universidade de aveiro

Inside A Collaborative Cell: Calibration, Perception And Safety Requirements

October 30, 2025

Daniela Rato
PhD Mechanical Engineering
Supervised by Prof. Miguel Oliveira, Prof. Vítor Santos and Prof. Angel Sappa

Motivation

- Collaborative robots are moving into dynamic, human-shared environments.
- Mechanical safety is not enough → perception-driven cobots are needed.
- Accurate perception requires multiple sensors from different modalities → robust extrinsic calibration is needed.
- Robust perception requires:
  - **Accurate extrinsic calibration** across heterogeneous sensors;
  - **Reliable 3D human pose estimation** under occlusion and real-world variability.

Research Objectives

- To investigate how the existing calibration framework can be applied and **extended** for use in **robotic collaborative cell** scenarios.
  - Adapt and enhance Atomics Transformation Optimization Method for robotic cells, including **RGB-D** and hand-eye RGB-D calibration.
  - Ensure **modularity** and **integration** into robot operating environments.
- To design and implement a **3D human pose estimation** framework tailored to **collaborative robotics**.
  - **Multi-camera RGB** approach tailored for collaborative robots.
  - Optimised for **robustness under occlusion** and real-world deployment.

Research Questions

- How can calibration algorithms be designed to support heterogeneous, **multi-modal sensor setups**, including both **fixed and mobile configurations**, while reducing reliance on manual procedures and maintaining spatial accuracy?
- What optimisation strategy enables consistent and **accurate extrinsic calibration** in **dynamic environments** where sensor positions or orientations may vary over time?
- How can **3D human pose estimation** systems based solely on RGB imagery be configured to offer sufficient **accuracy, robustness to occlusion**, **and computational efficiency** for integration into collaborative robotic environments?

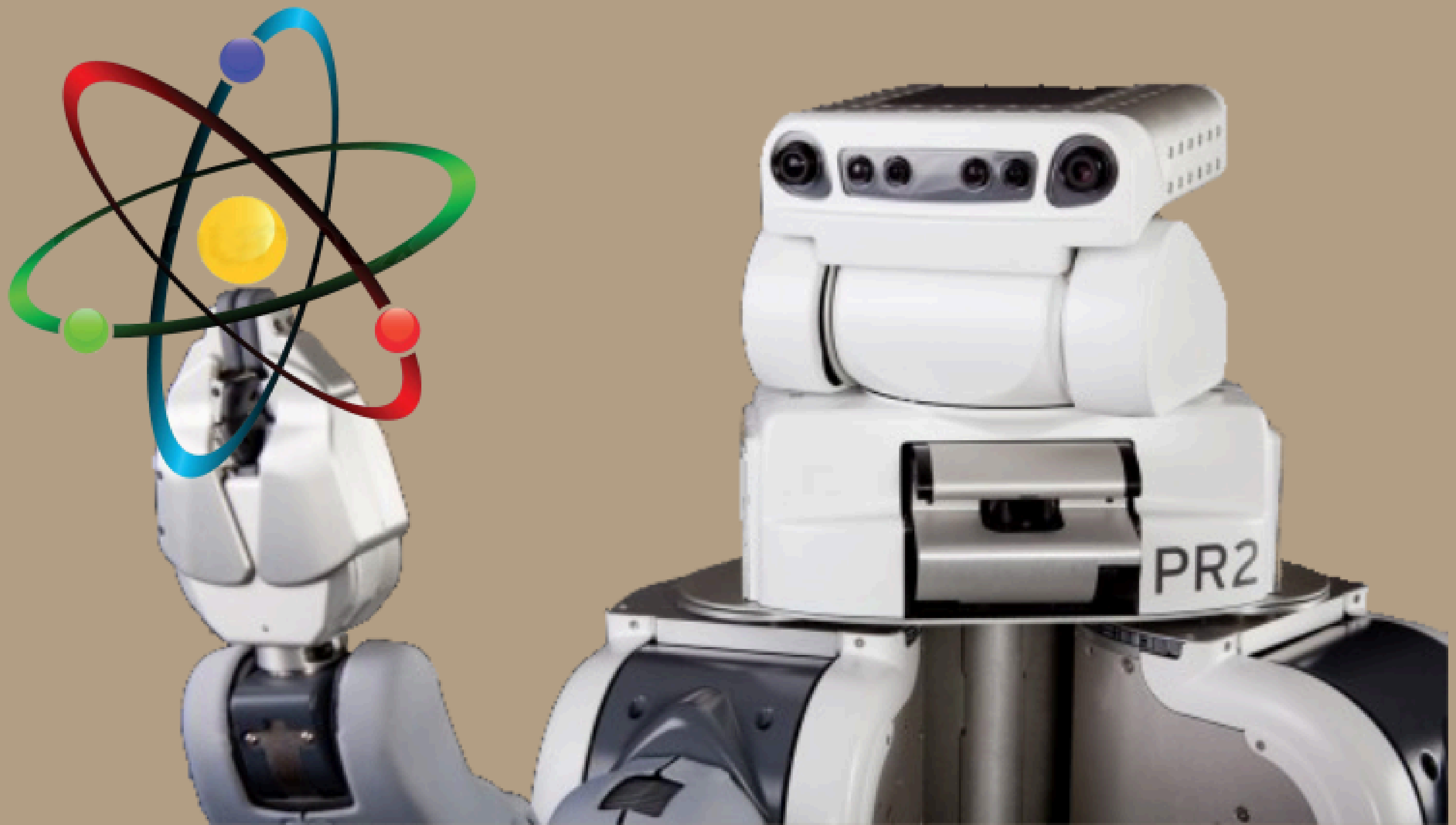
Research Questions

- How can calibration algorithms be designed to support heterogeneous, **multi-modal sensor setups**, including both **fixed and mobile configurations**, while reducing reliance on manual procedures and maintaining spatial accuracy?
- What optimisation strategy enables consistent and **accurate extrinsic calibration** in **dynamic environments** where sensor positions or orientations may vary over time?
- How can 3D human pose estimation systems based solely on RGB imagery be configured to offer sufficient **accuracy, robustness to occlusion**, **and computational efficiency** for integration into collaborative robotic environments?

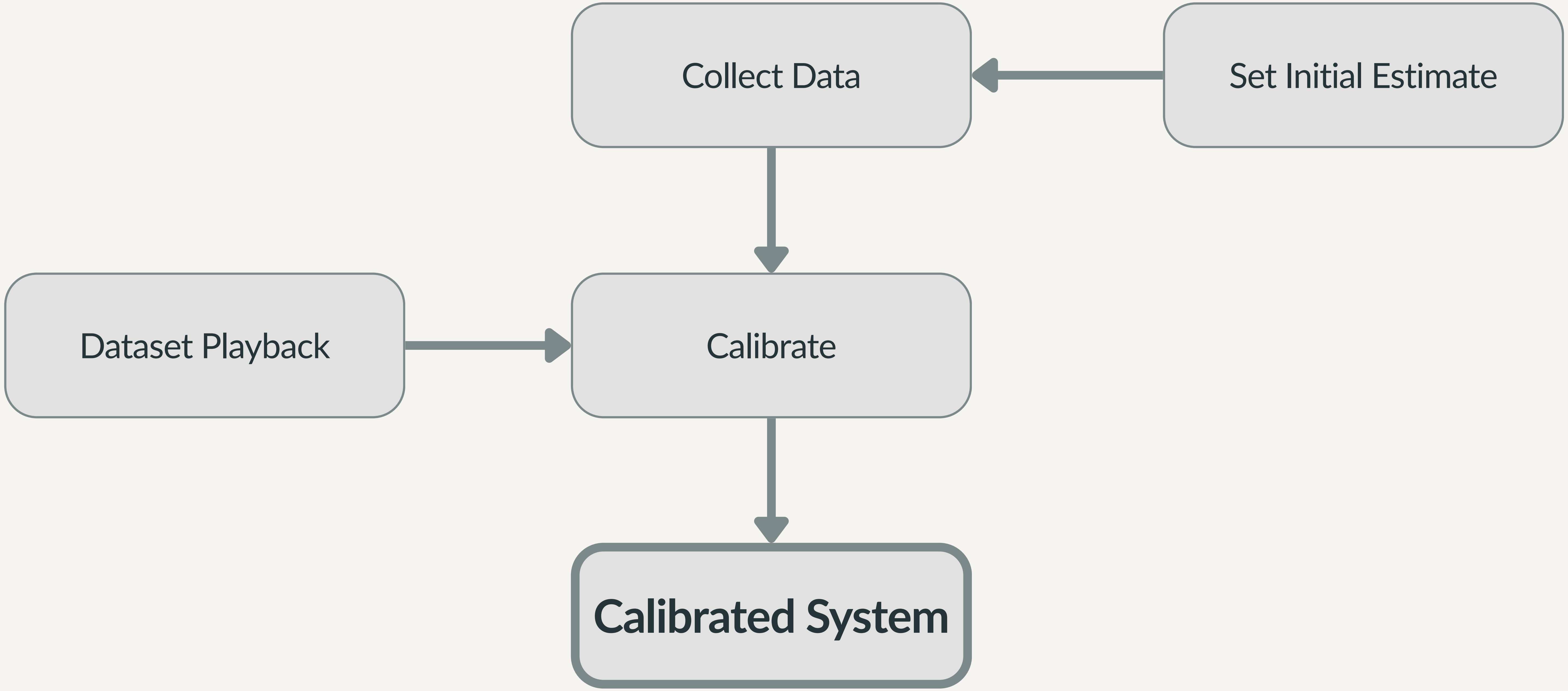
Hypothesis

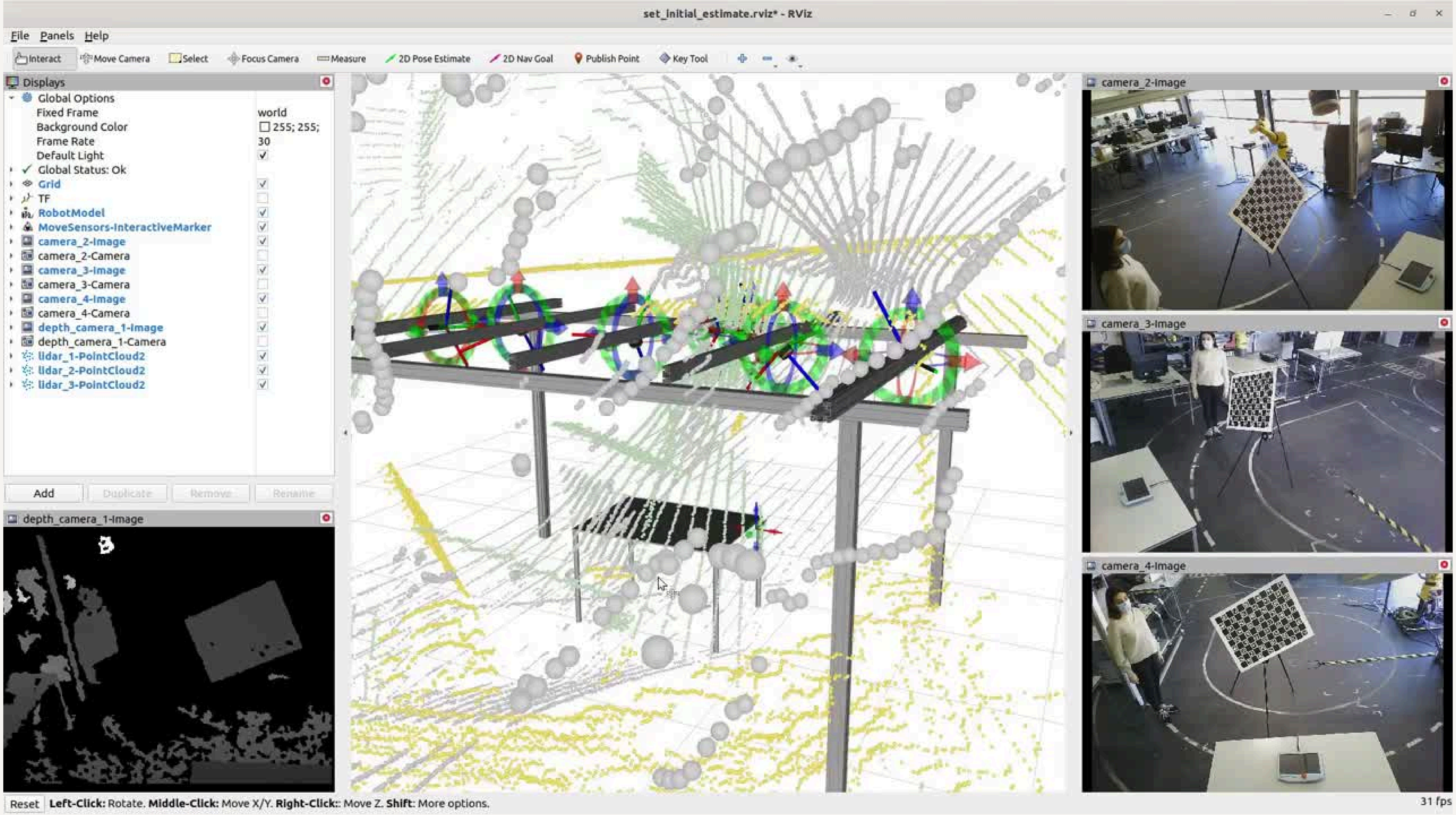
robust calibration + accurate pose estimation = a safer human-robot collaboration

ATOM:
Atomic Transformations Optimization Method



ATOM Framework




set_initial_estimate.rviz* - RViz
File
Panels
Help
Interact
Move Camera
Select
Focus Camera
Measure
2D Pose Estimate
2D Nav Goal
Publish Point
Key Tool
Displays
Global Options
Fixed Frame
world
Background Color
255; 255;
Frame Rate
30
Default Light
Global Status: Ok
Grid
TF
RobotModel
MoveSensors-InteractiveMarker
camera_2-Image
camera_2-Camera
camera_3-Image
camera_3-Camera
camera_4-Image
camera_4-Camera
depth_camera_1-Image
depth_camera_1-Camera
lidar_1-PointCloud2
lidar_2-PointCloud2
lidar_3-PointCloud2
Add
Duplicate
Remove
Rename
depth_camera_1-Image
camera_2-Image
camera_3-Image
camera_4-Image
Reset
Left-Click: Rotate. Middle-Click: Move X/Y. Right-Click:: Move Z. Shift: More options.
31 fps

speed: 4x

Calibration Formulation

Sensor to Pattern Approach

$$error = \underset{\{^{s_i}T^w\},\{^{p}T_c^w\}}{\arg\ \min} \sum_{S} \sum_{C} e\left(^{s_i}T^w, {}^{p}T_c^w, d_{s_i}, \lambda_{s_i}\right)$$

estimated transformation between sensor s_i and the world coordinate frame w

estimated transformation between the pattern p and w

set of sensors

set of collections = dataset

detection of the pattern by the sensor s_i

intrinsic parameters of sensor s_i

Depth Modality Extension

Orthogonal Error

$$e_{o_{[c,s]}} = \left[\left({}^{s}T_{c}^{p} \right)^{-1} . X_{[c,s]} \right]_z$$

estimated transformation between the sensor *s* and the pattern *p* for a collection *c*

detected pattern points for a collection *c* and sensor *s*

Longitudinal Error

$$e_{l_{[c,s,b]}} = \min_{q \in Q} \left(\left\| \left[q - \left({}^{s}T_{c}^{p} \right)^{-1} . X_{[c,s,b]} \right]_{xy} \right\|_F^2 \right)$$

Depth Modality Extension

Orthogonal Error

$$e_{o_{[c,s]}} = \left[\left({}^{s}T_{c}^{p} \right)^{-1} . X_{[c,s]} \right]_{z}$$

Longitudinal Error

$$e_{l_{[c,s,b]}} = \min_{q \in Q} \left(\left\| \left[q - \left({}^{s}T_{c}^{p} \right)^{-1} . X_{[c,s,b]} \right]_{xy} \right\|_{F}^{2} \right)$$

sampled pattern border point

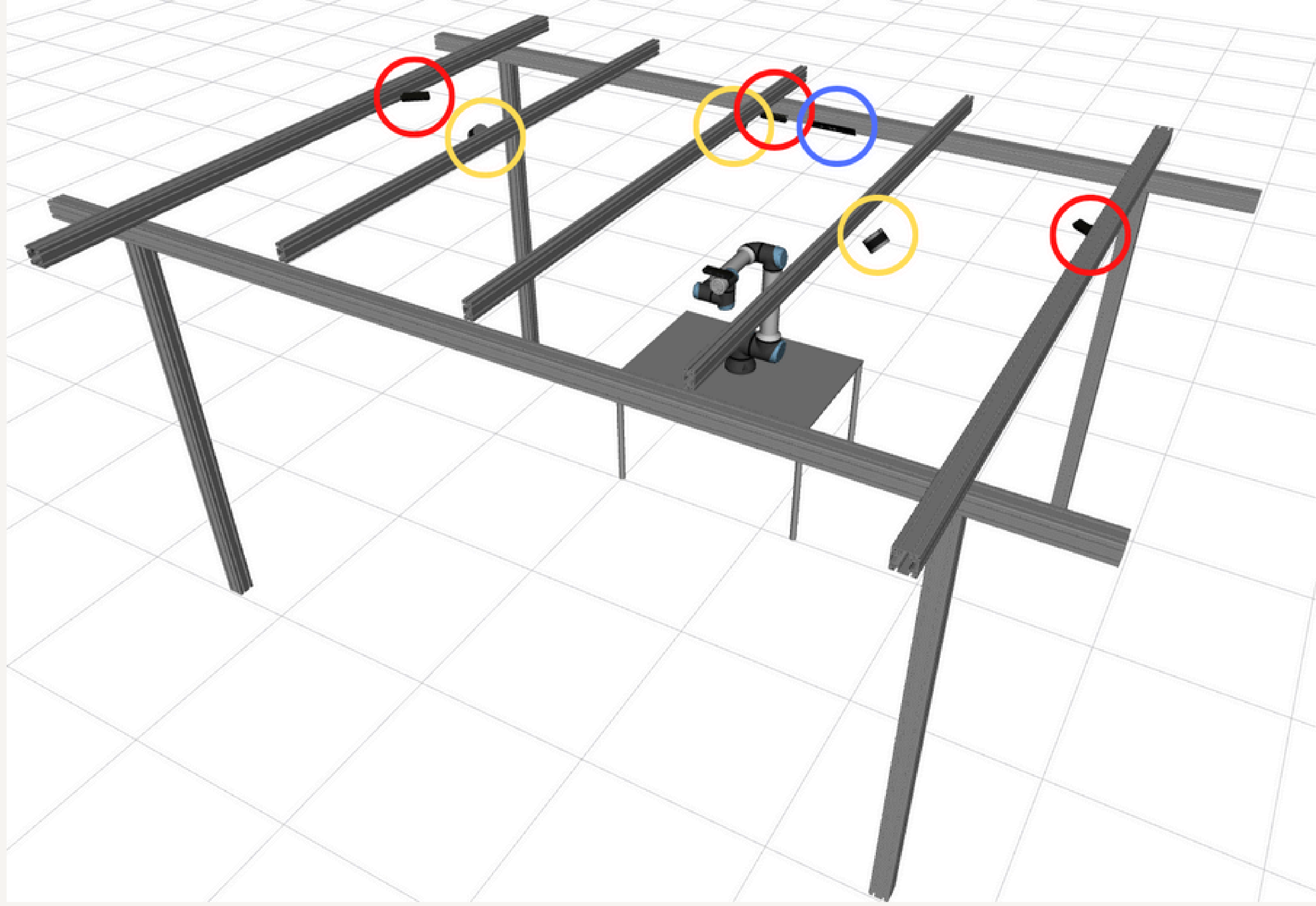
detected boundary points for a collection *c* and sensor *s*

Setup - Collaborative Cell

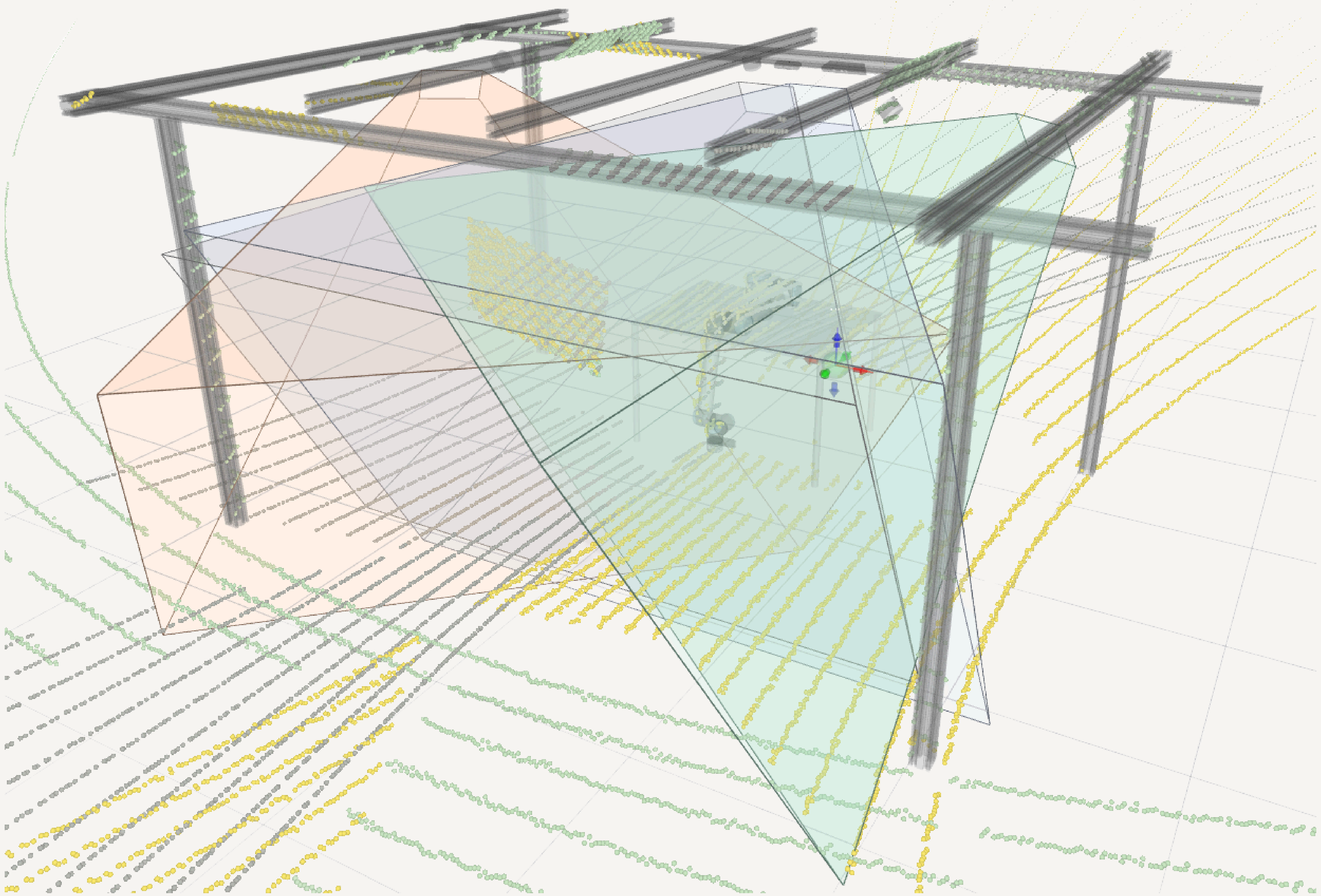
Real World

Red - RGB-D; Yellow - LiDAR; Blue - RGB

Simulation

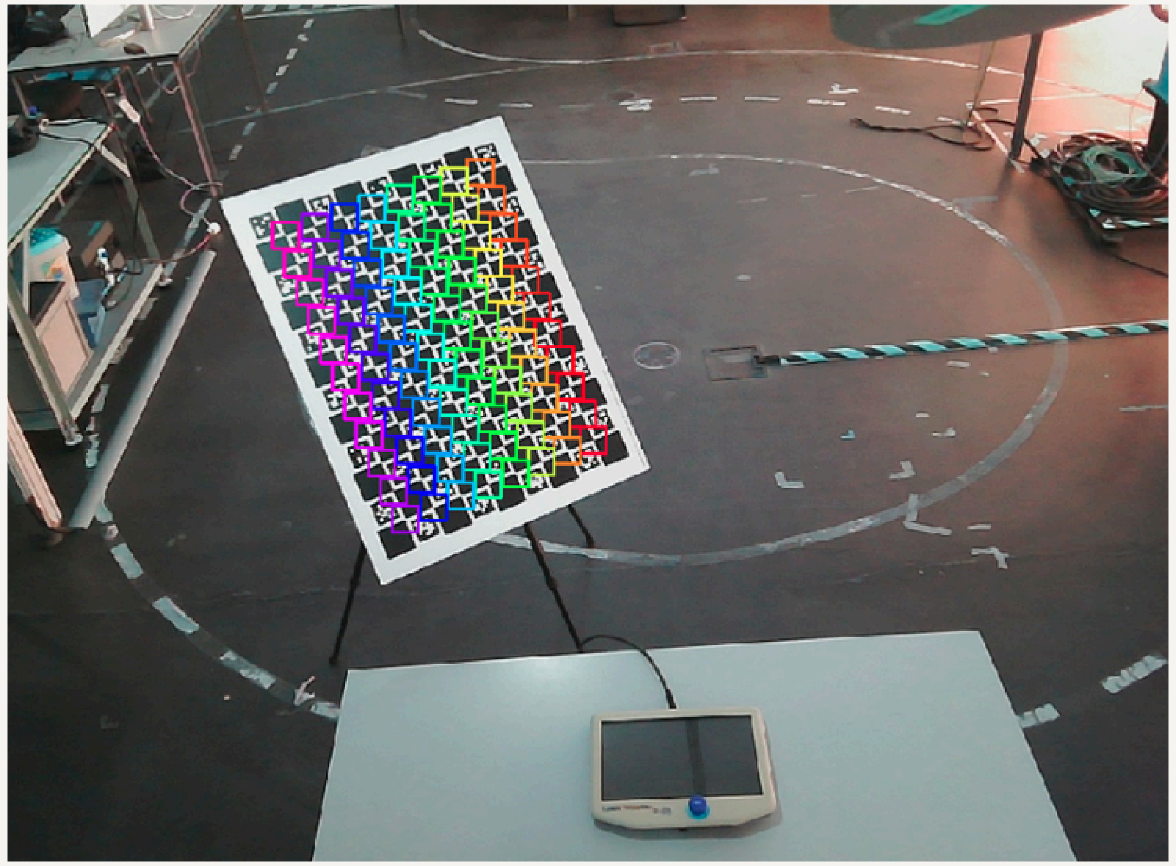


Fields-of-view

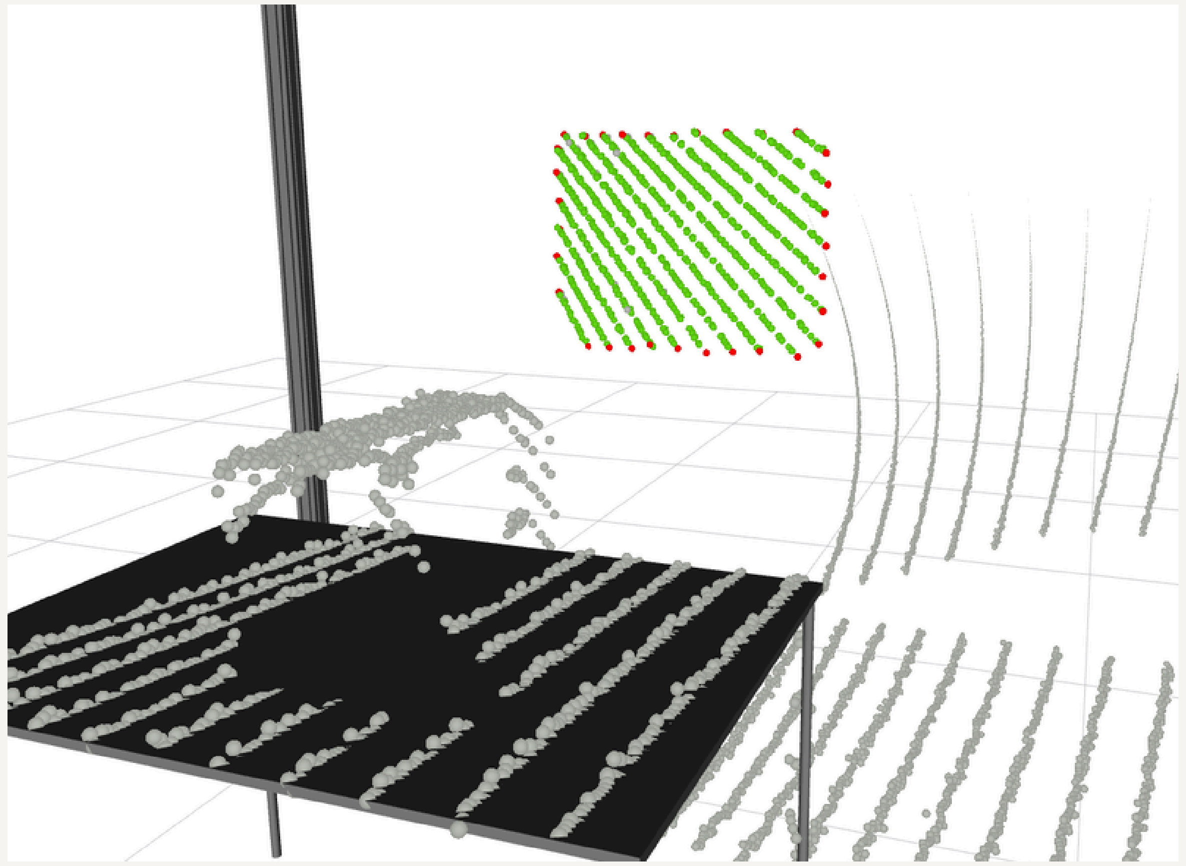


Example of labelled data

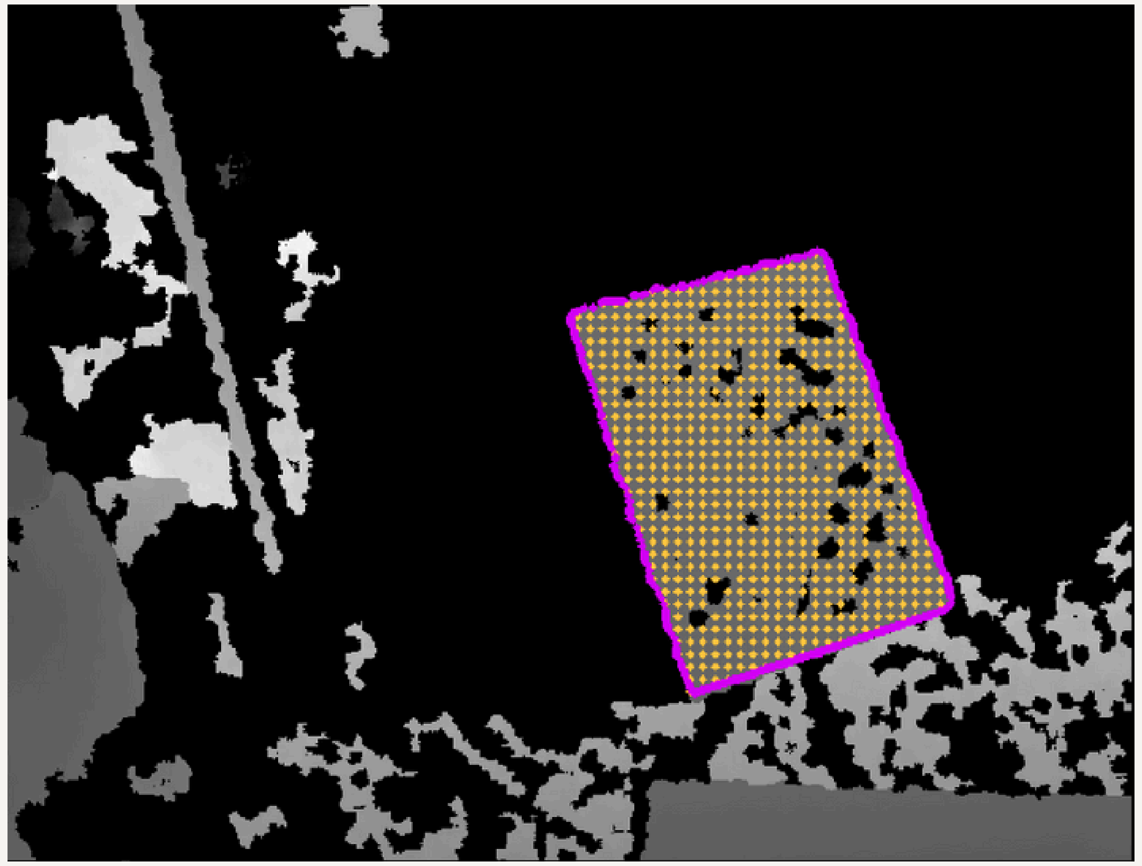
RGB



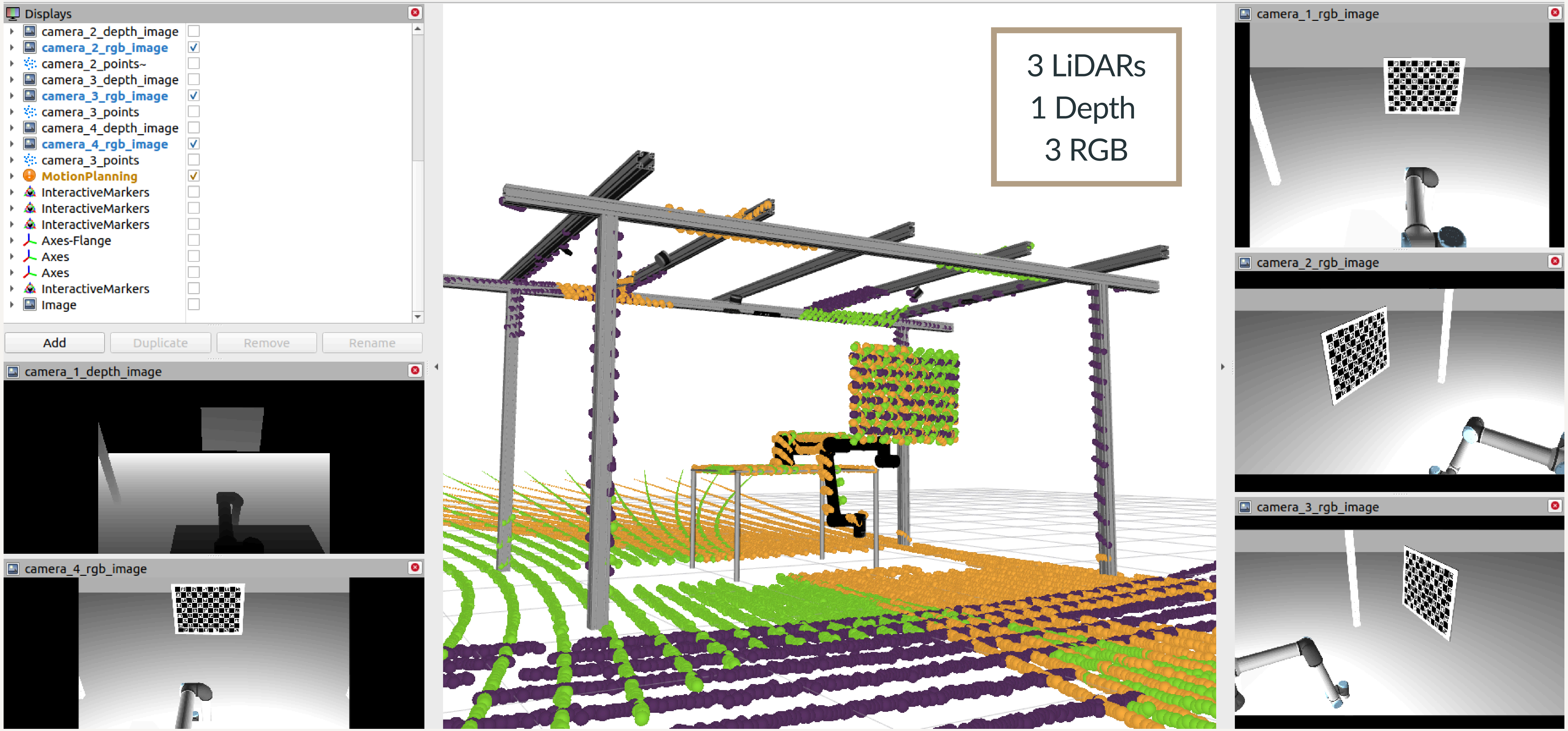
LiDAR



Depth



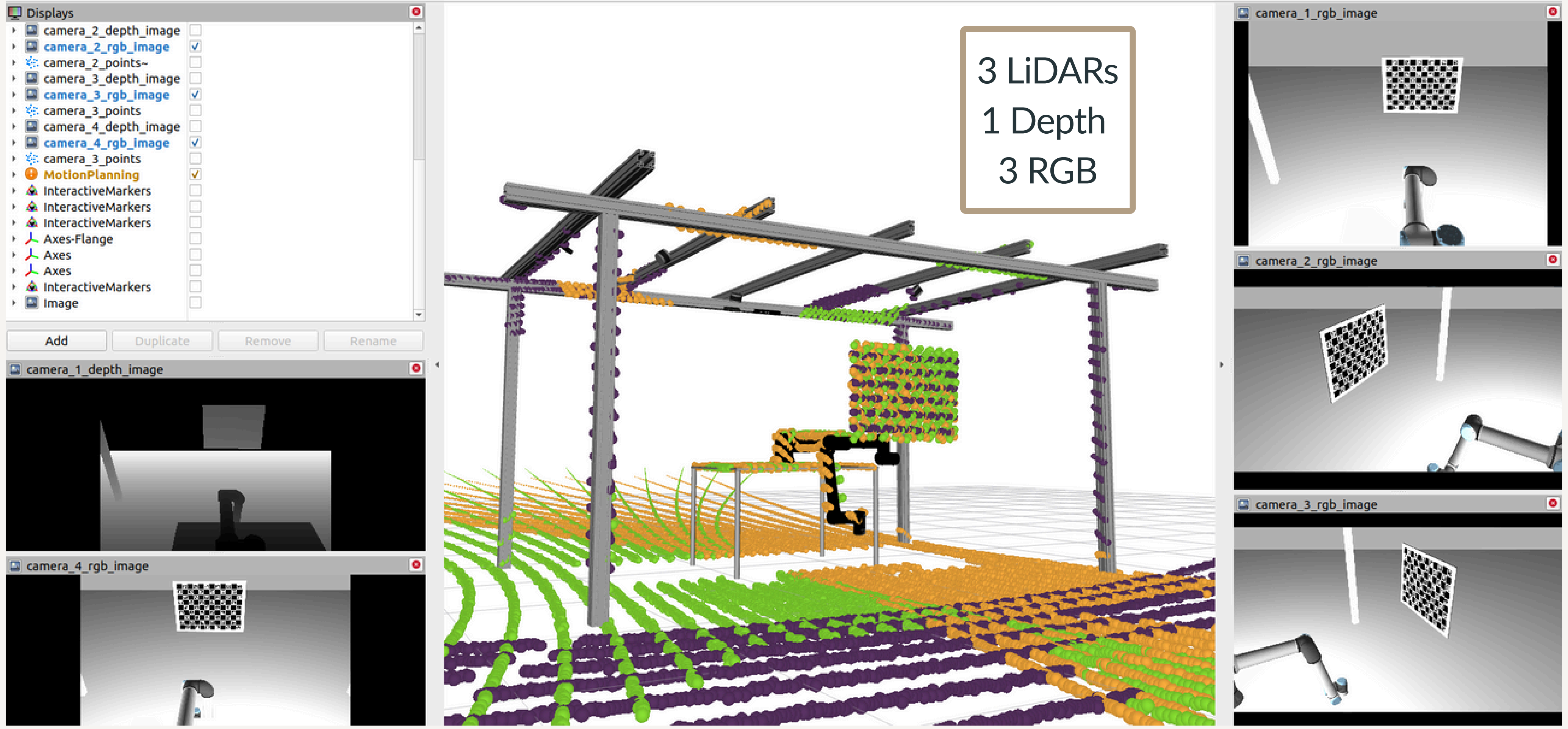
Case Study 1


Displays
camera_2_depth_image
camera_2_rgb_image
camera_2_points~
camera_3_depth_image
camera_3_rgb_image
camera_3_points
camera_4_depth_image
camera_4_rgb_image
camera_3_points
MotionPlanning
InteractiveMarkers
InteractiveMarkers
InteractiveMarkers
Axes-Flange
Axes
Axes
InteractiveMarkers
Image
Add
Duplicate
Remove
Rename
camera_1_depth_image
camera_4_rgb_image
3 LiDARs
1 Depth
3 RGB
camera_1_rgb_image
camera_2_rgb_image
camera_3_rgb_image

Case Study 1

Descriptions of the datasets used in this case study.

Type of data	Dataset	# collections	# partials	# complete
Simulation	train dataset	23	35	5
	test dataset	17	26	4
Real data	train dataset	29	61	6
	test dataset	14	29	4



Case Study 1

Comparative results

Average RGB-RGB root mean square reprojection errors in pixels.

Method	Simulation	Real
Ours	**0.6**	1.3
OpenCV	0.6	1.8
Kalibr	1.3	**0.8**

Average LiDAR-LiDAR root mean square reprojection errors in millimetres.

Method	Simulation	Real
Ours	**33.0**	53.6
ICP Initial Avg	249.4	173.4
ICP Initial Best	36.5	162.8
ICP Aligned Avg	33.9	**53.3**
ICP Aligned Best	91.7	109.9

Average LiDAR-depth root mean square reprojection errors in pixels.

Method	Simulation	Real
Ours	**1.3**	**1.8**
ICP Initial Avg	21.1	115.1
ICP Initial Best	17.8	77.9
ICP Aligned Avg	2.8	5.2
ICP Aligned Best	2.5	5.5

Average pairwise root mean square reprojection errors in pixels.

Pair	Simulation	Real
LiDAR-RGB	2.6	3.1
Depth-RGB	3.4	4.0

Case Study 1

Comparative results

Average **RGB-RGB** root mean square reprojection errors in pixels.

Method	Simulation	Real
Ours	**0.6**	1.3
OpenCV	0.6	1.8
Kalibr	1.3	**0.8**

Average **LiDAR-LiDAR** root mean square reprojection errors in millimetres.

Method	Simulation	Real
Ours	**33.0**	53.6
ICP Initial Avg	249.4	173.4
ICP Initial Best	36.5	162.8
ICP Aligned Avg	33.9	**53.3**
ICP Aligned Best	91.7	109.9

Average **LiDAR-depth** root mean square reprojection errors in pixels.

Method	Simulation	Real
Ours	**1.3**	**1.8**
ICP Initial Avg	21.1	115.1
ICP Initial Best	17.8	77.9
ICP Aligned Avg	2.8	5.2
ICP Aligned Best	2.5	5.5

Average pairwise root mean square reprojection errors in pixels.

Pair	Simulation	Real
LiDAR-RGB	2.6	3.1
Depth-RGB	3.4	4.0

Case Study 1

Comparative results

Average RGB-RGB root mean square reprojection errors in pixels.

Method	Simulation	Real
Ours	**0.6**	1.3
OpenCV	0.6	1.8
Kalibr	1.3	**0.8**

Average LiDAR-LiDAR root mean square reprojection errors in millimetres.

Method	Simulation	Real
Ours	**33.0**	53.6
ICP Initial Avg	249.4	173.4
ICP Initial Best	36.5	162.8
ICP Aligned Avg	33.9	**53.3**
ICP Aligned Best	91.7	109.9

Average LiDAR-depth root mean square reprojection errors in pixels.

Method	Simulation	Real
Ours	**1.3**	**1.8**
ICP Initial Avg	21.1	115.1
ICP Initial Best	17.8	77.9
ICP Aligned Avg	2.8	5.2
ICP Aligned Best	2.5	5.5

Average pairwise root mean square reprojection errors in pixels.

Pair	Simulation	Real
LiDAR-RGB	2.6	3.1
Depth-RGB	3.4	4.0

Case Study 1

Comparative results

Average RGB-RGB root mean square reprojection errors in pixels.

Method	Simulation	Real
Ours	**0.6**	1.3
OpenCV	0.6	1.8
Kalibr	1.3	**0.8**

Average LiDAR-LiDAR root mean square reprojection errors in millimetres.

Method	Simulation	Real
Ours	**33.0**	53.6
ICP Initial Avg	249.4	173.4
ICP Initial Best	36.5	162.8
ICP Aligned Avg	33.9	**53.3**
ICP Aligned Best	91.7	109.9

Average LiDAR-depth root mean square reprojection errors in pixels.

Method	Simulation	Real
Ours	**1.3**	**1.8**
ICP Initial Avg	21.1	115.1
ICP Initial Best	17.8	77.9
ICP Aligned Avg	2.8	5.2
ICP Aligned Best	2.5	5.5

Average pairwise root mean square reprojection errors in pixels.

Pair	Simulation	Real
LiDAR-RGB	2.6	3.1
Depth-RGB	3.4	4.0

Case Study 1

Comparative results

Average RGB-RGB root mean square reprojection errors in pixels.

Method	Simulation	Real
Ours	**0.6**	1.3
OpenCV	0.6	1.8
Kalibr	1.3	**0.8**

Average LiDAR-LiDAR root mean square reprojection errors in millimetres.

Method	Simulation	Real
Ours	**33.0**	53.6
ICP Initial Avg	249.4	173.4
ICP Initial Best	36.5	162.8
ICP Aligned Avg	33.9	**53.3**
ICP Aligned Best	91.7	109.9

Average LiDAR-depth root mean square reprojection errors in pixels.

Method	Simulation	Real
Ours	**1.3**	**1.8**
ICP Initial Avg	21.1	115.1
ICP Initial Best	17.8	77.9
ICP Aligned Avg	2.8	5.2
ICP Aligned Best	2.5	5.5

Average pairwise root mean square reprojection errors in pixels.

Pair	Simulation	Real
LiDAR-RGB	2.6	3.1
Depth-RGB	3.4	4.0

Case Study 1

Comparative results

Average RGB-RGB root mean square reprojection errors in pixels.

Method	Simulation	Real
Ours	**0.6**	1.3
OpenCV	0.6	1.8
Kalibr	1.3	**0.8**

Average LiDAR-LiDAR root mean square reprojection errors in millimetres.

Method	Simulation	Real
Ours	**33.0**	53.6
ICP Initial Avg	249.4	173.4
ICP Initial Best	36.5	162.8
ICP Aligned Avg	33.9	**53.3**
ICP Aligned Best	91.7	109.9

Average LiDAR-depth root mean square reprojection errors in pixels.

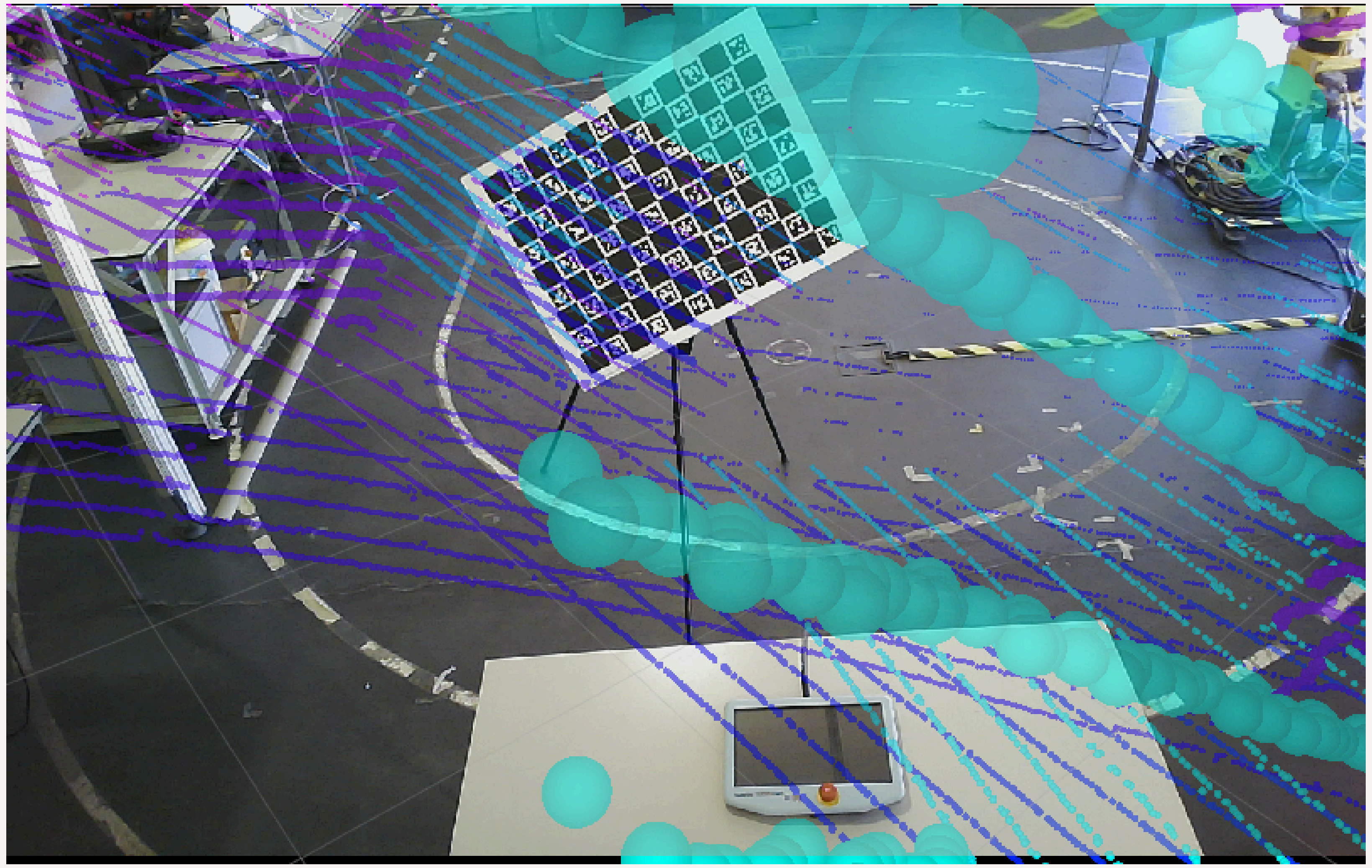
Method	Simulation	Real
Ours	**1.3**	**1.8**
ICP Initial Avg	21.1	115.1
ICP Initial Best	17.8	77.9
ICP Aligned Avg	2.8	5.2
ICP Aligned Best	2.5	5.5

Average pairwise root mean square reprojection errors in pixels.

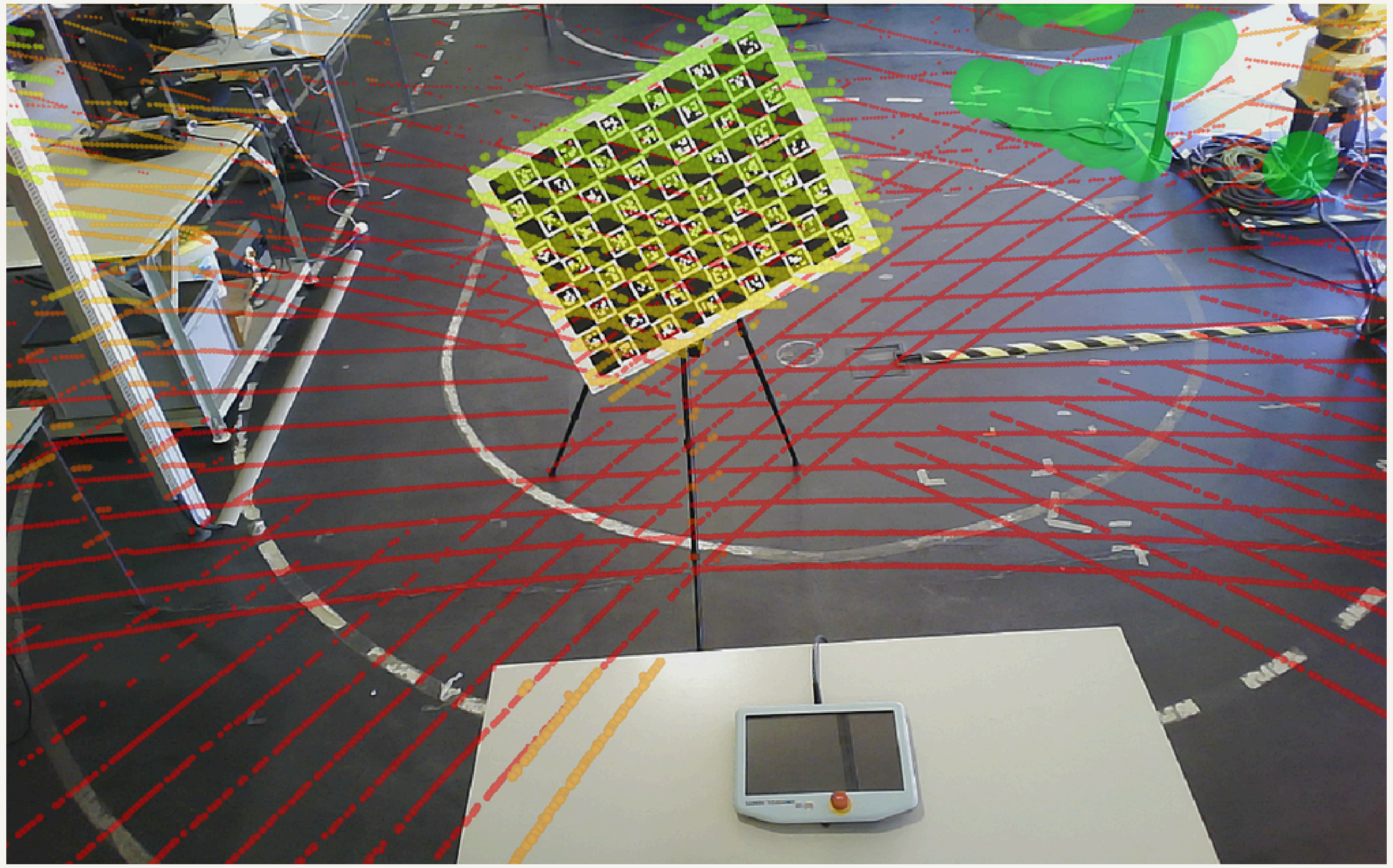
Pair	Simulation	Real
LiDAR-RGB	2.6	3.1
Depth-RGB	3.4	4.0

Case Study 1

Qualitative Results



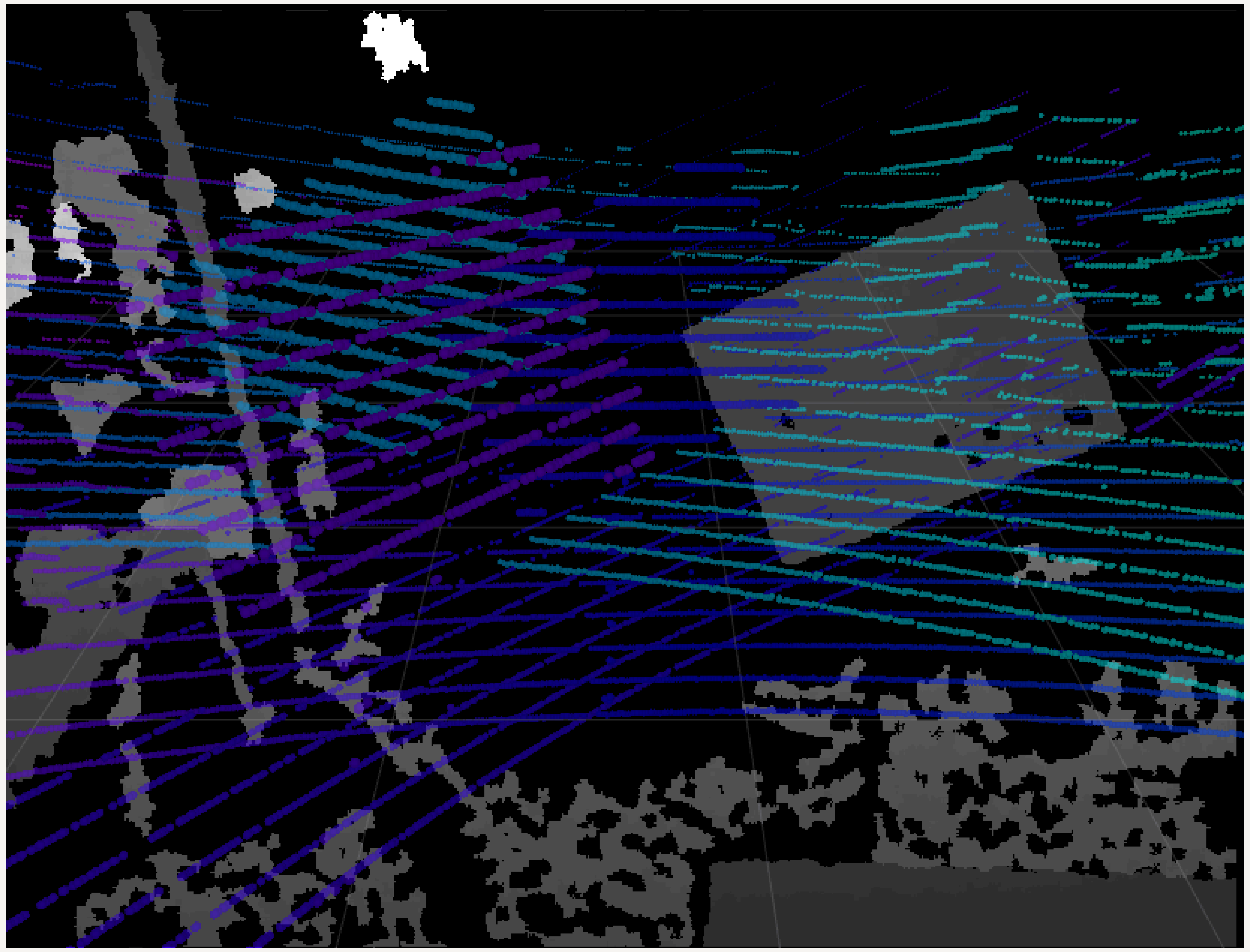
Projection of LiDAR's point clouds to an RGB image **before** calibration.



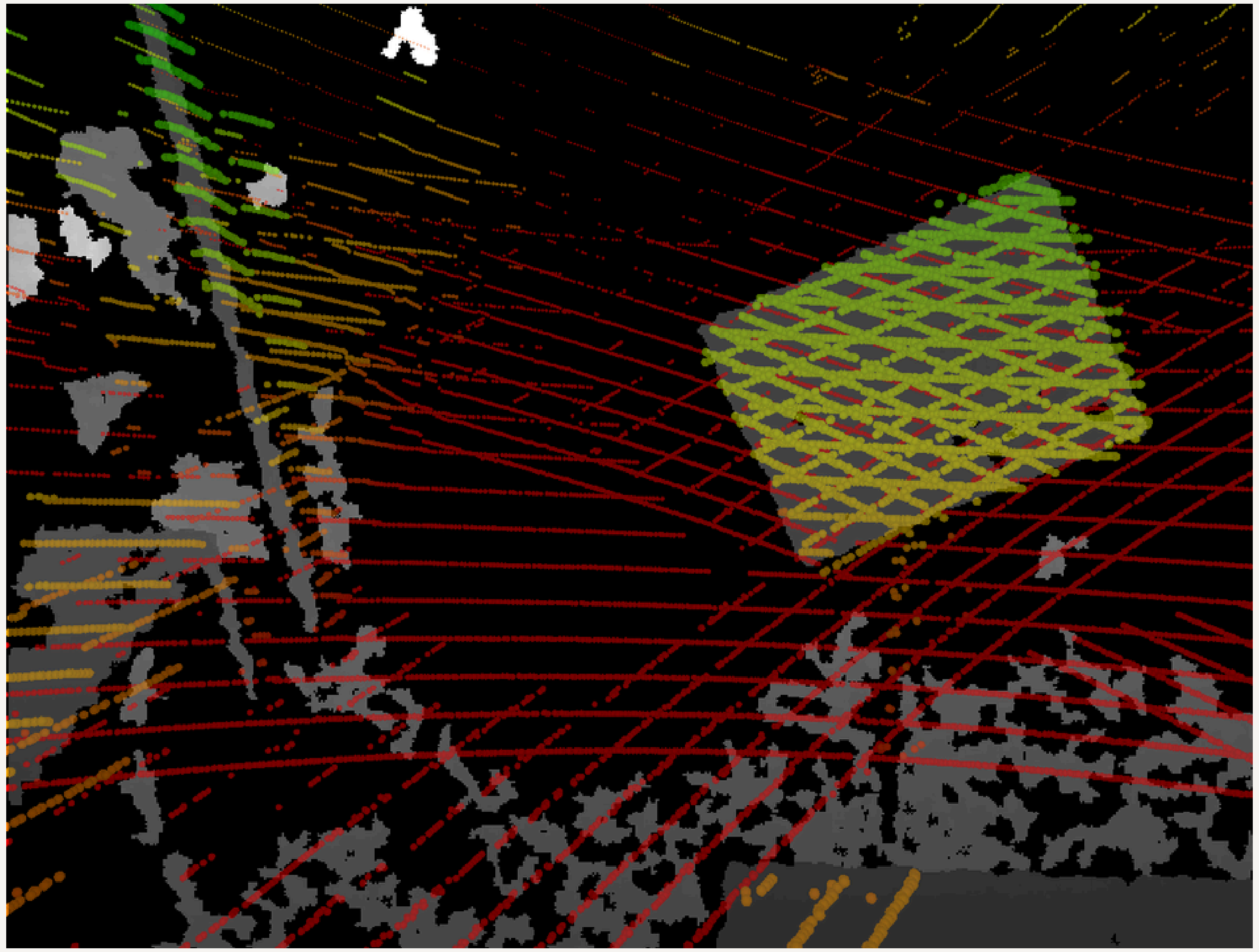
Projection of LiDAR's point clouds to an RGB image **after** calibration.

Case Study 1

Qualitative Results



Projection of LiDAR's point clouds to an RGB image **before** calibration..



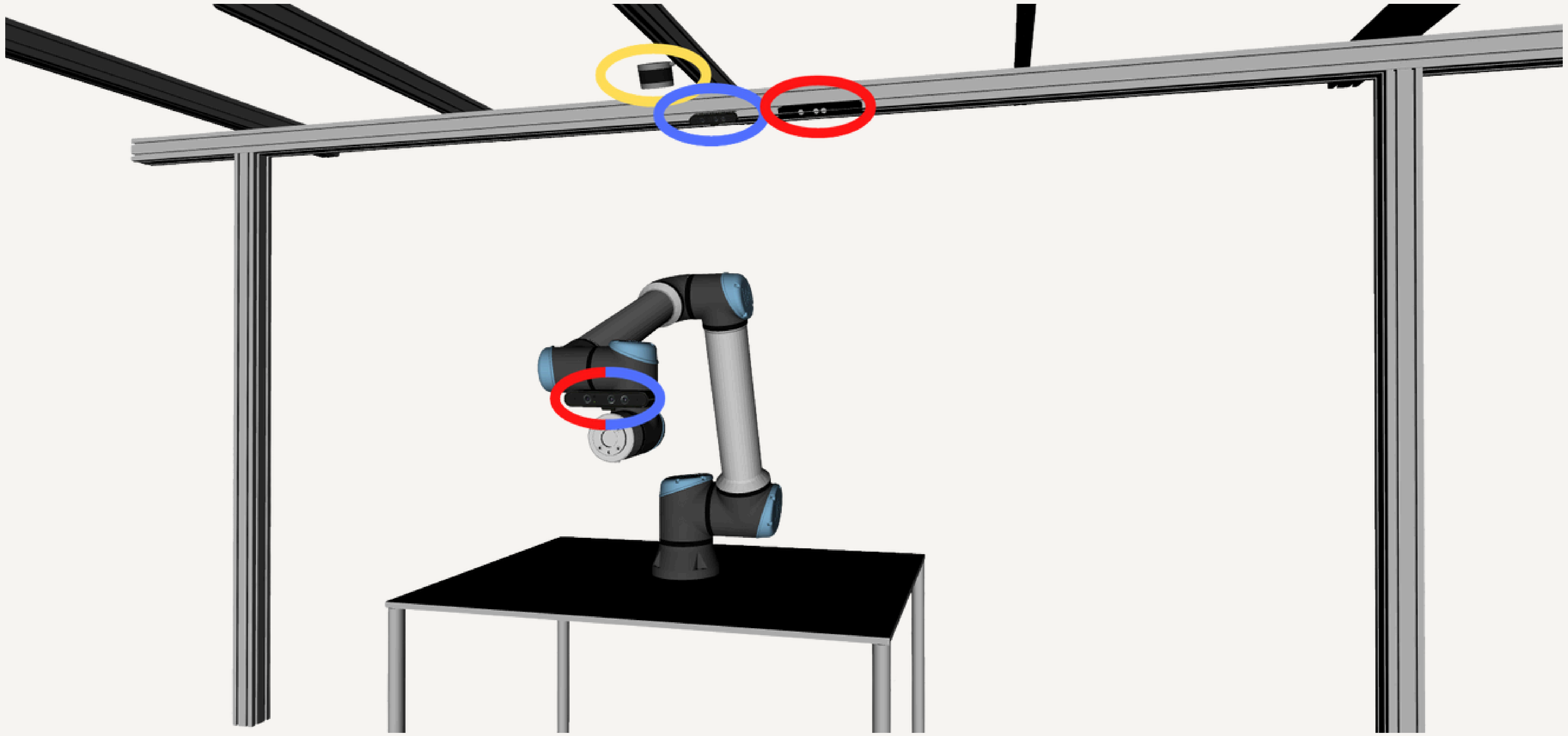
Projection of LiDAR's point clouds to an RGB image **after** calibration.

Case Study 2

Experiment definition

Descriptions of the datasets used in this experiment.

Type of data	Dataset	# collections	# partials	# complete
Simulation	train dataset	22	4	22
	test dataset	10	2	9
Real data	train dataset	29	58	11
	test dataset	13	26	5



Red - Depth; Yellow - LiDAR; Blue - RGB

Sensors

1 LiDAR
1 RGB-D (hand-eye)
1 RGB
1 Depth

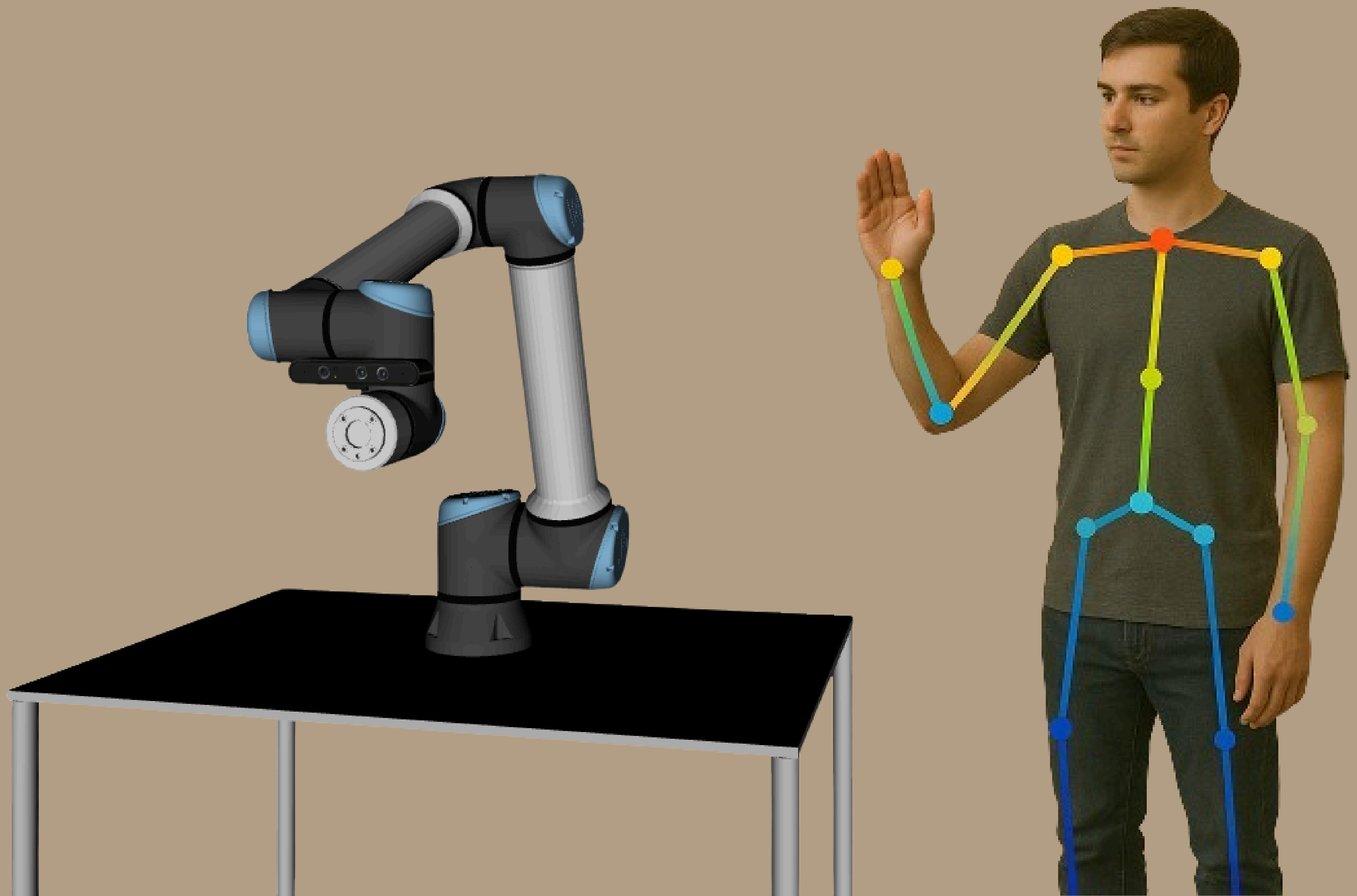
Case Study 2

Quantitative Results

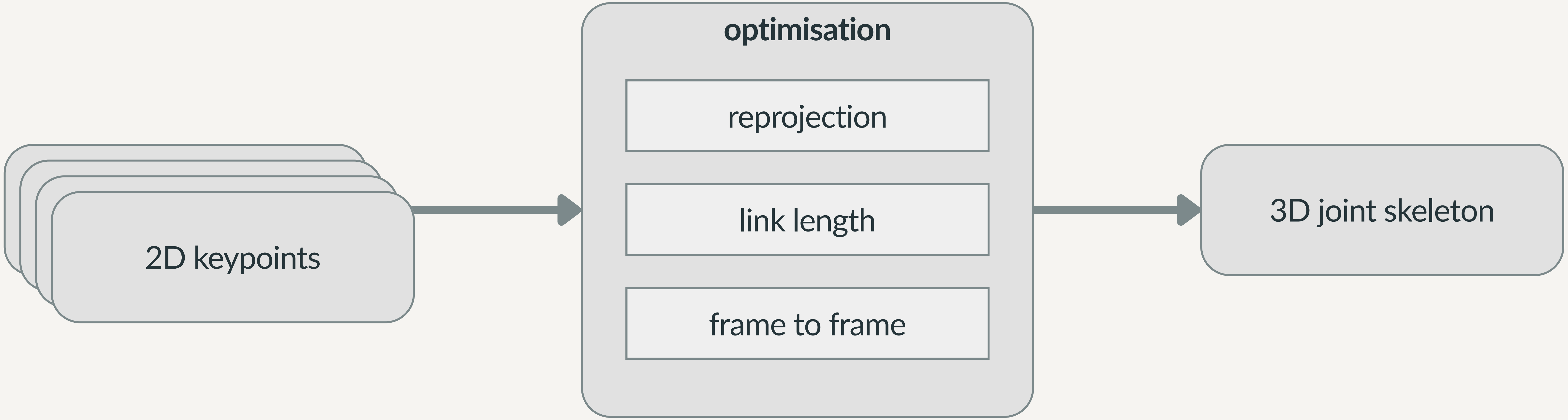
Average pairwise reprojection errors in pixels.

Pair	Simulation	Real
RGB-RGB	0.7	3.7
LiDAR-RGB	1.9	5.5
LiDAR-Depth	1.9	4.7
Depth-RGB	1.9	4.6
Depth-Depth	1.5	3.6

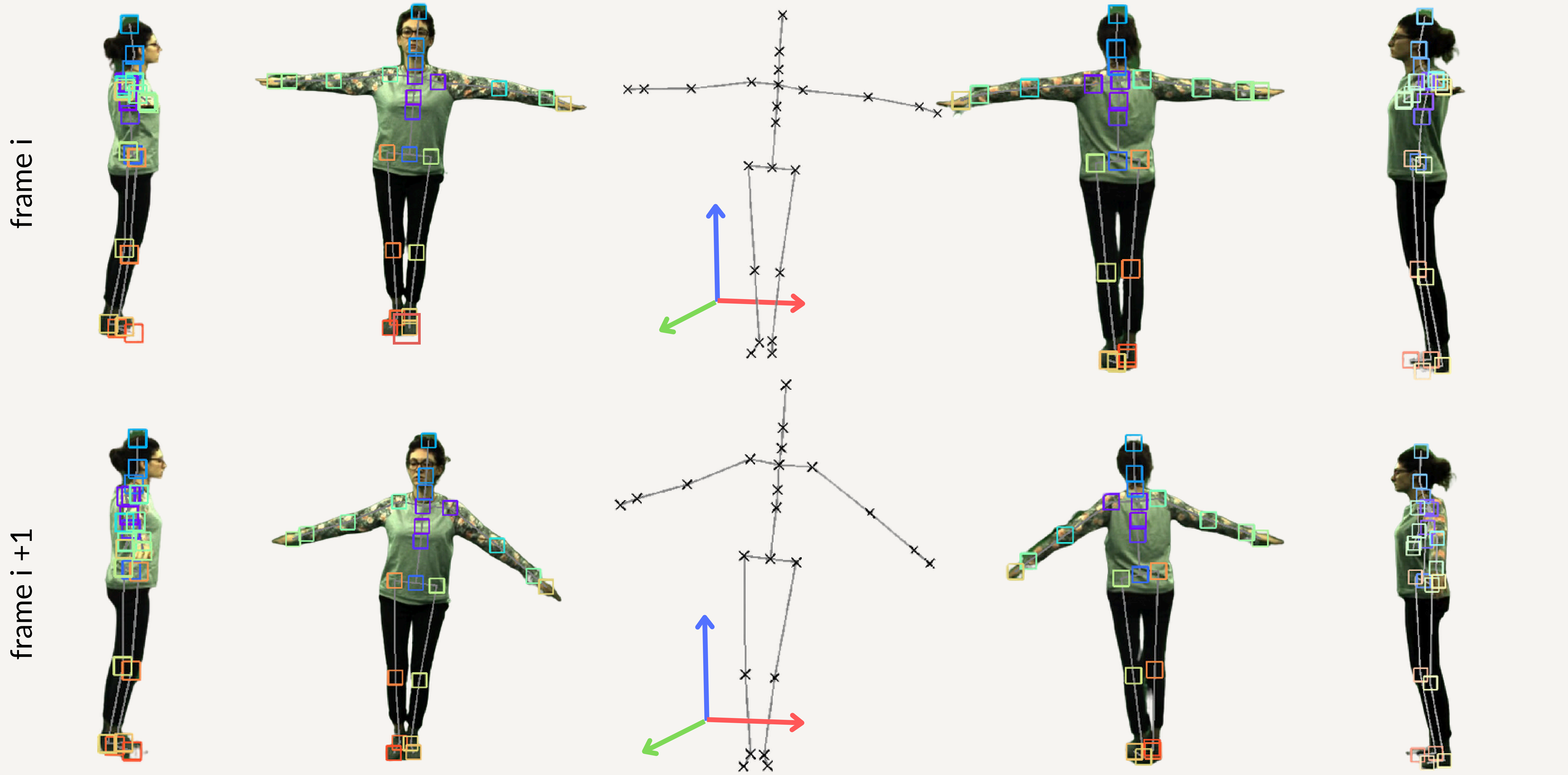
Human Pose Estimation



3D Human Pose Estimation Pipeline

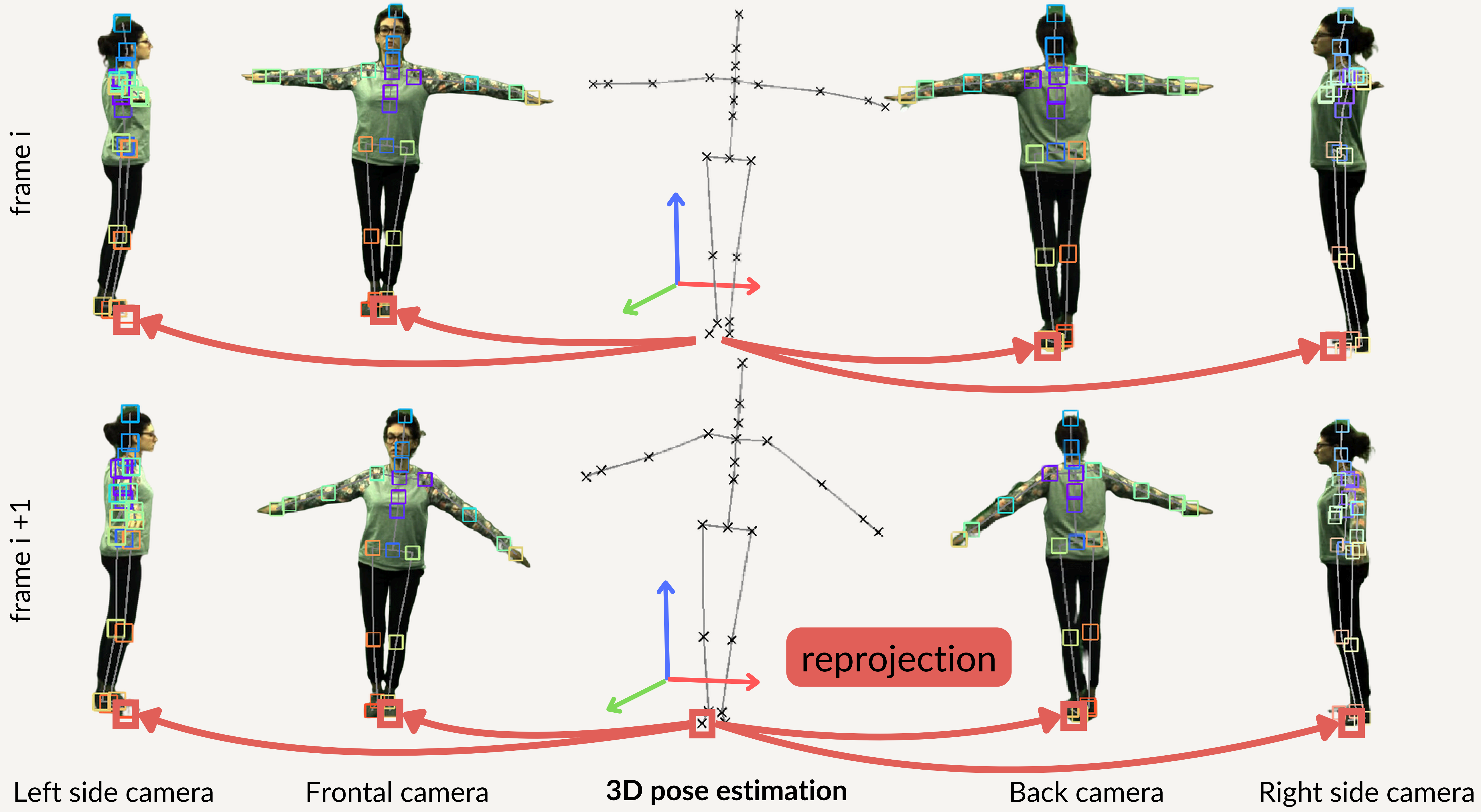


3D Human Pose Estimation

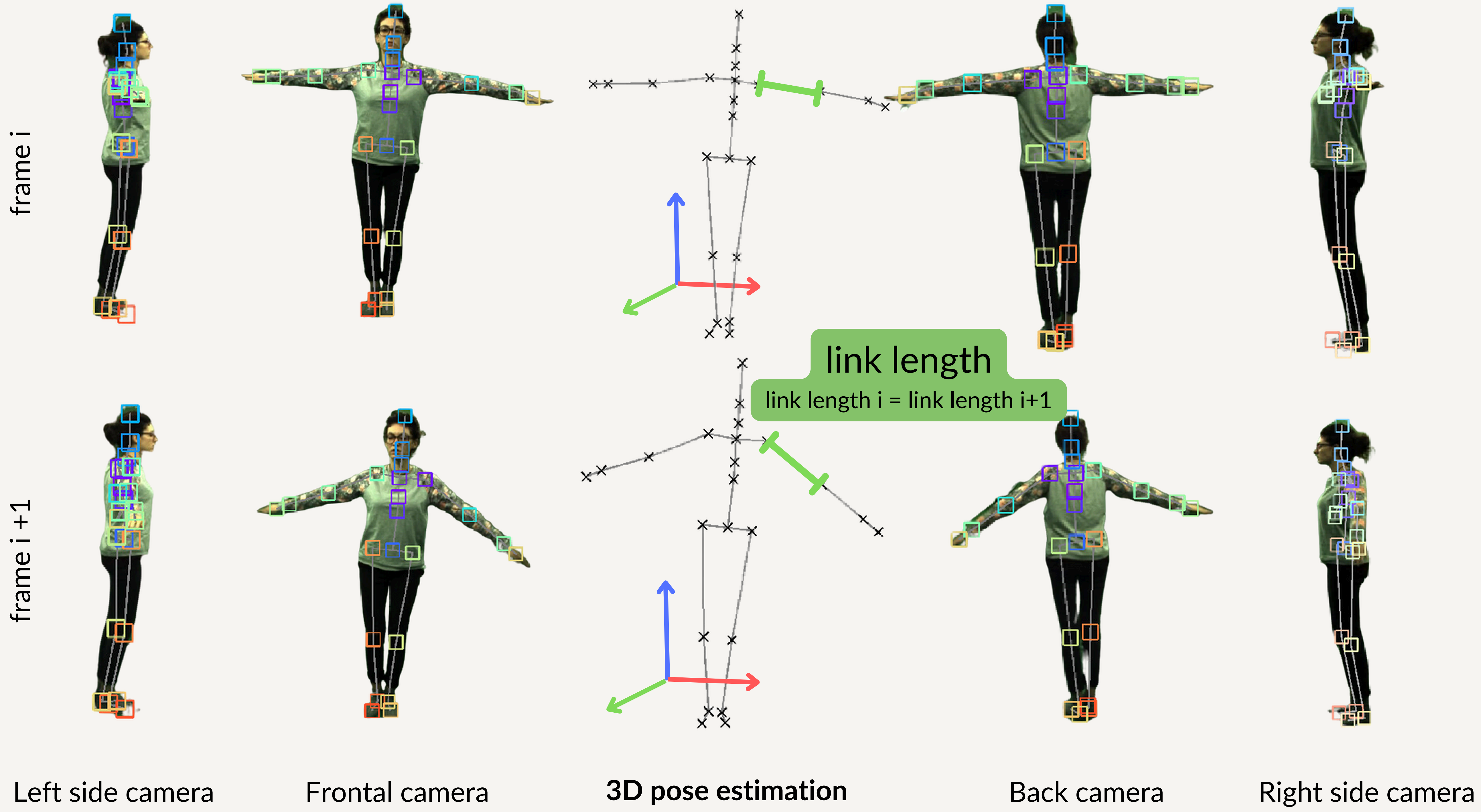


Left side camera Frontal camera **3D pose estimation** Back camera Right side camera

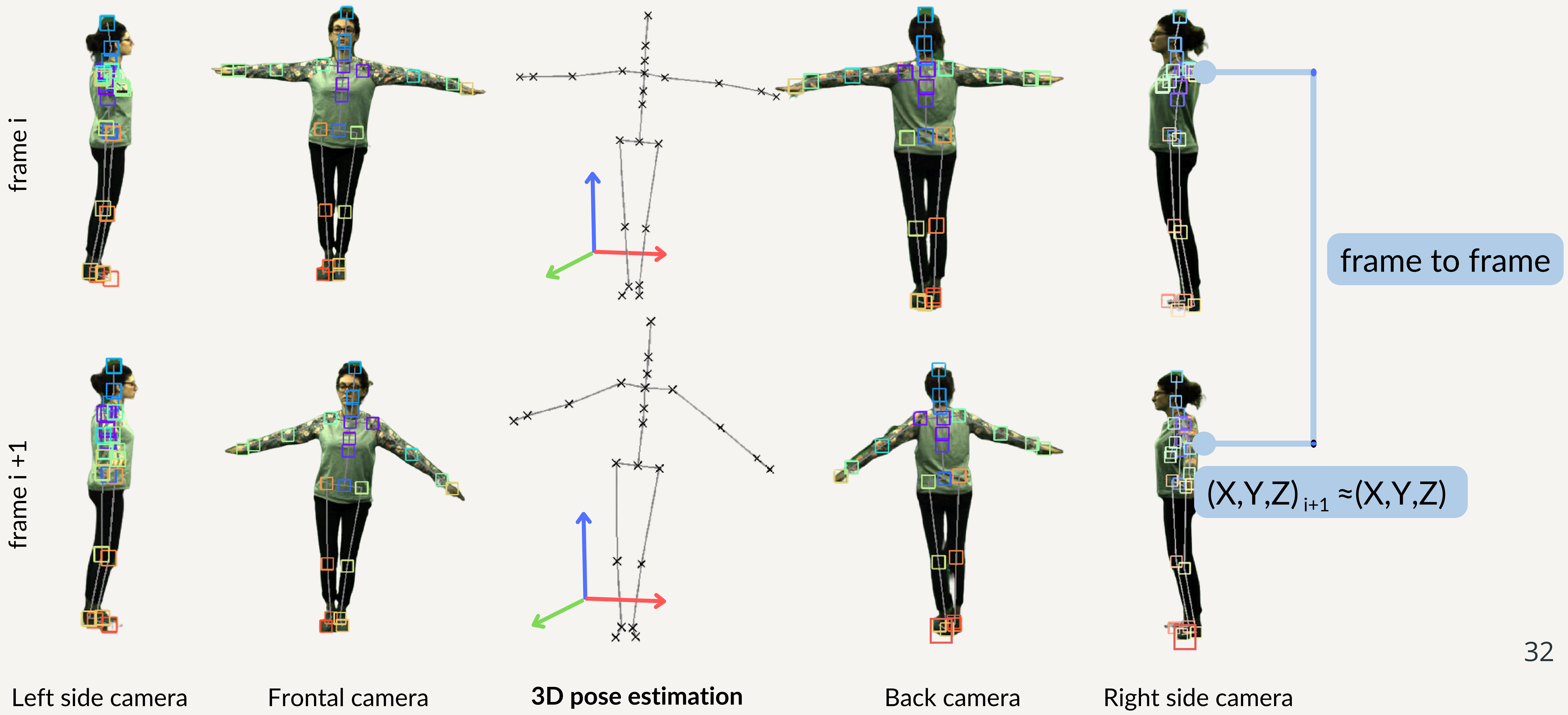
3D Human Pose Estimation



3D Human Pose Estimation



3D Human Pose Estimation



3D Human Pose Estimation

Objective Function

$$f_{obj} = \arg\min_{(X,Y,Z)_{i,f}} \sum_j e_{ll} + \sum_i \sum_f (e_{rp} + e_{ff})$$

3D Human Pose Estimation

Objective Function

$$f_{obj} = \underset{(X,Y,Z)_{i,f}}{\arg\ \min} \sum_{j} e_{ll} + \sum_{i} \sum_{f} (e_{rp} + e_{ff})$$

Reprojection Error

$$e_{rp} = \left\| proj\left((X,Y,Z)_{i,f}, \lambda_i\right) - d_{j,f,i} \right\| . c_{j,f,i}$$

joint coordinates for each frame *f* and each image *i*

intrinsic parameters

confidence value for each joint j in each frame f and each image i

2D detection coordinates

3D Human Pose Estimation

Objective Function

$$f_{obj} = \arg \min_{(X,Y,Z)_{i,f}} \sum_j e_{ll} + \sum_i \sum_f (e_{rp} + e_{ff})$$

3D Human Pose Estimation

Objective Function

$$f_{obj} = \arg\min_{(X,Y,Z)_{i,f}} \sum_j e_{ll} + \sum_i \sum_f (e_{rp} + e_{ff})$$

Link length Error

$$e_{ll} = \sqrt{\frac{\sum_f \left(l_{j,f} - \bar{l}_j\right)^2}{F}}$$

average link length for joint *j*

link length for joint *j* and frame *f*

total number of frames

3D Human Pose Estimation

Objective Function

$$f_{obj} = \arg\min_{(X,Y,Z)_{i,f}} \sum_j e_{ll} + \sum_i \sum_f (e_{rp} + e_{ff})$$

3D Human Pose Estimation

Objective Function

$$f_{obj} = \underset{(X,Y,Z)_{i,f}}{\arg\ \min} \sum_j e_{ll} + \sum_i \sum_f (e_{rp} + e_{ff})$$

Frame-to-frame Error

$$e_{ff} = \begin{cases} \|(X,Y,Z)_{i,f} - (X,Y,Z)_{j,f-1}\|, & \text{if j occluded} \\ 0, & \text{otherwise} \end{cases}$$

joint coordinates for frame *f* and each image *i*

joint coordinates for frame *f-1* and each image *i*

Quantitative Results

Experiment Details

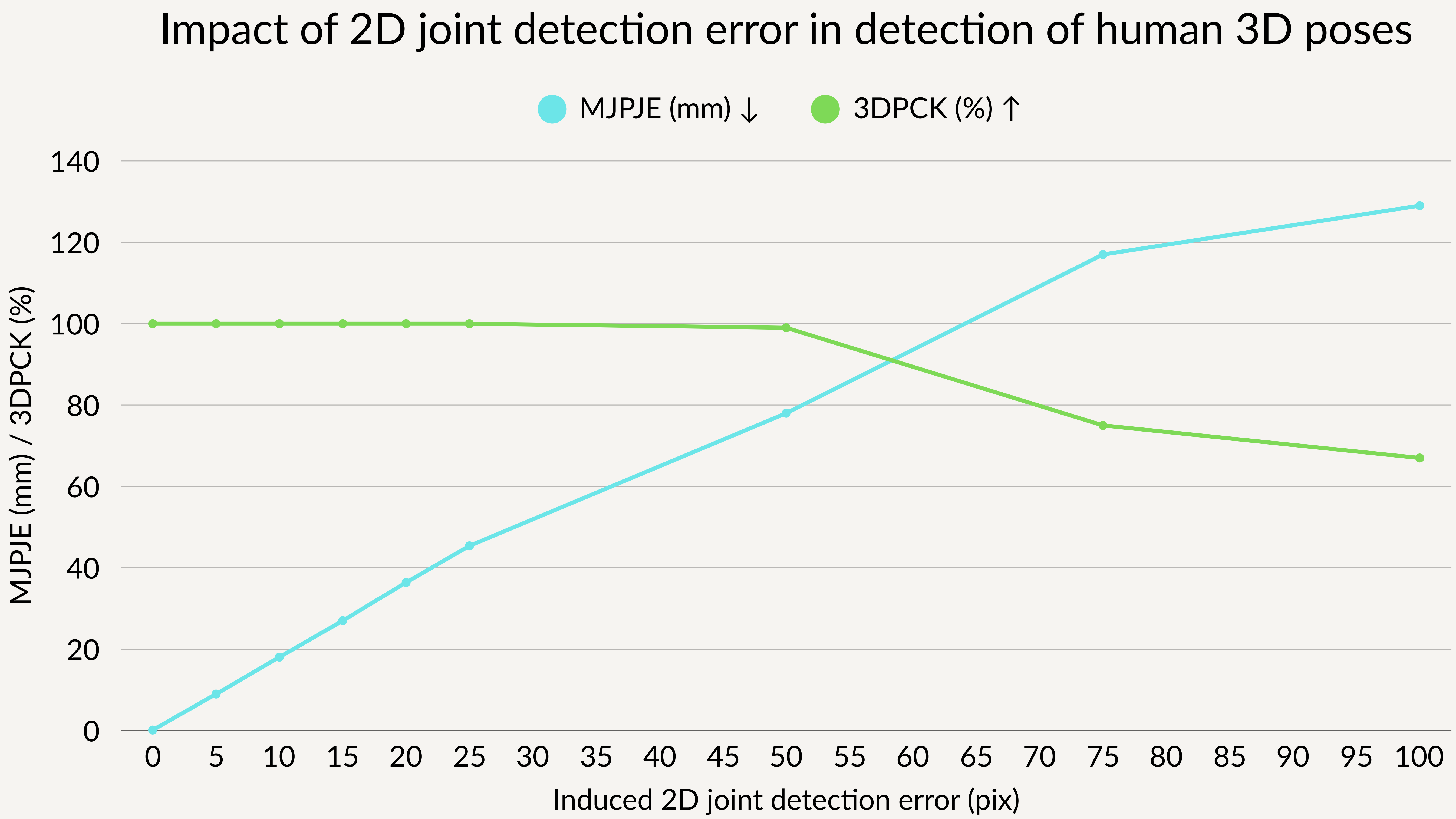
Dataset: MPI-INF-3DHP
Skeleton: 23 joints
Cameras: 4

Results reported by other state-of-the-art 2D to 3D lifting approaches.

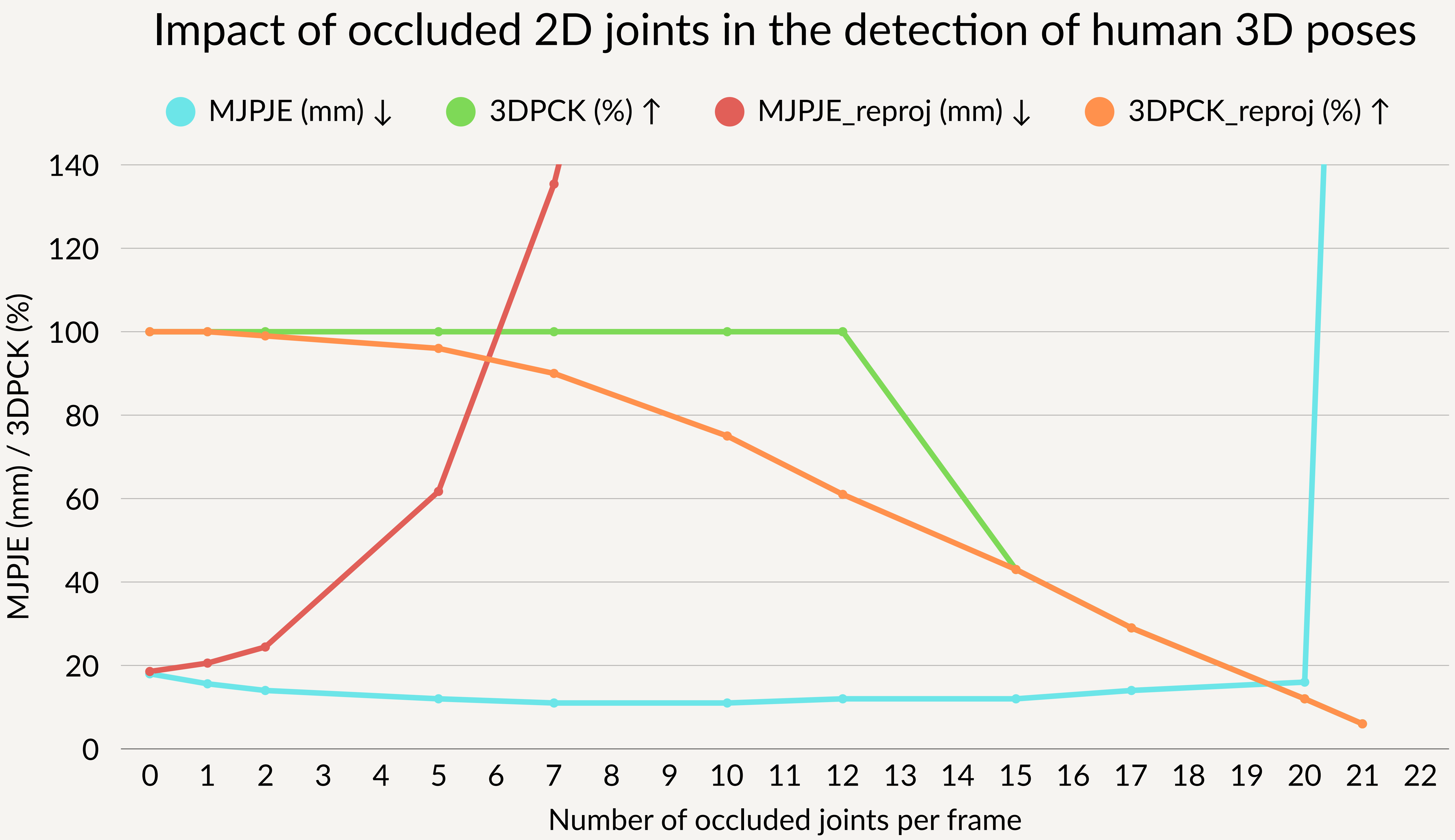
Methodology	MPJPE (mm) ↓
Kocabas et al. (CVPR 2019)	109.0
Bouazizi et al. (AVSS 2021)	93.0
Pavvlo et al. (CVPR 2019)	86.6
Bouazizi et al. (FG 2021)	65.9
Jiang et al. (WACV 2024)	55.2
Zhao et al. (CVPR 2023)	27.8
Yu et al. (CVPR 2023)	27.8
Ours (20px error)	36.4
Ours (10px error)	18.1

Mean Per Joint Position Error (MPJPE) in millimeters.

Quantitative Results



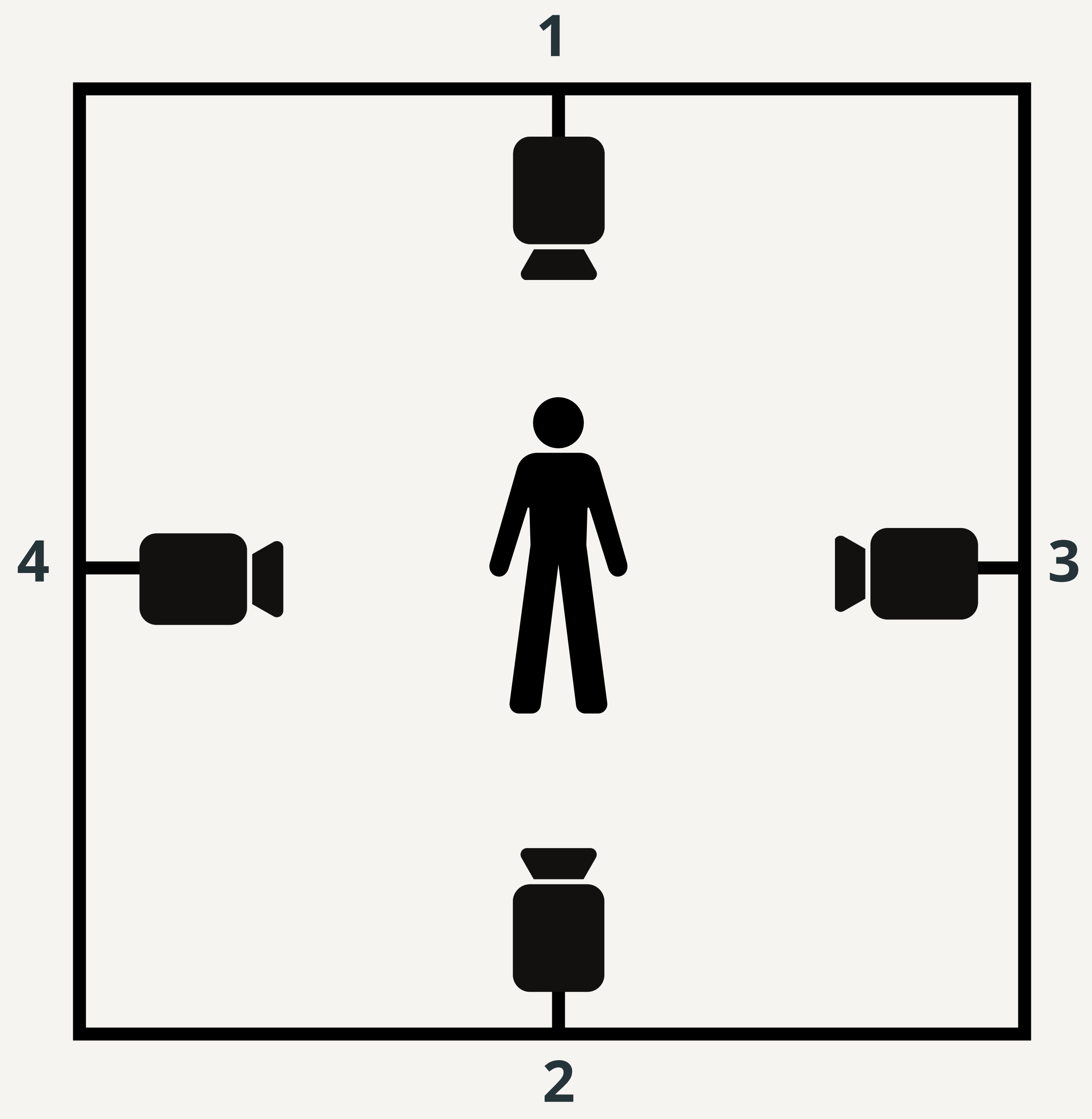
Quantitative Results



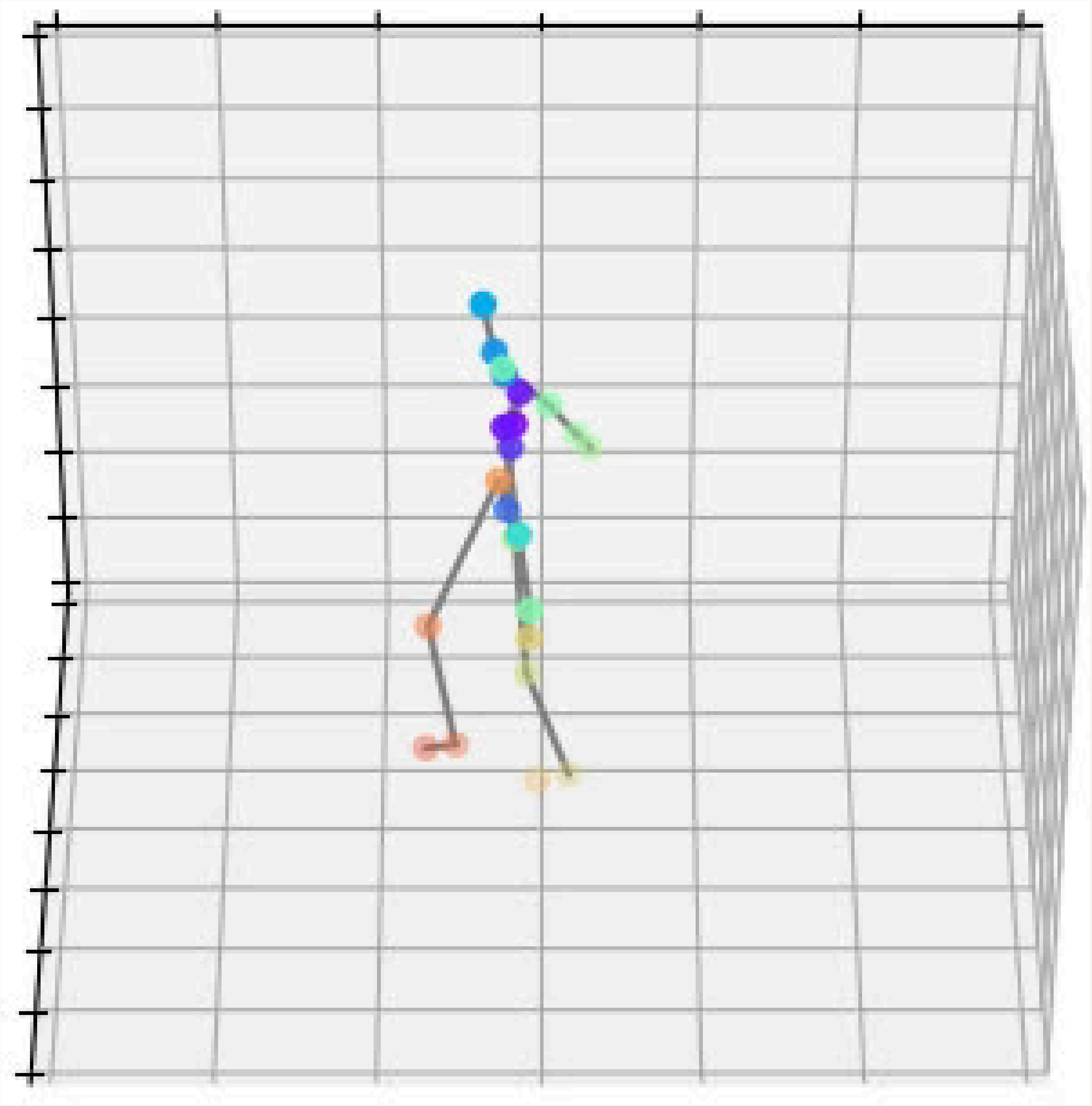
Quantitative Results

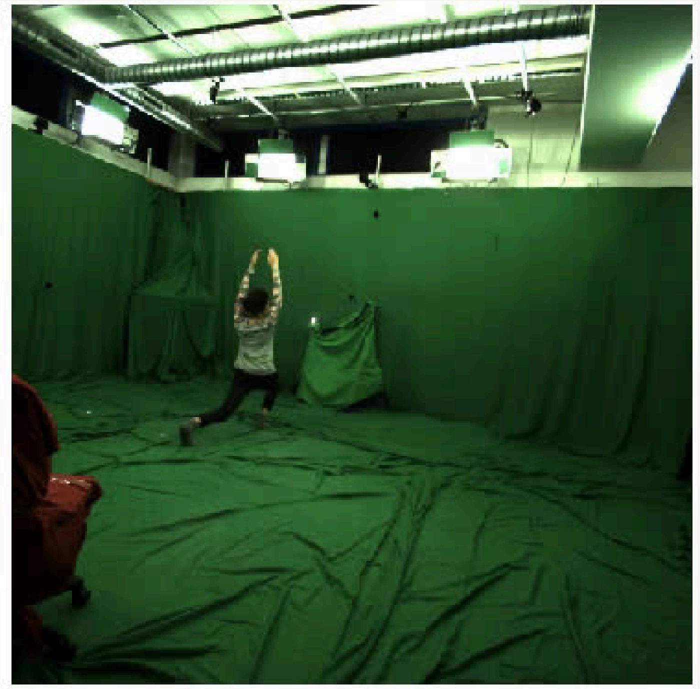
Impact of the number of cameras in the MPJPE (mm) and 3DPCK (%).

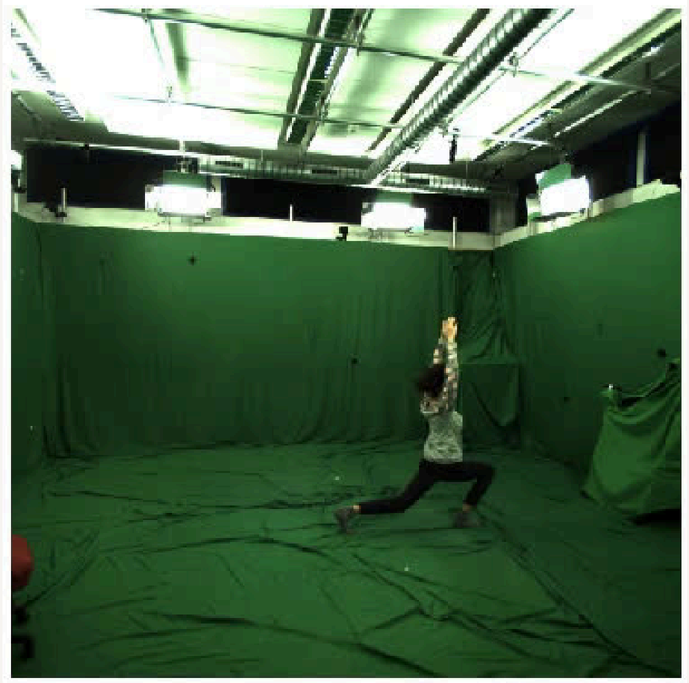
# cameras	MPJPE	3DPCK
2	47.4	96.2
3	13.8	100
4	11.6	100



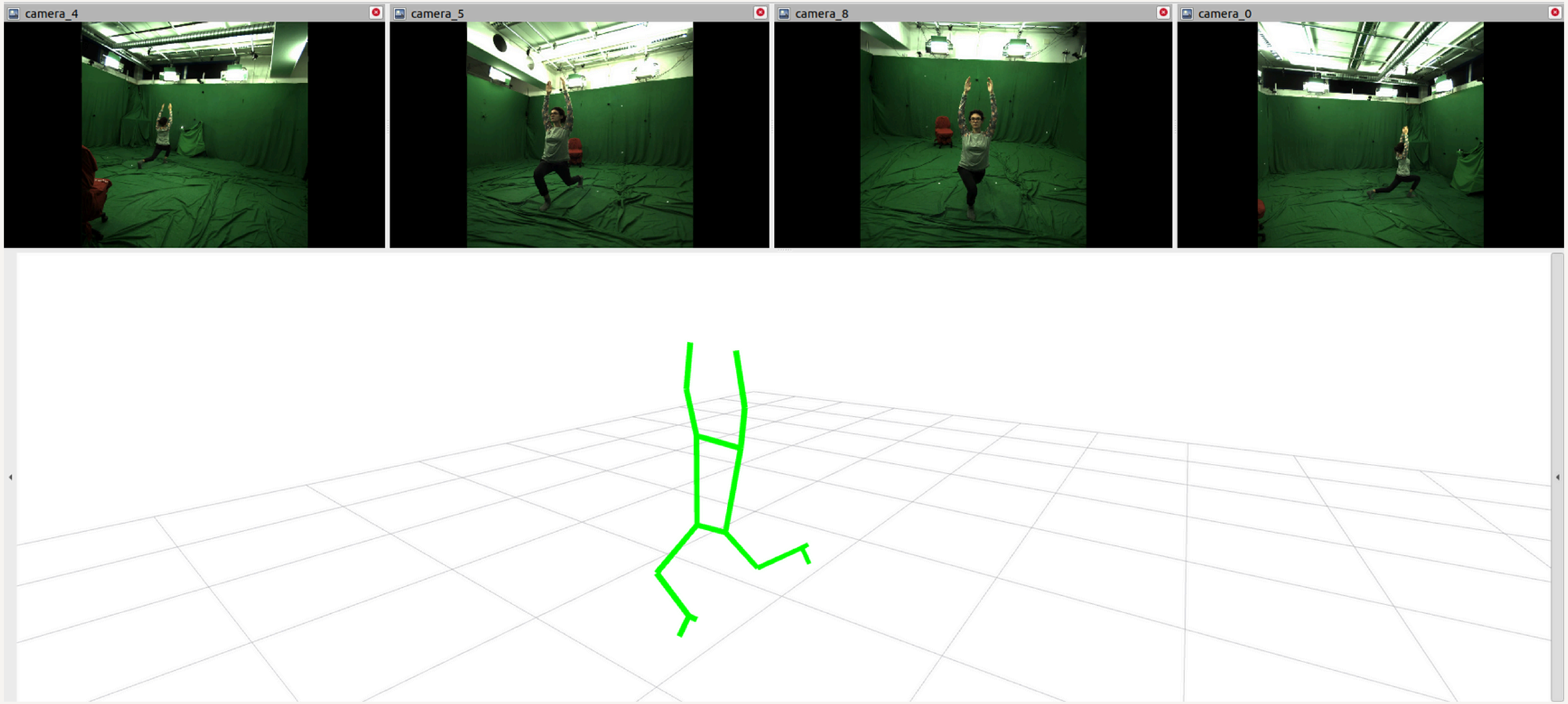
Qualitative Results







ROS Integration



Conclusions and Future Work

Contributions and Implications

Contributions

- Extended ATOM framework to support **RGB-D** cameras, including **hand–eye calibration**.
- Rigorous testing in simulated and real robotic setups, achieving **accuracy and robustness under motion**.
- Multi-camera RGB pipeline for **3D human pose estimation** in collaborative workspaces.
- Open-source implementation and integration in ROS, adaptable to **industrial collaborative cells**.

Contributions and Implications

Contributions

- Extended ATOM framework to support **RGB-D** cameras, including **hand–eye calibration**.
- Rigorous testing in simulated and real robotic setups, achieving **accuracy and robustness under motion**.
- Multi-camera RGB pipeline for **3D human pose estimation** in collaborative workspaces.
- Open-source implementation and integration in ROS, adaptable to **industrial collaborative cells**.

Implications

Scientific: Bridges two core **perception** challenges (calibration and pose estimation) in **one framework**.

Practical: Supports **safer**, more **flexible**, and **human-aware** robot collaboration.

Thesis Publications

- **Rato D.**, Oliveira M., Santos V., Gomes M., Sappa A. (2022), A sensor-to-pattern calibration framework for multi-modal industrial collaborative cells. In: **Journal of Manufacturing Systems**, doi: 10.1016/j.jmsy.2022.07.006

- **Rato D.**, Oliveira M., Santos V., Sappa A., Raducanu B. (2024), Multi-View 2D to 3D Lifting Video-Based Optimization: A Robust Approach for Human Pose Estimation with Occluded Joint Prediction. In: 2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (**IROS 2024**), doi: 10.1109/IROS58592.2024.10802200

- **Rato D.,** Oliveira M., Santos V., Sappa A., New Methodology to Calibrate Depth Sensors in Multi-Modal Dynamic Setups. Submitted: **IEEE Access**

Conclusions and Future Work

Conclusions

- **Robust calibration** in complex robotic systems is feasible with the extended ATOM framework.
- Multi-camera RGB pose estimation significantly improves **accuracy and robustness under occlusion**.
- Both are essential building blocks for **perception-driven human–robot collaboration**.

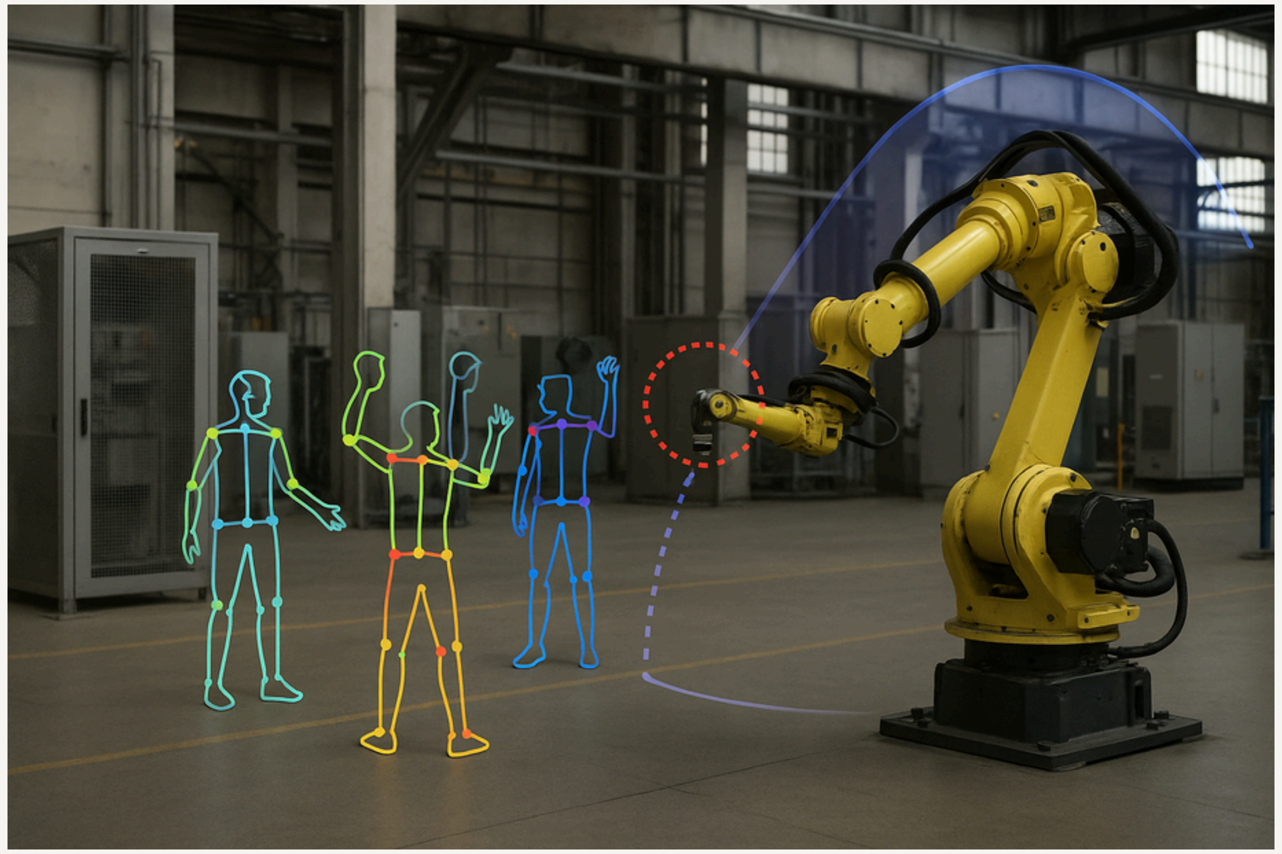
Conclusions and Future Work

Conclusions

- **Robust calibration** in complex robotic systems is feasible with the extended ATOM framework.
- Multi-camera RGB pose estimation significantly improves **accuracy and robustness under occlusion**.
- Both are essential building blocks for **perception-driven human–robot collaboration**.

Future Work

- Online / **adaptive calibration** for dynamic setups.
- **Multi-person pose estimation** with temporal consistency.
- Integration with **robot control** for predictive and safety-aware behaviour.
- **Industrial benchmarking** under diverse, real-world conditions.



Thank You

The author acknowledges the support of the Project Augmented Humanity [POCI-01-0247-FEDER-046103] and the CYTED Network: Ibero-American Thematic Network on ICT Applications for Smart Cities (RF-518RT0559).

This work was supported by FCT - Fundação para a Ciência e Tecnologia, I.P. by project reference 2021.04792.BD and DOI identifier https://doi.org/10.54499/2021.04792.BD.

fct Fundação para a Ciência e a Tecnologia

UNIÃO EUROPEIA
Fundo Social Europeu

COMPETE 2020

Lisb@2020

PORTUGAL 2020

CENTRO 2020