Perception devices and systems for monitoring humans and production process in Industry 4.0 scenarios

PhD Student: Laura Romeo (laura.romeo@stiima.cnr.it, laura.romeo@poliba.it)

Tutors:

Prof. Anna Gina Perri

Ing. Roberto Marani

DEI Doctoral Research Seminars

Bari, 23/02/2022

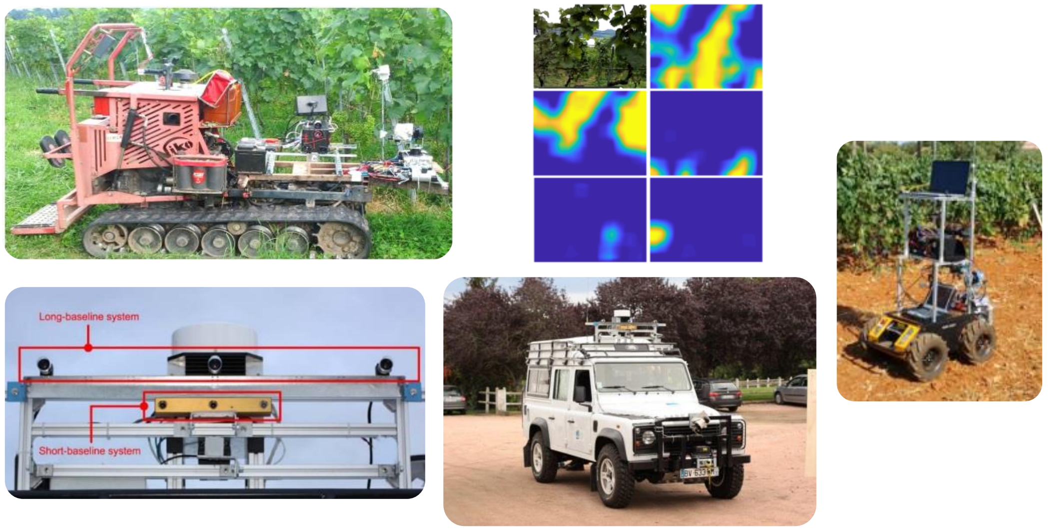
Outline

- Introduction
- System Overview
- Azure Kinect: Specifics and Analysis
- Setup Assessment:
  - Multi-vision System
  - Action Recognition
- Conclusion and Future Works

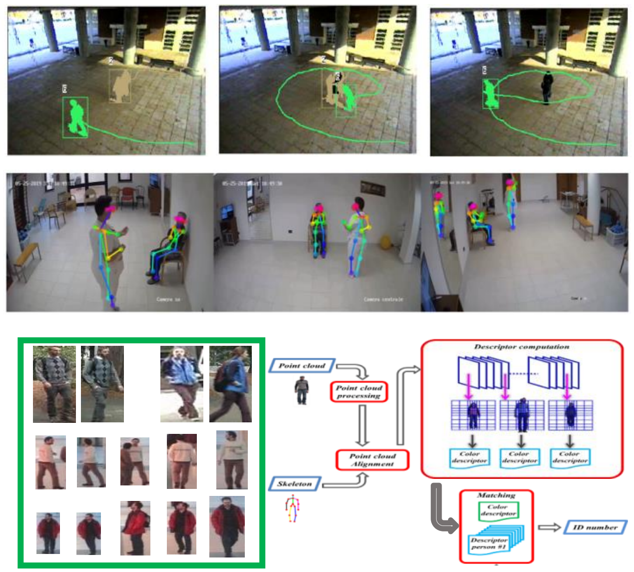
laura.romeo@stiima.cnr.it, laura.romeo@poliba.it

Introduction

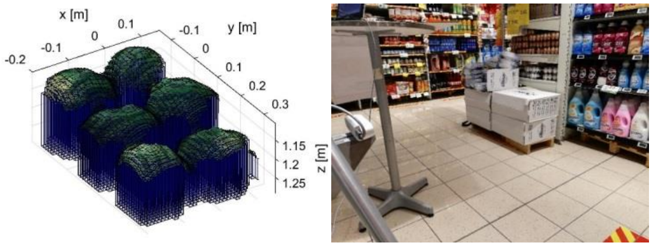
Agricultural Robotics



Ambient Intelligence



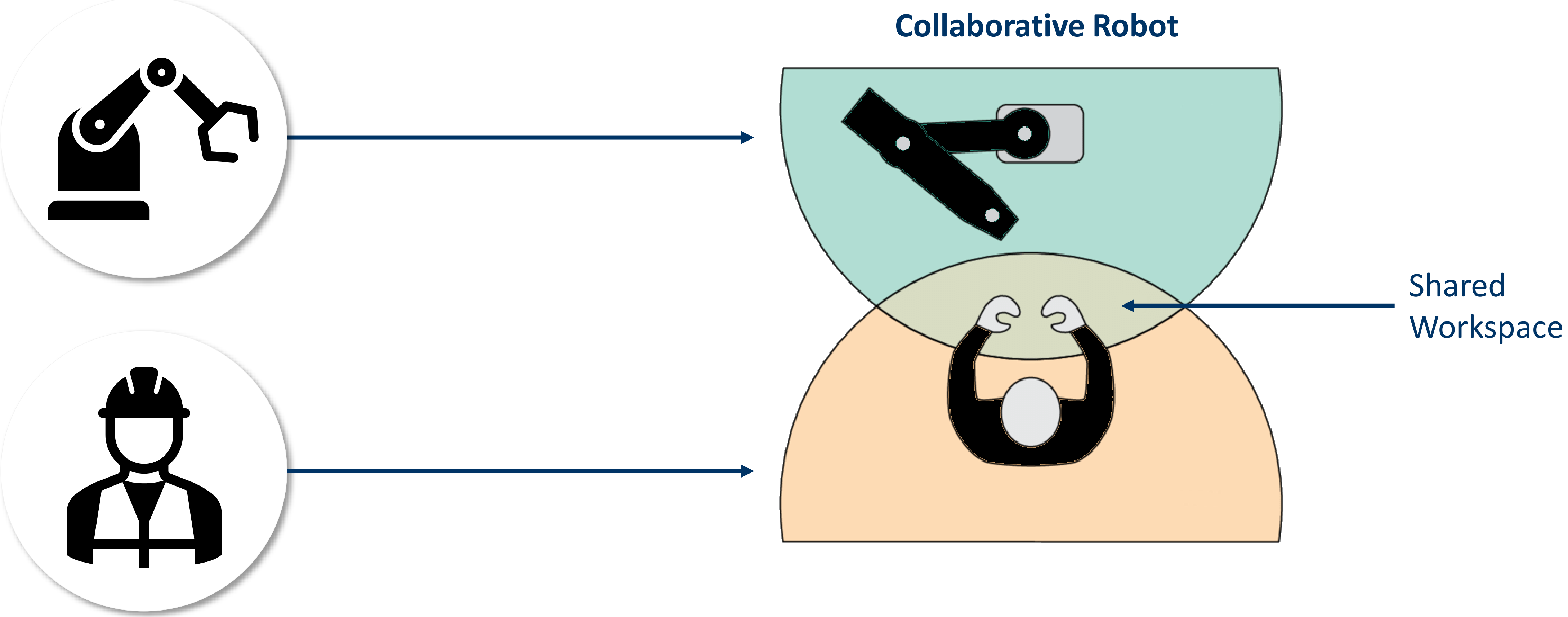
Smart Retail



Robot and Sensor Networks

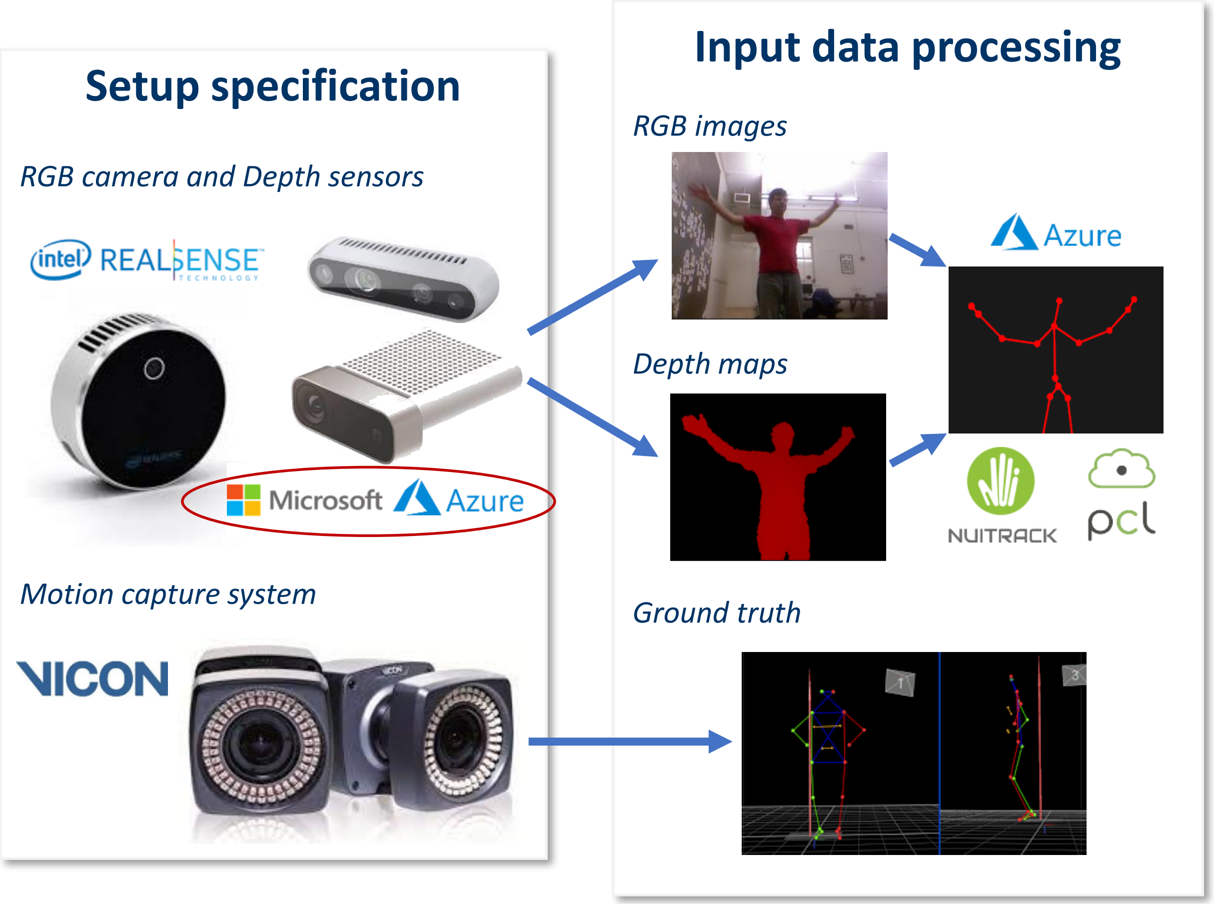
laura.romeo@stiima.cnr.it, laura.romeo@poliba.it

Introduction



laura.romeo@stiima.cnr.it, laura.romeo@poliba.it

Introduction



RGB camera:

- CMOS, CCD standard sensors

Depth camera:

- Stereo
- Time of Flight (ToF)

Skeleton extraction:

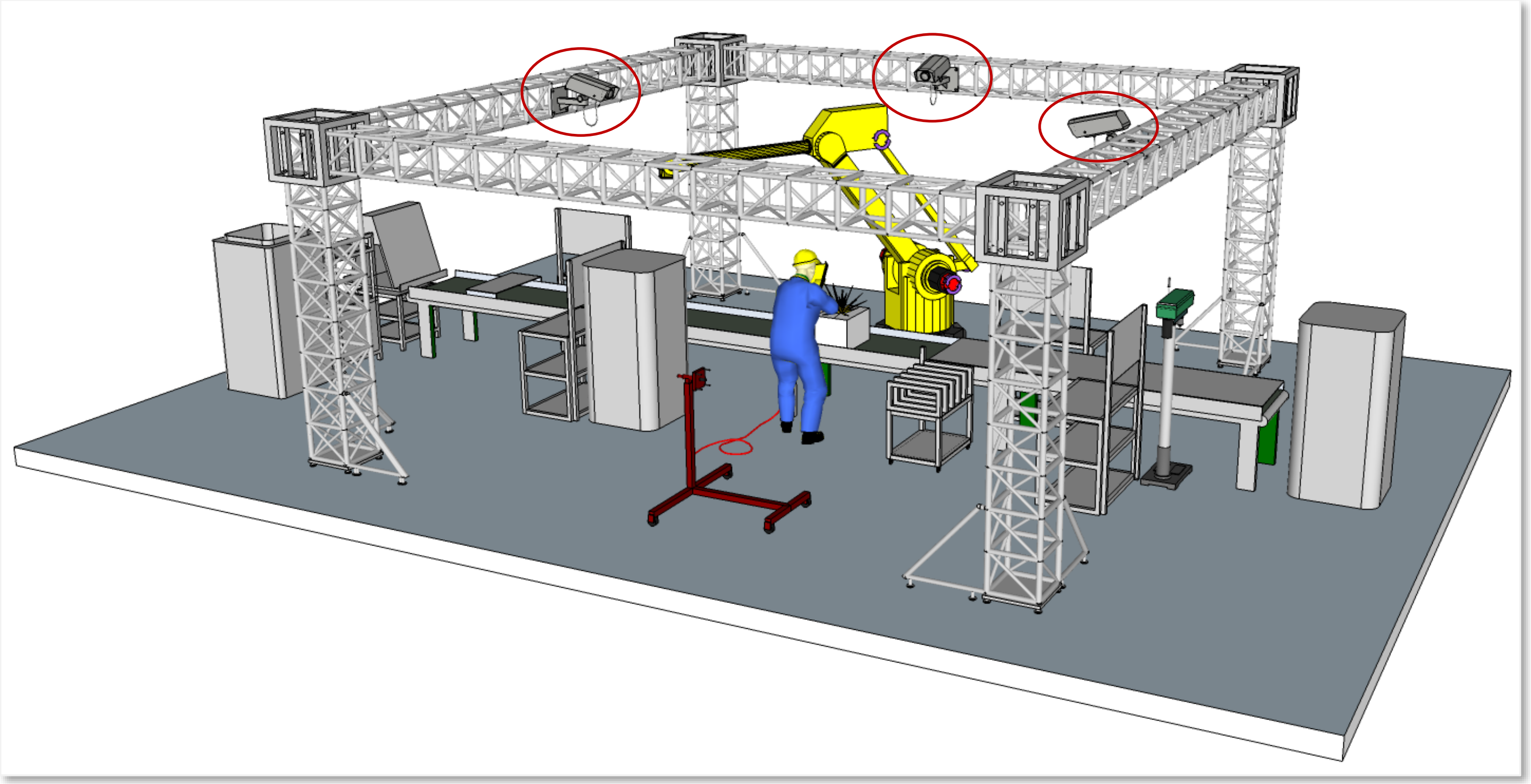
- VICON
- Azure Body Tracking
- Nutrack
- …

laura.romeo@stiima.cnr.it, *laura.romeo@poliba.it*

System Overview

Aim:

- Provide information about the dynamic occupancy (even predicting movements) of humans within the cell
- Establish a new way of communication between cobots and operators



Romeo, L., Petitti, A., Marani, R., & Milella, A. (2020). "Internet of robotic things in smart domains: applications and challenges". *Sensors*, 20(12), 3355.

laura.romeo@stiima.cnr.it, *laura.romeo@poliba.it*

System Overview

Microsoft Azure Kinect SDK

- RGB camera with Depth sensor
- Spatial microphone
- Motion sensors (Gyroscope and Accelerometer)
- Advanced AI sensors that provide sophisticated computer vision

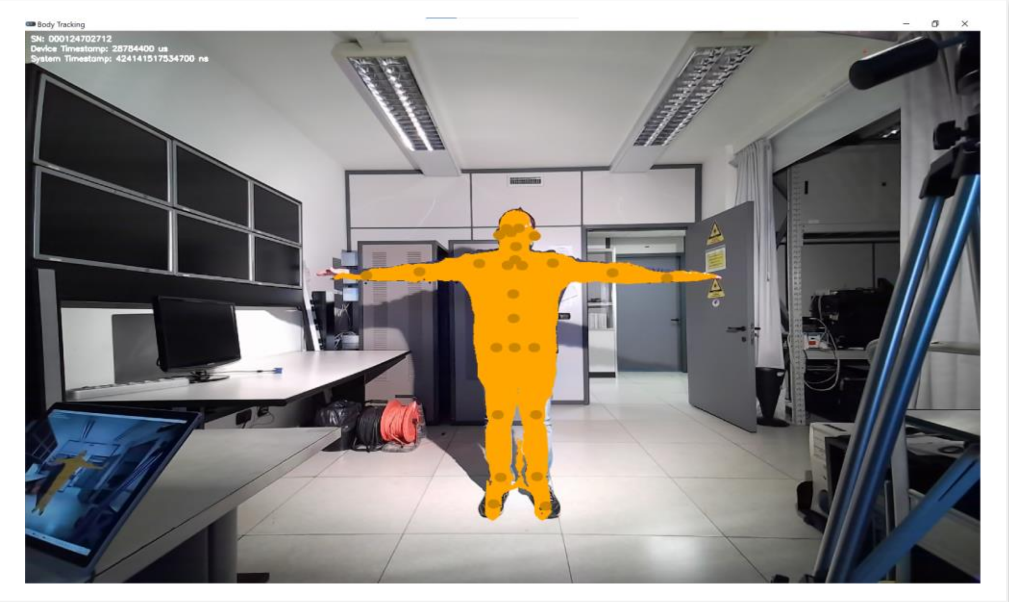
laura.romeo@stiima.cnr.it, laura.romeo@poliba.it

System Overview

Azure Kinect Body Tracking SDK

Tracking of bodies in 3D when used with the Azure Kinect DK hardware.

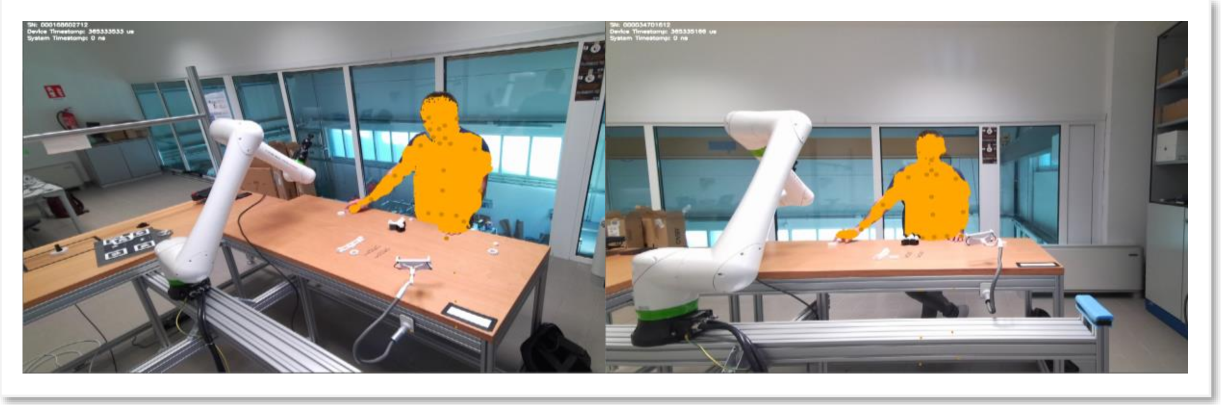
- Body segmentation
- Recognition of 32 skeletal joints for each body
- Extraction of anatomically correct skeleton for each partial or full body in FOV
- Unique identity for each body



laura.romeo@stiima.cnr.it, laura.romeo@poliba.it

System Overview

1. Azure Kinect analysis and performance evaluation
2. Multi vision system calibration
3. Image and Skeleton acquisition for action recognition

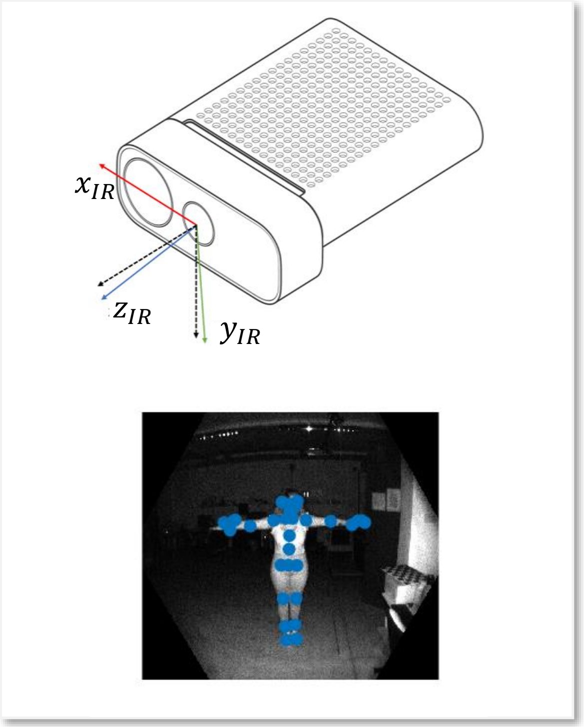


laura.romeo@stiima.cnr.it, laura.romeo@poliba.it

Azure Kinect: Specifics and Analysis

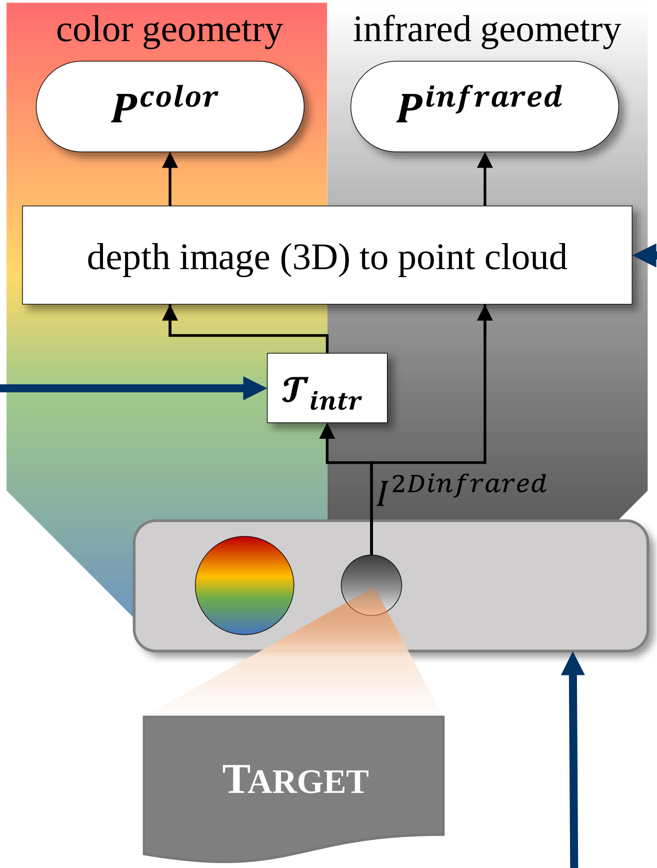
Skeletal joints coordinates **extracted** using the IR sensor.

- *k4a_calibration_3d_to_2d()*: converts the 3D coordinates in 2D pixel coordinates
- *k4a_transformation_depth_image_to_color_camera()*: converts the 3D coordinates from depth resolution to color resolution
- *k4a_calibration_2d_to_2d()*: converts the 2D pixel coordinates from depth resolution to color resolution, and vice-versa



laura.romeo@stiima.cnr.it, laura.romeo@poliba.it

Point Cloud realization



depth_image_to_point_cloud()

depth_image_to_color_camera()

Microsoft Azure Kinect

laura.romeo@stiima.cnr.it, laura.romeo@poliba.it

Azure Kinect: Specifics and Analysis

Point Cloud realization

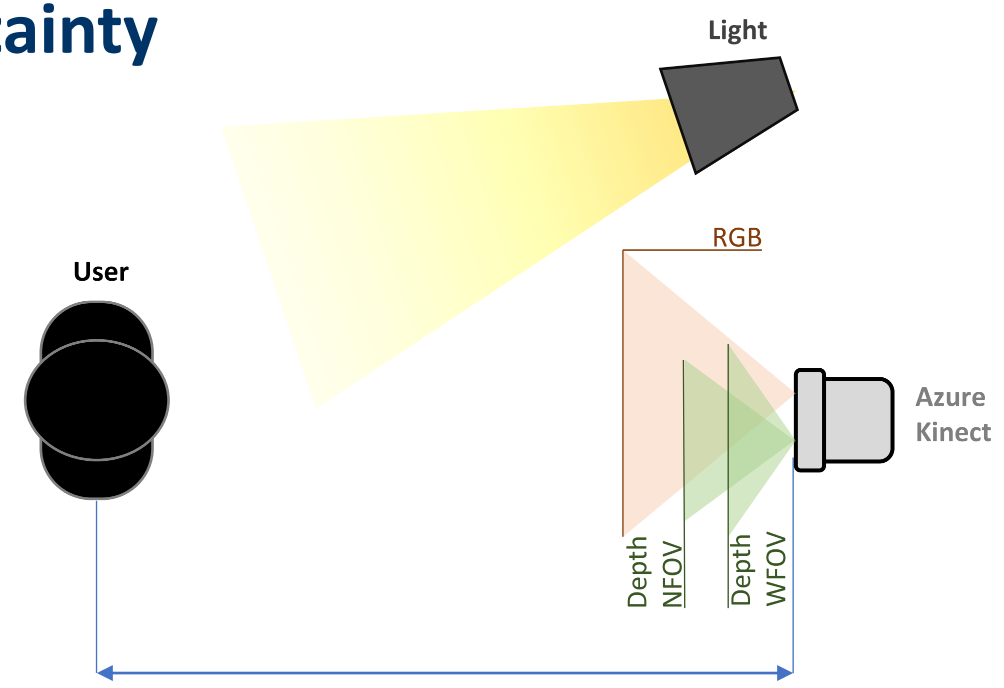
- Elaboration from the IR sensor
- Set of 3D coordinate points in a specific coordinate system
- Color, depth, intensity information for each point

laura.romeo@stiima.cnr.it, laura.romeo@poliba.it

Azure Kinect: Specifics and Analysis

STIIMA | Politecnico di Bari

Skeletal joints uncertainty



Standard Deviation computed by changing:

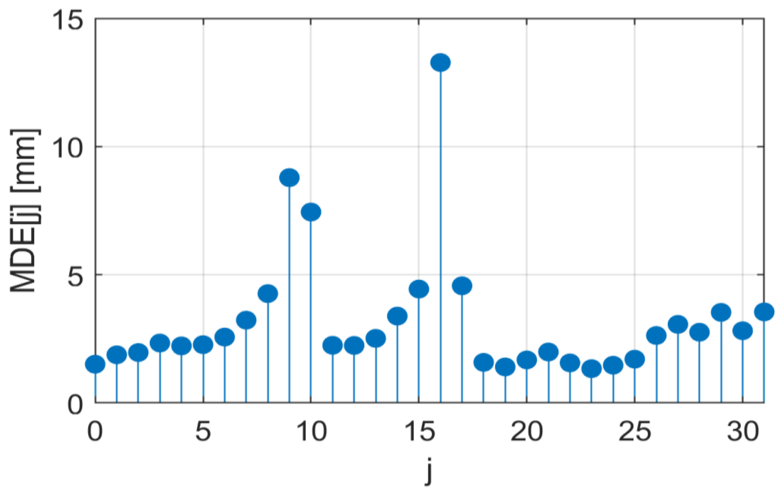
- *Infrared Resolution*
- *Ambient light conditions*
- *Body occlusions*
- *Subject-camera distance*

laura.romeo@stiima.cnr.it, *laura.romeo@poliba.it*

Azure Kinect: Specifics and Analysis

Example of MDE (Mean Distance Error) of the 32 joints computed by the Azure Kinect and its body-tracking library with the following configuration:

- Narrow depth resolution
- No Occlusions
- Lights at 1750 Lux (direct illumination)
- User at 1-meter distance



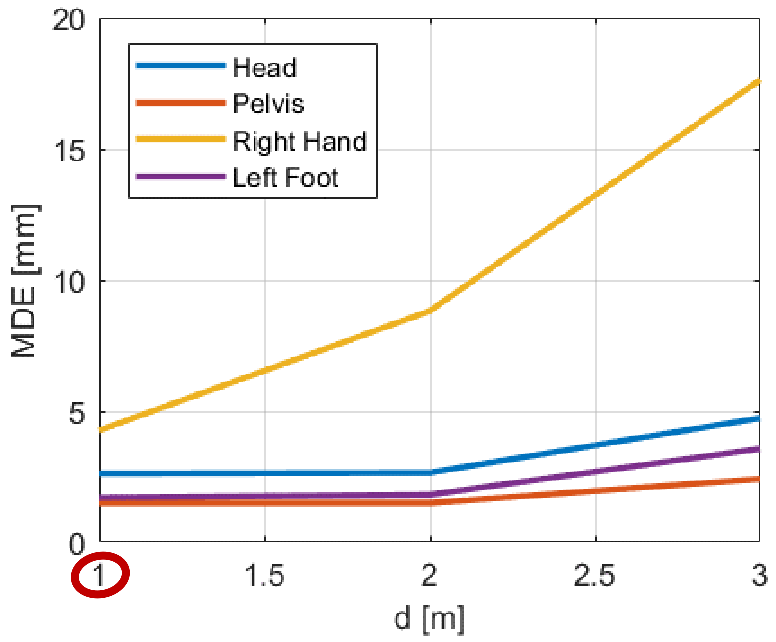
Four representative Joints:

- Head joint (j = 26)
- Pelvis joint (j = 0)
- Left-hand joint (j = 8)
- Right-foot joint (j = 25)

laura.romeo@stiima.cnr.it, laura.romeo@poliba.it

Azure Kinect: Specifics and Analysis

Best configuration



- Narrow Resolution
- No Occlusions
- 1750 lux
- 1 meter distance

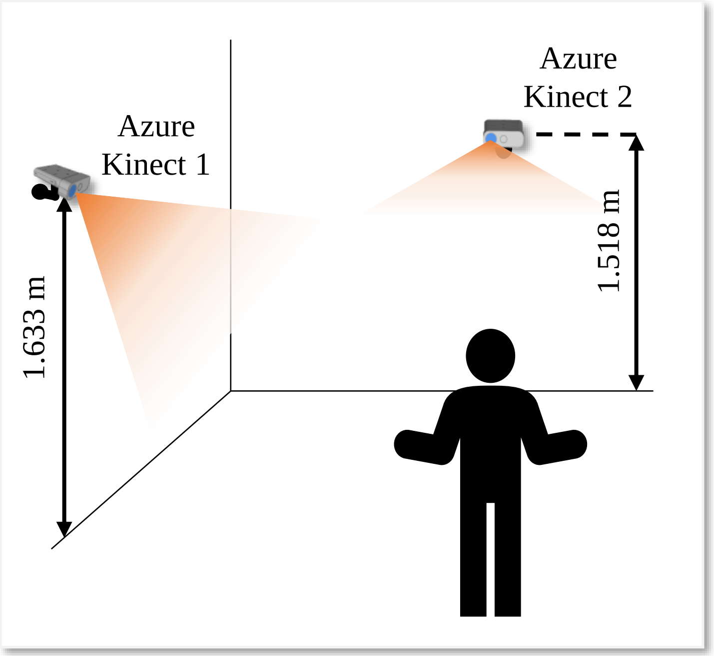
- MDEs increase as the user-camera distance grows
- Wide Resolution increases the MDE, which is about tripled of the case with Narrow Resolution
- Direct illumination of 1750 lux generally downs the performance of the body tracking
- The joints show higher values of uncertainty when the skeleton is partially occluded
- Left-hand joint shows the worst MDEs under all testing conditions

1 *mm* < MDE < 53 *mm*
MEAN = 8 *mm*
STD = 6 *mm*

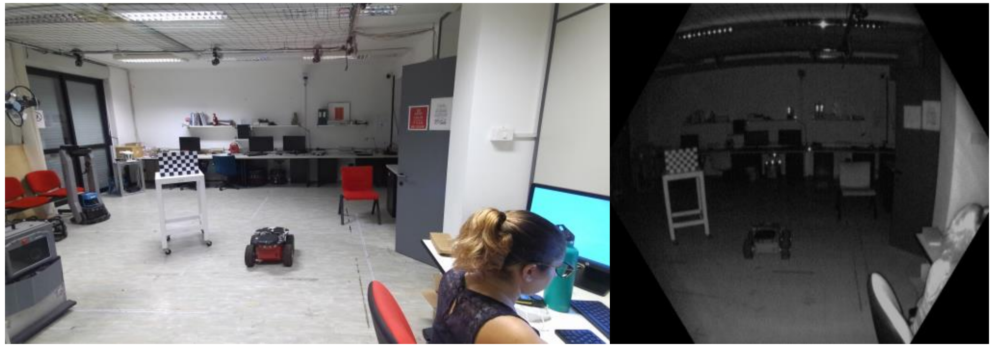
Romeo, L., Marani, R., Malosio, M., Perri, A. G., & D'Orazio, T. (2021, June). "Performance analysis of body tracking with the Microsoft Azure Kinect". *In 2021 29th Mediterranean Conference on Control and Automation* (MED) (pp. 572-577). IEEE.

laura.romeo@stiima.cnr.it, laura.romeo@poliba.it

Setup assessment: multi-vision system



- Vision system of two cameras
- Calibration assessed depending on the use-case:
  - ✓ Skeletal joints
  - ✓ Point Cloud with color geometry
  - ✓ Point Cloud with infrared geometry



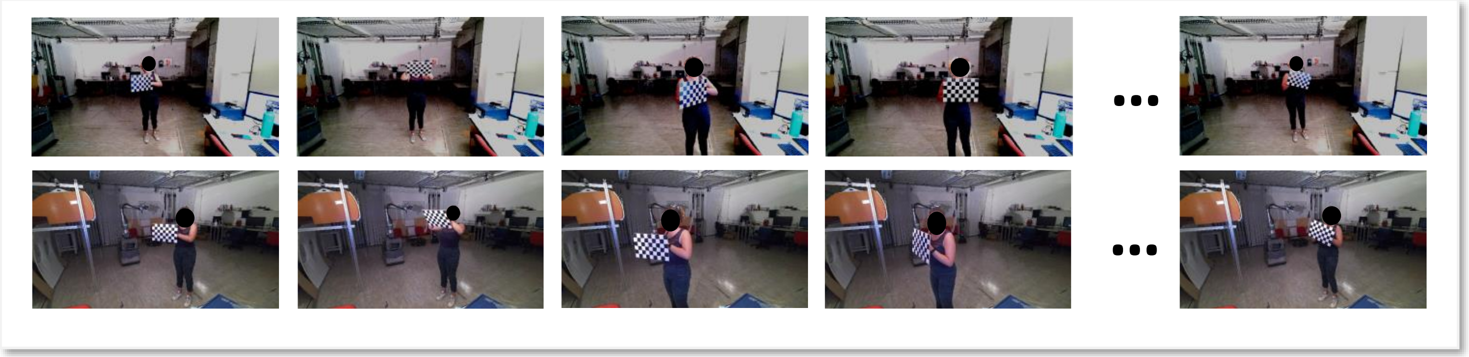
laura.romeo@stiima.cnr.it, laura.romeo@poliba.it

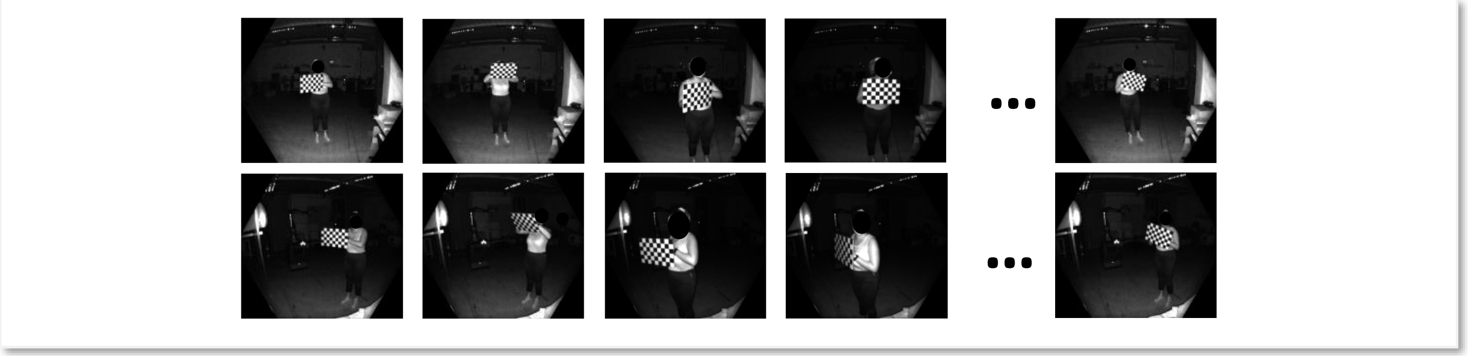
Setup assessment: multi-vision system

Camera Calibration

- Interface between 3D world and 2D camera images
- Minimization of metric distance inaccuracy
- Methodologies based on a feature target

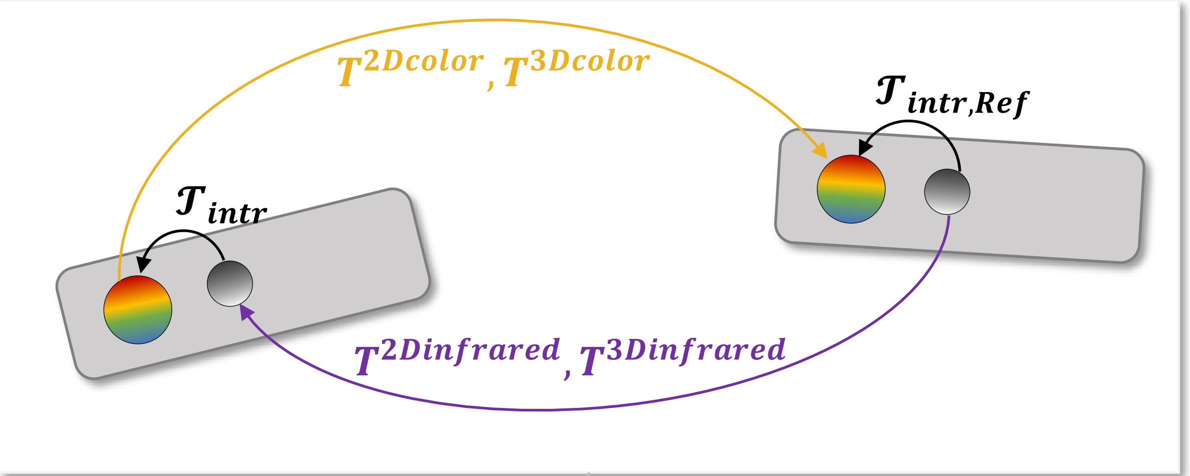
- Chessboard acquisition:
  - ✓ black and white squares
  - ✓ frames with different positions and orientations
  - ✓ RGB and IR images





laura.romeo@stiima.cnr.it, laura.romeo@poliba.it

Setup assessment: multi-vision system



- **2D calibration** (RGB and IR separately):
  - ✓ Chessboard frames from both cameras
  - ✓ Matching chessboard square corners among frames
  - ✓ Trasformation matrix as output
- **3D calibration** (RGB and IR separately):
  - ✓ Chessboard frames from both cameras
  - ✓ SDK function for 3D projection
  - ✓ 3D geometric transformation estimation
  - ✓ Transformation matrix as output

Romeo, L., Marani, R., Perri, A.G., & D'Orazio, T. "Multiple Azure Kinect Calibration for three-dimensional dense Point Clouds and reliable Skeletons". *IEEE Transaction on Instrumentation and Measurement*. *(to be submitted)*

laura.romeo@stiima.cnr.it, laura.romeo@poliba.it

Setup assessment: multi-vision system

Point Cloud Calibration (RGB)

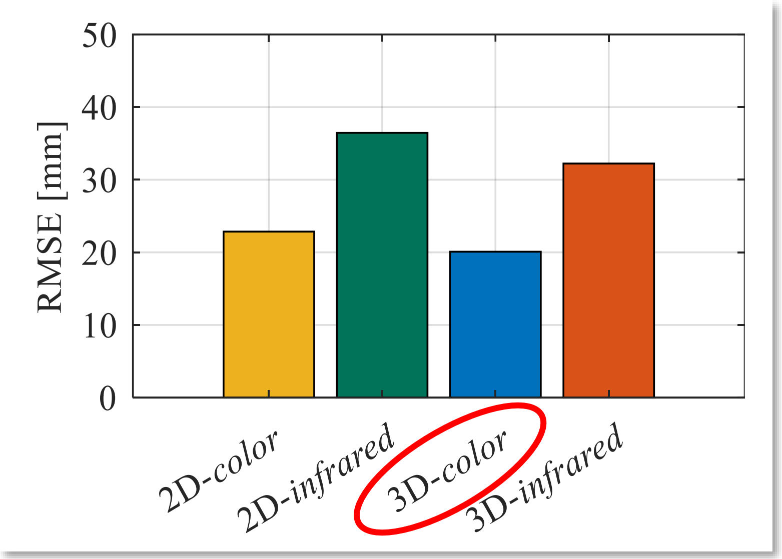
- Best result: 3D calibration with RGB images
- Worst result: 2D calibration with IR images

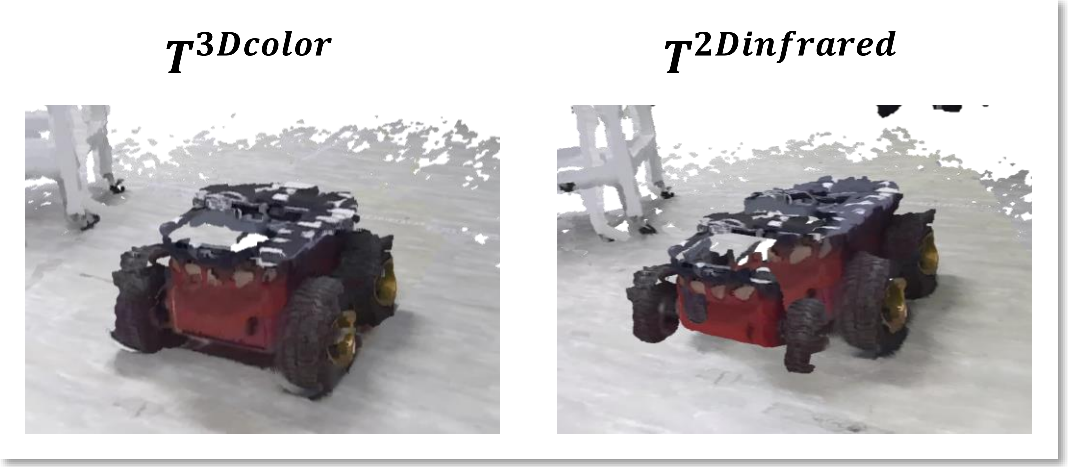
Pros:

- Large set of coordinates
- High point cloud density

Cons:

- Low coordinate precision
- Massive data storage



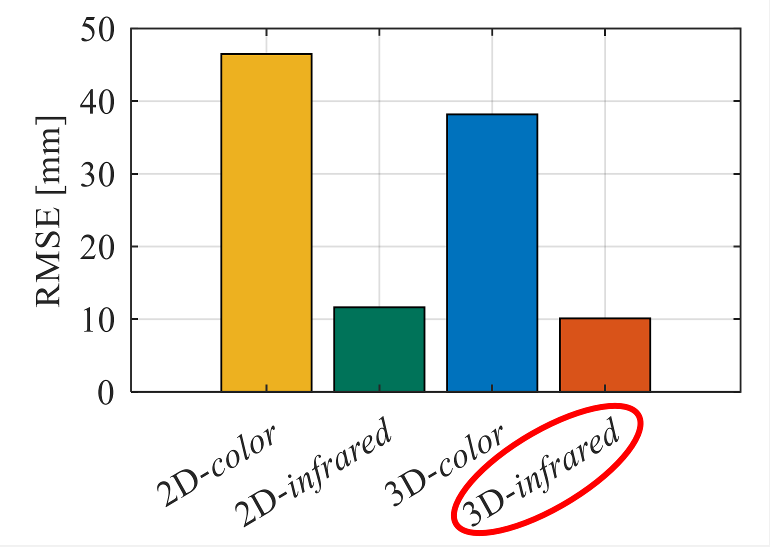


laura.romeo@stiima.cnr.it, laura.romeo@poliba.it

Setup assessment: multi-vision system

Point Cloud Calibration (IR)

- ➢ Best result: 3D calibration with IR images
- ➢ Worst result: 2D calibration with RGB images



Pros:

- High coordinate precision
- Low data storage

Cons:

- Small set of coordinates
- Low point cloud density

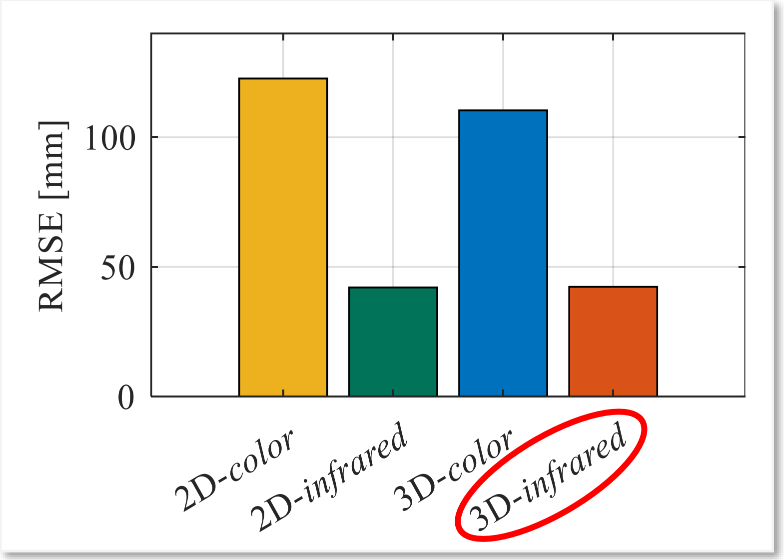
$T^{3Dinfrared}$ $T^{2Dcolor}$

laura.romeo@stiima.cnr.it, laura.romeo@poliba.it

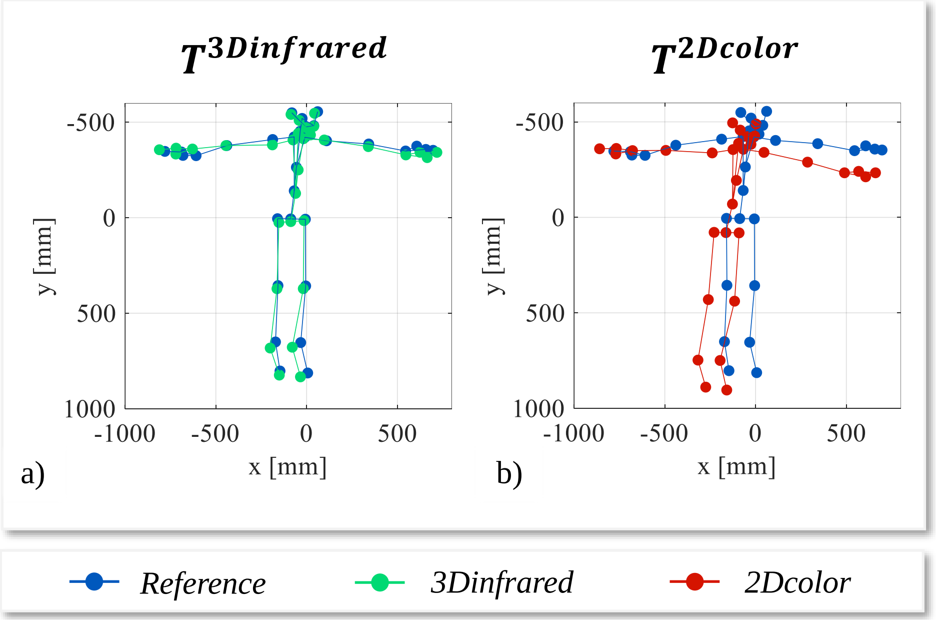
Setup assessment: multi-vision system

Skeleton Calibration (IR)

- Best result: 3D calibration with IR images
- Worst result: 2D calibration with RGB images



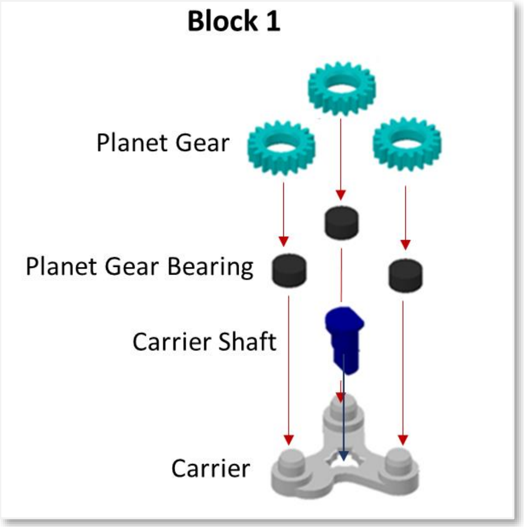
- Results highly comparable with IR point cloud calibration
- Uncertainty due to the pseudo-random error from Azure Kinect Body Tracking SDK

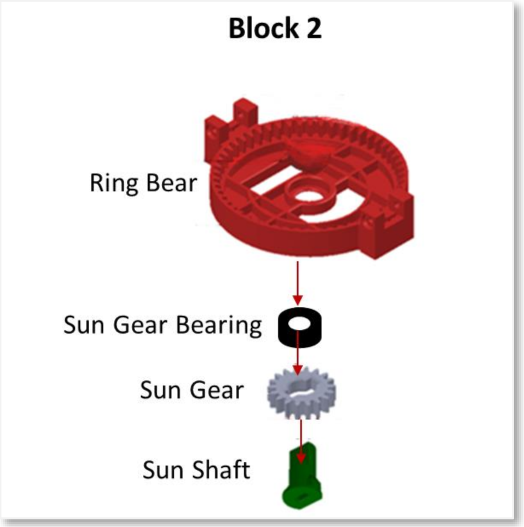


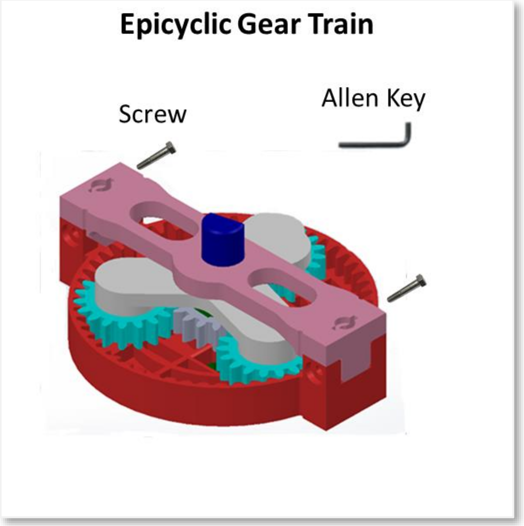
laura.romeo@stiima.cnr.it, laura.romeo@poliba.it

Epicyclic Gear Train (EGT)

- Visual recognition of human performing a specific task
- Epicyclic Gear Train to be assembled by the operator
- Action recognition information sent to the collaborative robot

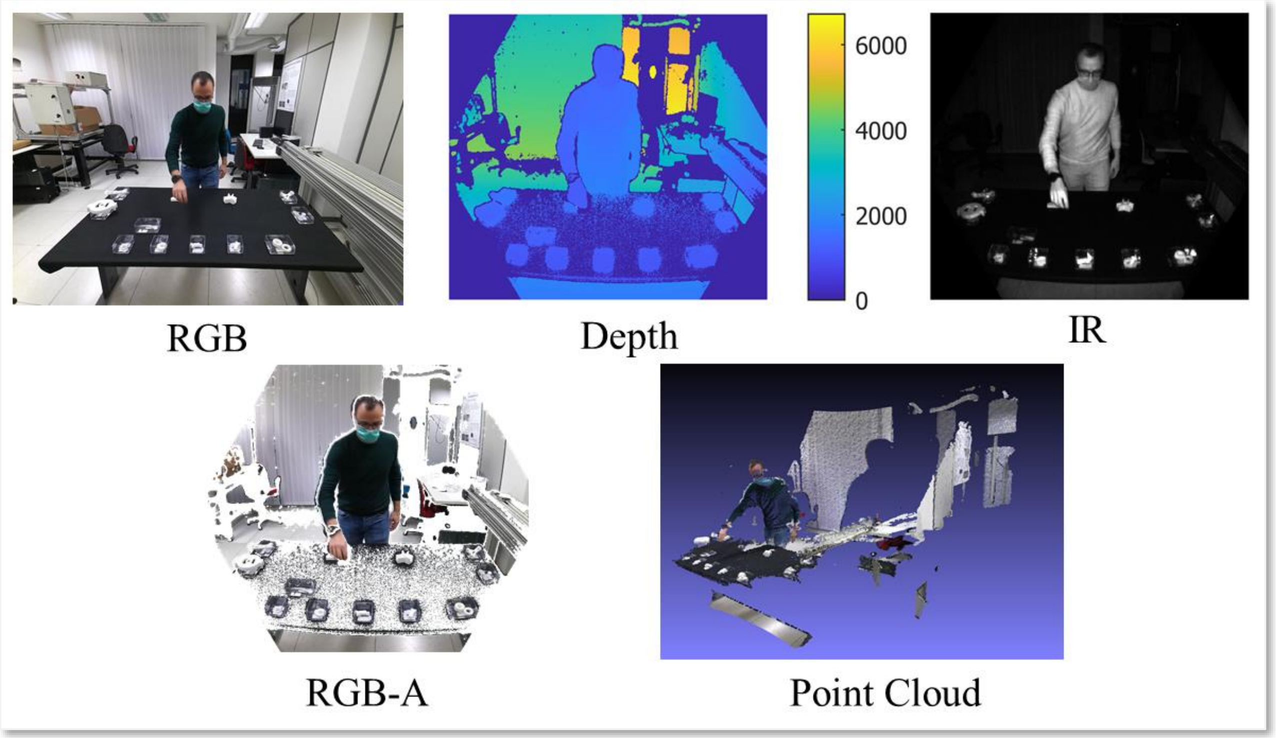






laura.romeo@stiima.cnr.it, laura.romeo@poliba.it

Setup assessment: action recognition



- Creation of a dataset for action recognition
- Recording of the assembling process using an Azure Kinect
- Dataset composed of:
  - ✓ RGB images
  - ✓ IR images
  - ✓ Depth images
  - ✓ Point Cloud with IR resolution
  - ✓ RGB images in IR environment
  - ✓ Skeleton information

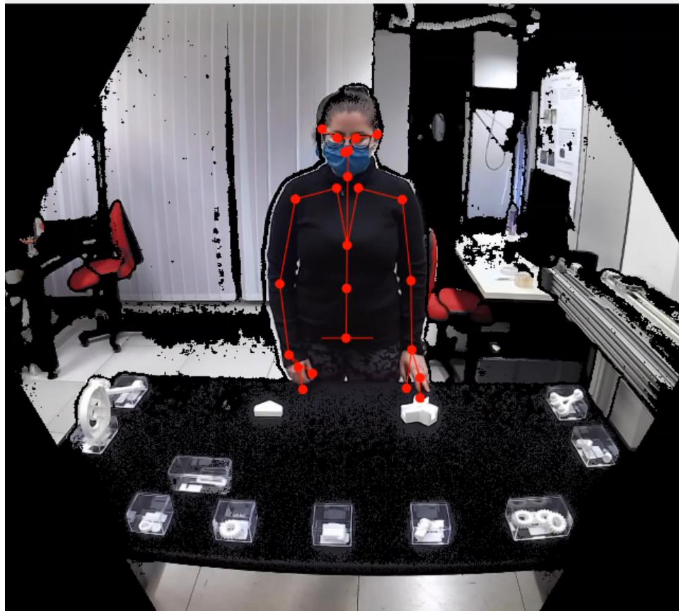
Romeo, L., Marani, R., D'Orazio, T., & Cicirelli, G. "RGB-D Dataset for Human Action Recognition in Assembly Task". *Data in Brief*. *(to be submitted)*

laura.romeo@stiima.cnr.it, laura.romeo@poliba.it

Setup assessment: action recognition

- Image Segmentation
- Analysis of the dataset
- Extraction of relevant information
- Deep learning implementation (CNN, GCNN, LSTM, …)

Frame n. 18. Action 0 (0)



	Quantity	Components	Class Action
Block 1	1	Carrier	**1**
	3	Planet Gear Bearing	**2**
	3	Planet Gear	**3**
	1	Carrier Shaft	**4**
Block 2	1	Sun Shaft	**5**
	1	Sun Gear	**6**
	1	Sun Gear Bearing	**7**
	1	Ring Bear	**8**
Block 3	1	Block 2 on Block 1	**9**
	1	Ring Shaft	**10**
	2	Screws	**11**
	1	Allen Key	**12**

laura.romeo@stiima.cnr.it, *laura.romeo@poliba.it*

Conclusion and Future Works

- Implementation of a monitoring system in a real-case scenario
- Analysis of video devices and their uncertainties
- Calibration methodologies for the estimation of 3D coordinates in a multi-vision system
- Action Segmentation and Recognition on a real Dataset

Future Works:

- Development of the most suitable neural network for the segmented Dataset, for Action Recognition
- Extraction of the most relevant feature to better exploit the created dataset
- Implementation of the Action Recognition system also using a second camera, in a calibrated setup

laura.romeo@stiima.cnr.it, laura.romeo@poliba.it

Thanks for the attention!

PhD Student: Laura Romeo (laura.romeo@stiima.cnr.it, laura.romeo@poliba.it)

Tutors:

Prof. Anna Gina Perri

Ing. Roberto Marani