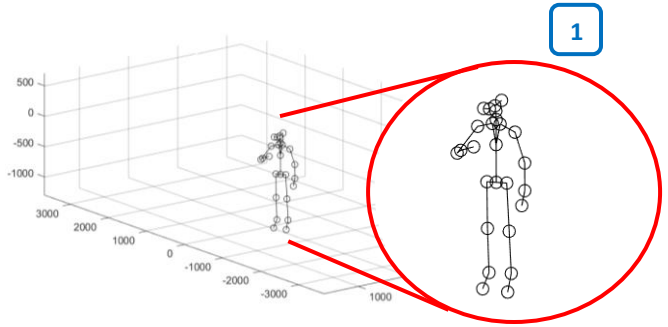
K1 and K2 view

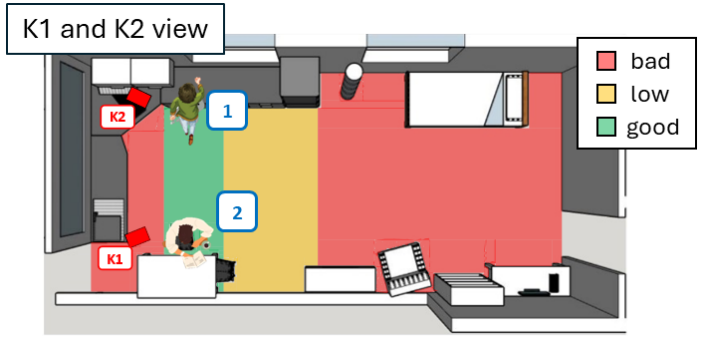


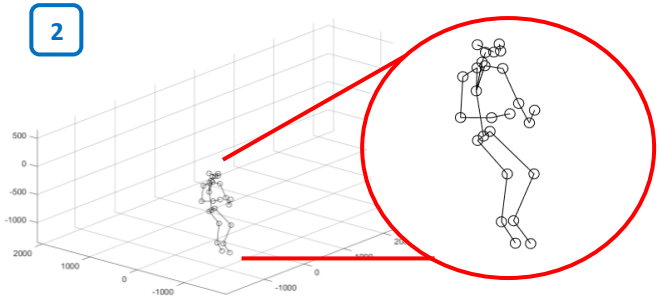
K3 and K4 view

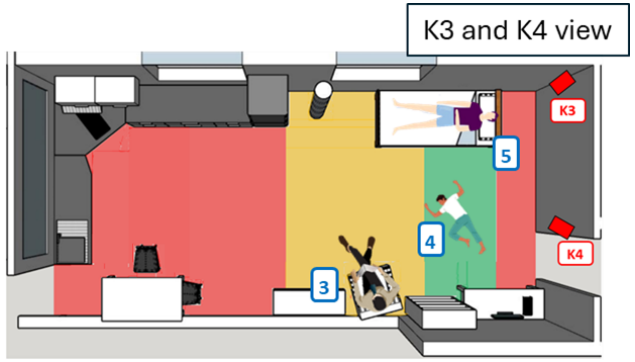


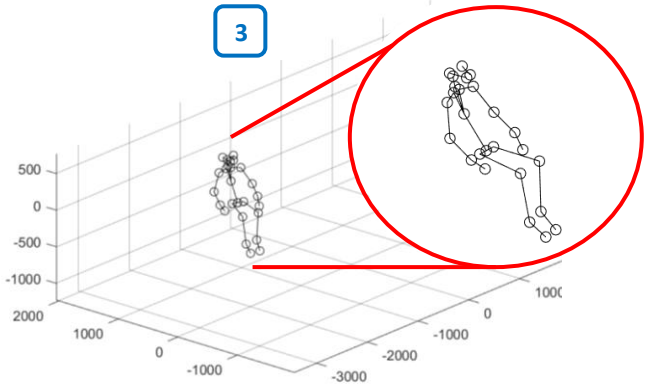
Discussion

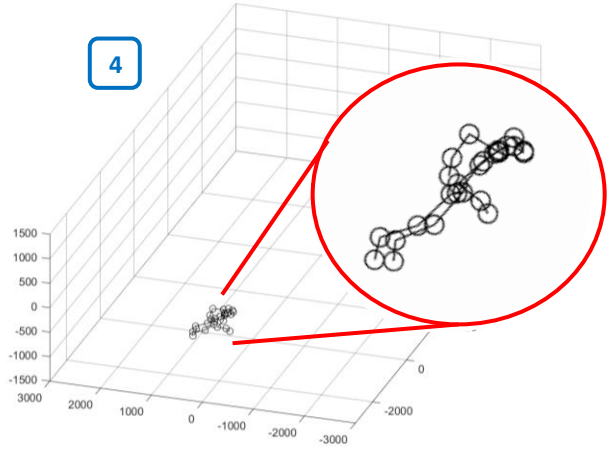

1

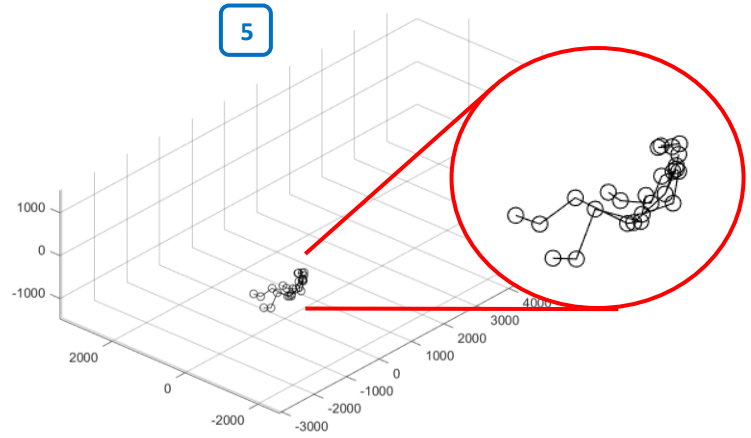

K1 and K2 view
bad
low
good
K2
1
2
K1


2


K3 and K4 view
K3
5
4
K4
3


3


4


5

Future Directions

- Combine the Kinect output only if the subject's distance from the Kinect is under 1.75 m
- Add new control in case the subject's pose is lying in bed or fallen to the ground
- Use a different strategy to combine the skeleton in the case of lying in bed or fallen to the ground
- Improve the algorithm for combining data from multiple Kinects

Thank you for your attention !