ETH zürich

pd|z Product Development Group Zurich
Produktentwicklungsgruppe Zürich

Surgeon Hand Pose Tracking for a Digital Surgery Platform

Nilesh Balu

Final Presentation

17th June 2025

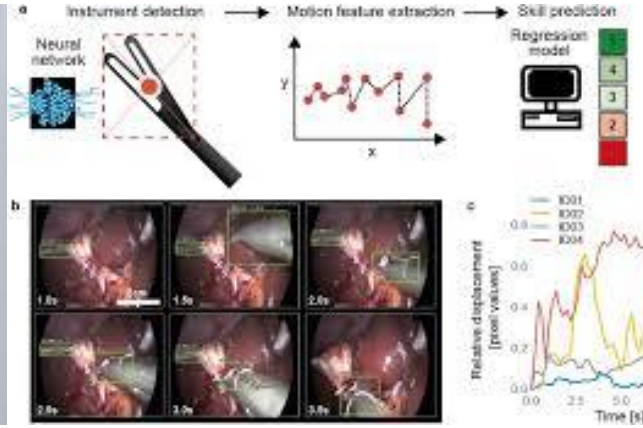
Motivation

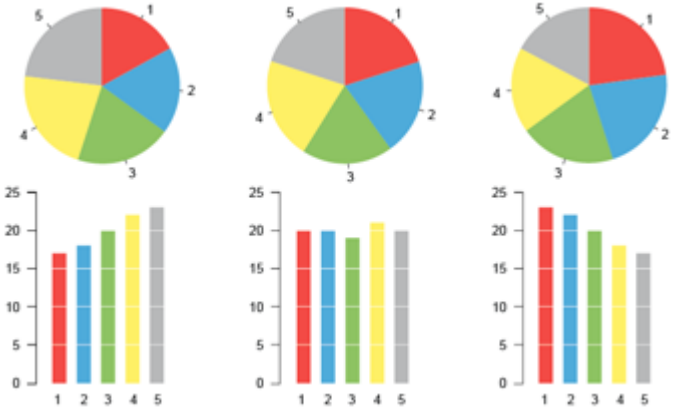
Digital Surgery Platform

Surgical Training

Objective Performance Evaluation

Automated Feedback





Digital Twin

The Current System

Tracks Tool Pose and Anatomical Features of the Patient in 3D

This Thesis: Track Hand and Body Pose of the Surgeon in 3D

Goals

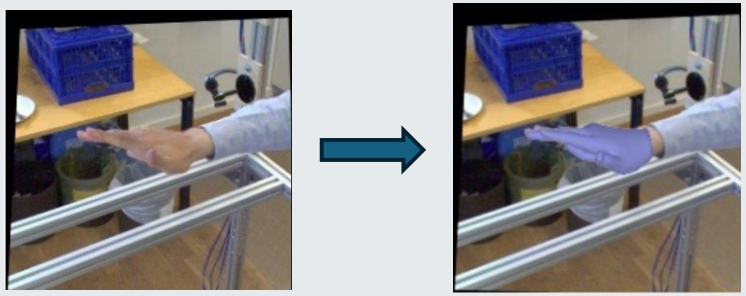
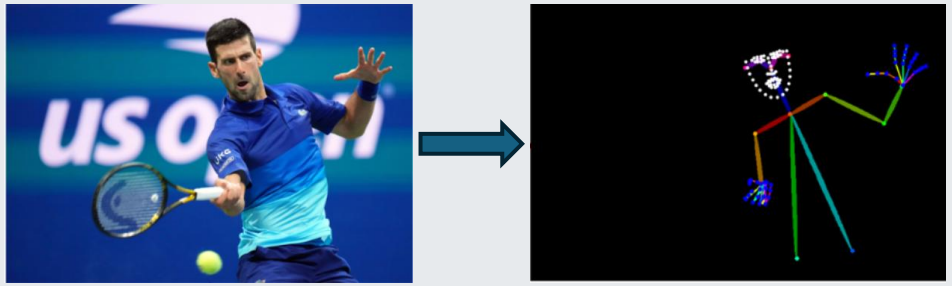
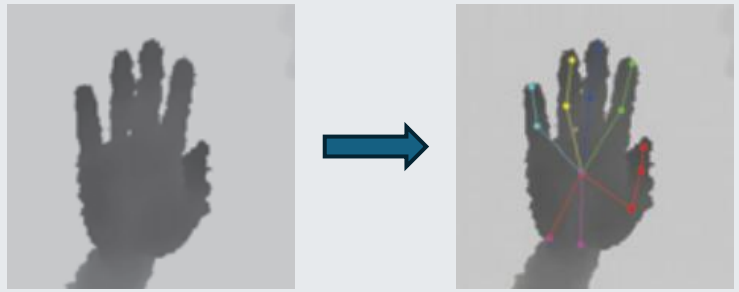
Estimate Hand or Body Pose in 3D up to a 10 mm accuracy

Robust to Occlusions caused by Surgical Tools

Portable and Real Time Capable

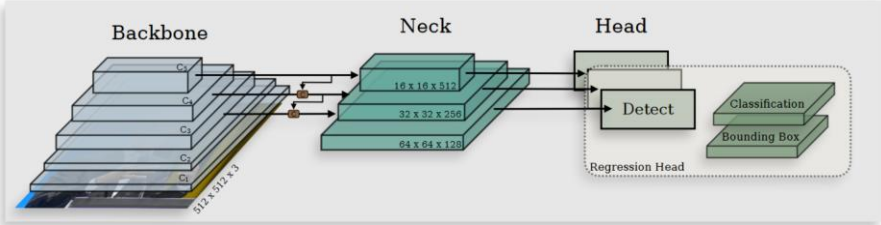
Validate Results

Pose Estimation Models

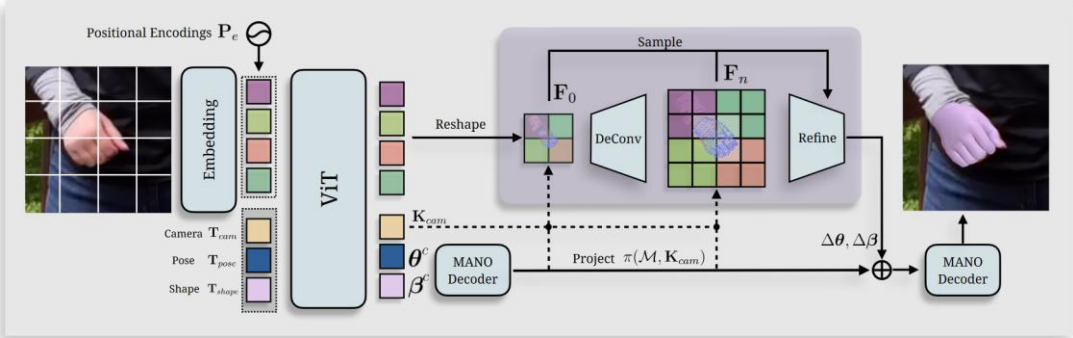
	Hand Pose	Whole Body Pose
Single RGB Image	**WiLoR** 	**DWPose** 
Single Depth Image	**AWR** 	

WiLoR

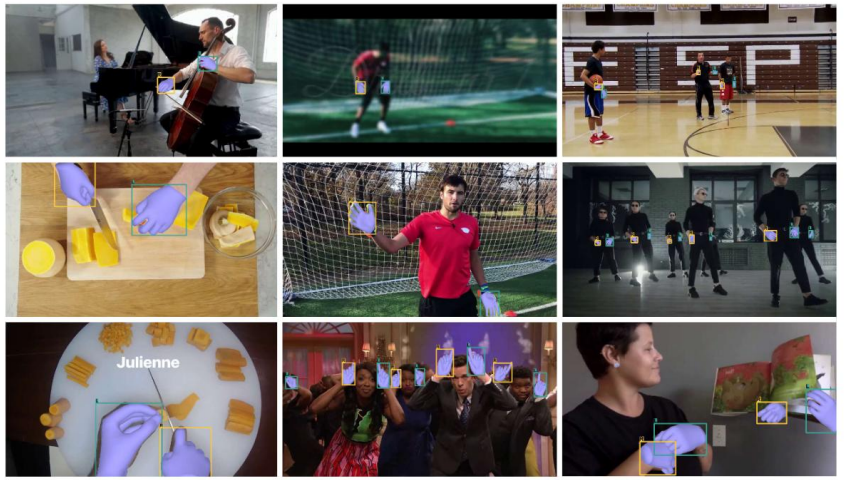
1. Detection: Low-latency hand detector



2. Reconstruction: Fits the MANO Hand model on the image

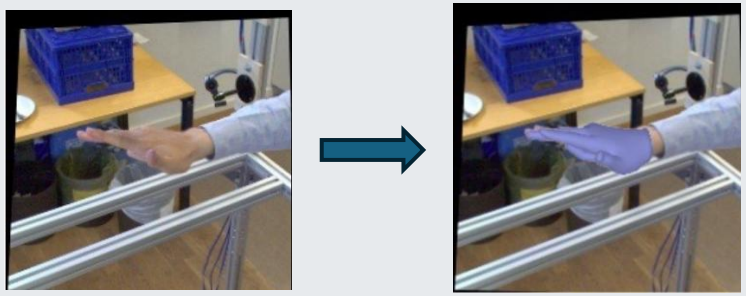
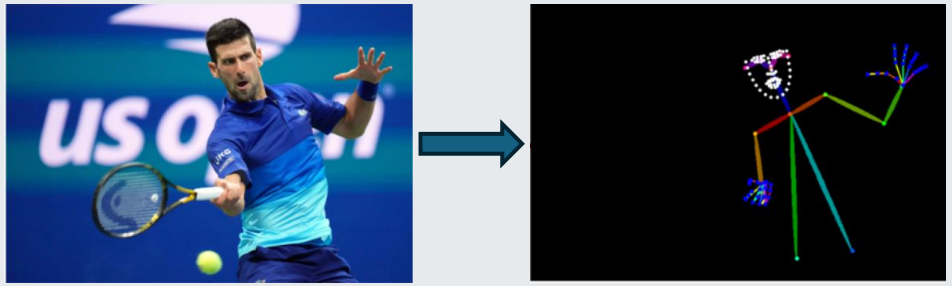
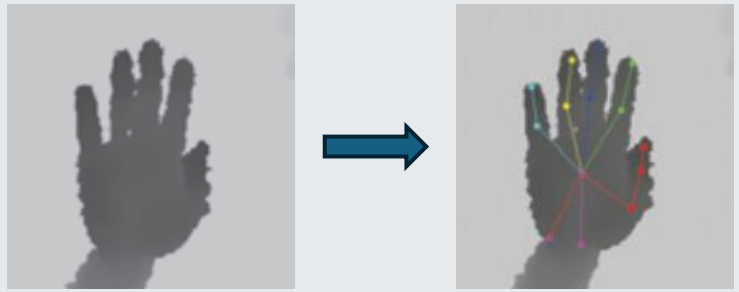


3. Training: Trained on the WHIM dataset



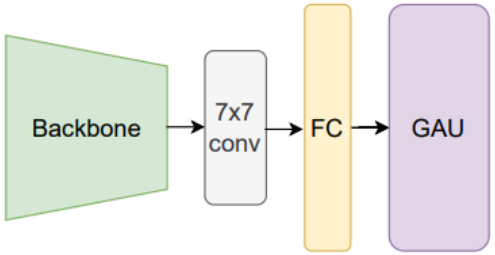
- 3rd person and ego-centric views
- Large number of occlusions present
- > 1 million images

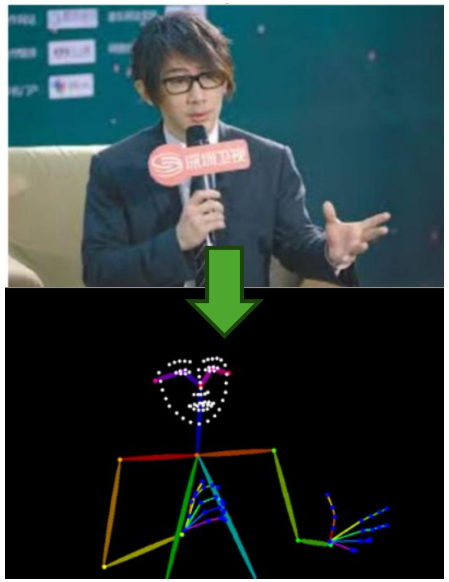
Pose Estimation Models

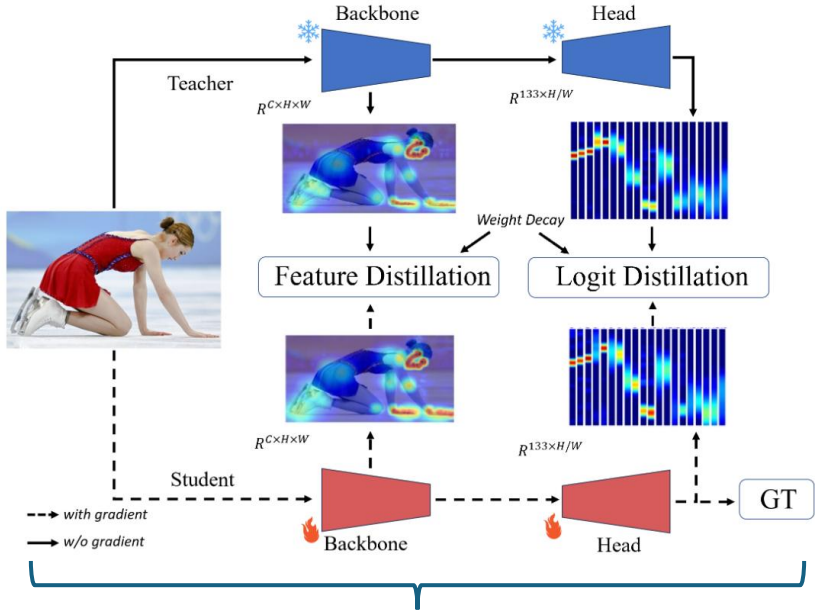
	Hand Pose	Whole Body Pose
Single RGB Image	**WiLoR** 	**DWPose** 
Single Depth Image	**AWR** 	

DWPose

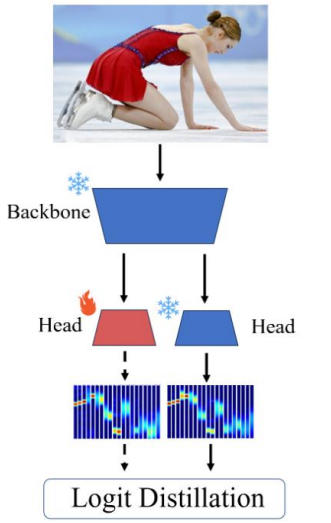
Teacher Network: RTMPose







Stage 1: Teacher-Student Distillation

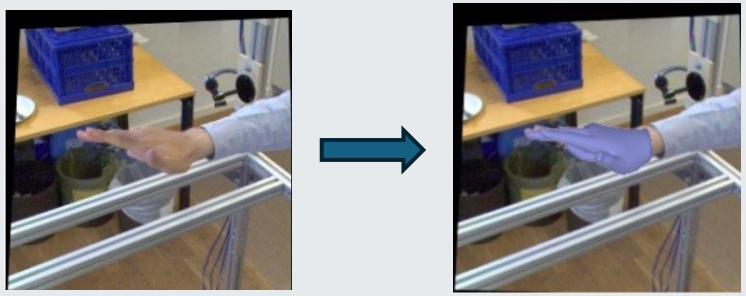
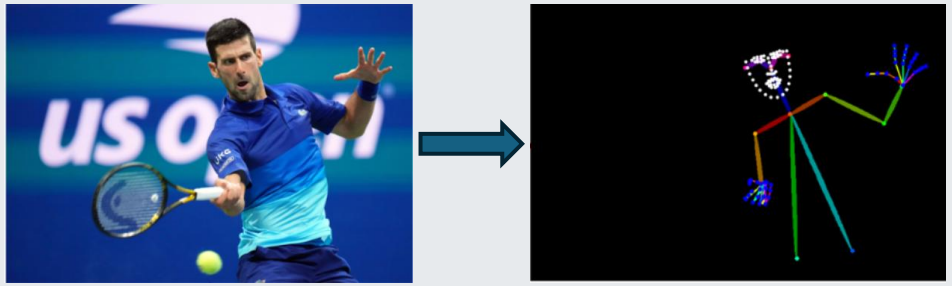
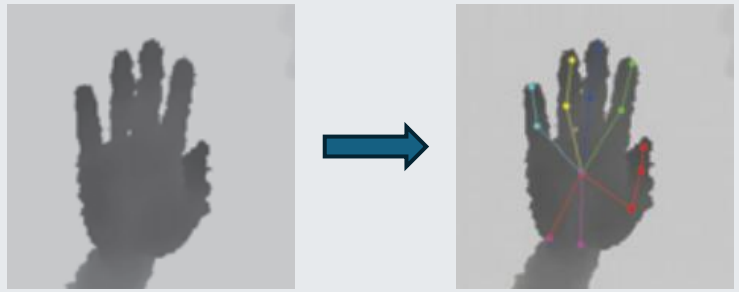


Stage 2: Head-Aware Self Distillation

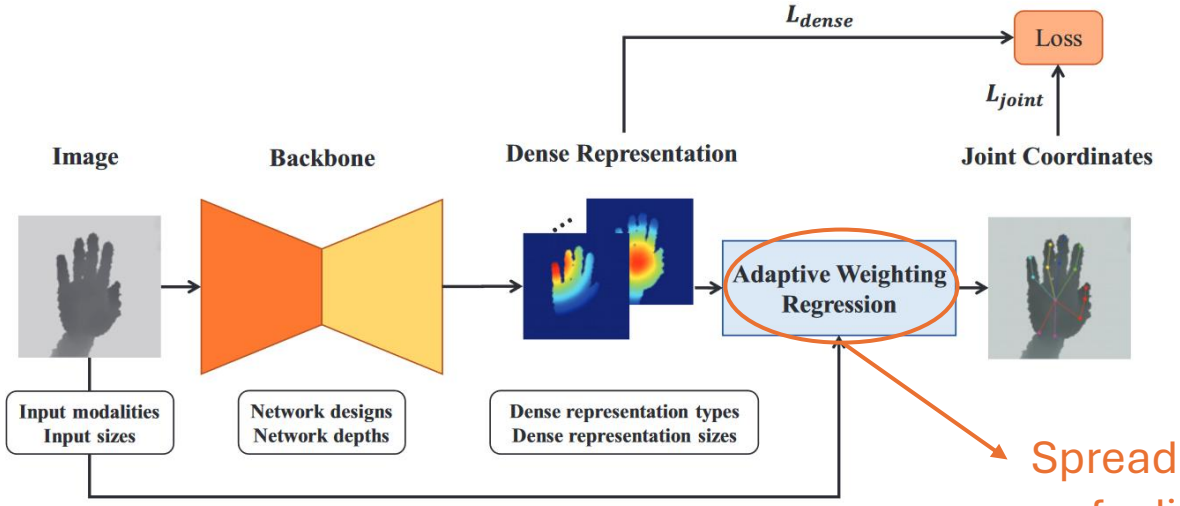
Dataset: COCO-WholeBody

3rd Person View; Occlusions

Pose Estimation Models

	Hand Pose	Whole Body Pose
Single RGB Image	**WiLoR** 	**DWPose** 
Single Depth Image	**AWR** 	

AWR



Spreads information of adjacent joints

Datasets:

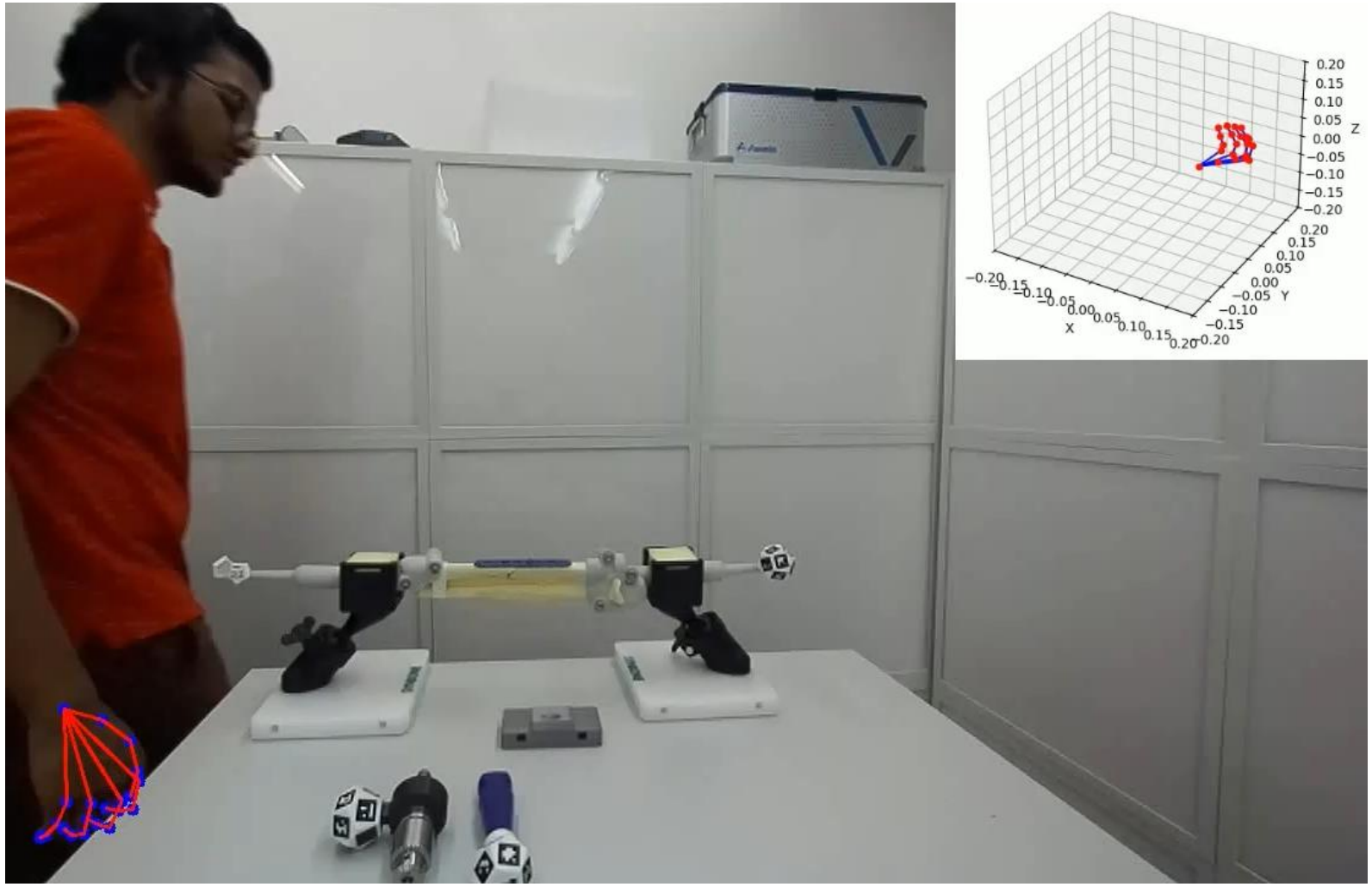
- NYU
- ICVL
- MSRA

3rd Person Views

Not Many Occlusions

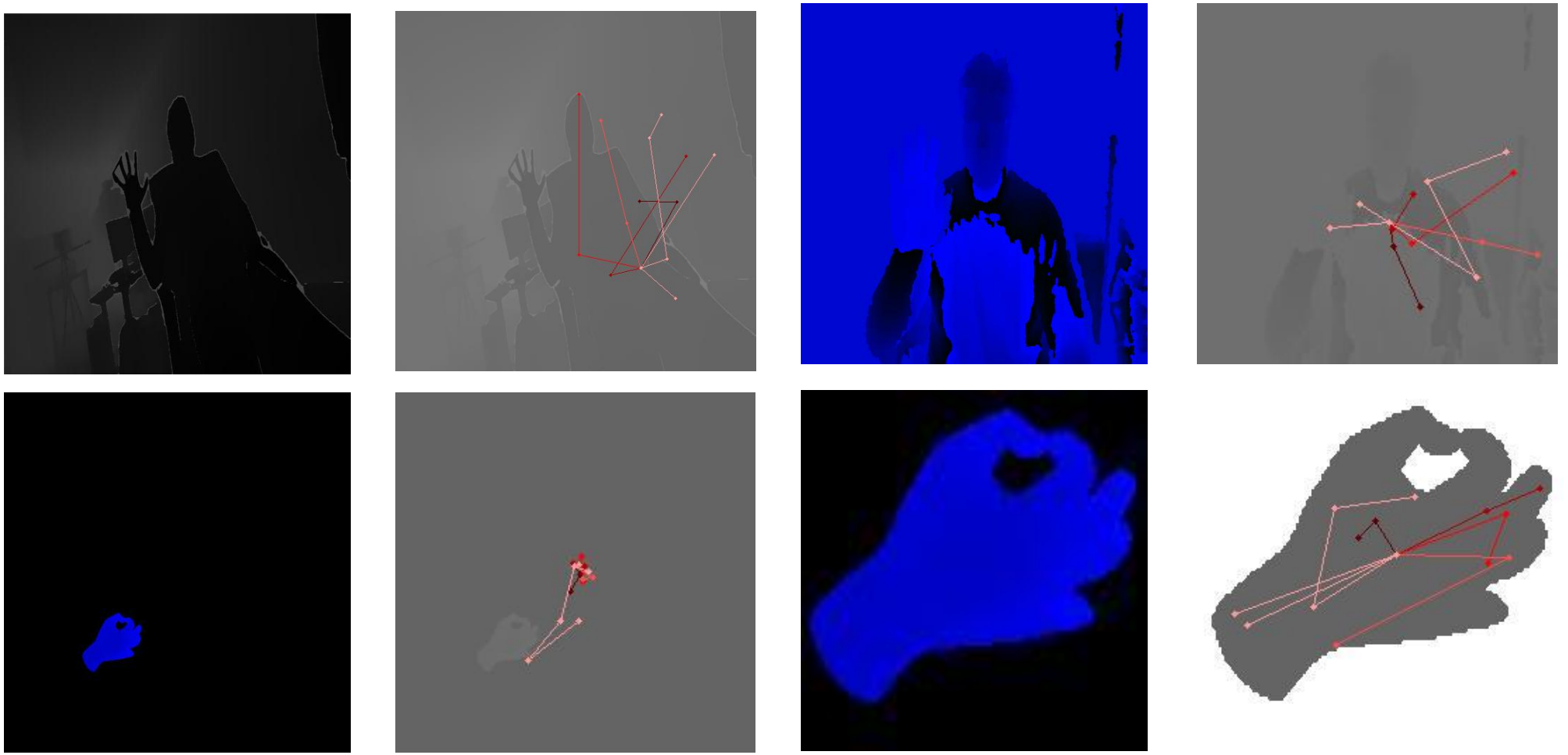
Robustness to Occlusions: Adaptive Weighting Regression

Results: WiLoR



Results: DWPose

Results: Issues with AWR



Goals

Estimate Hand or Body Pose in 3D up to a 10 mm accuracy

Only 2.5 D

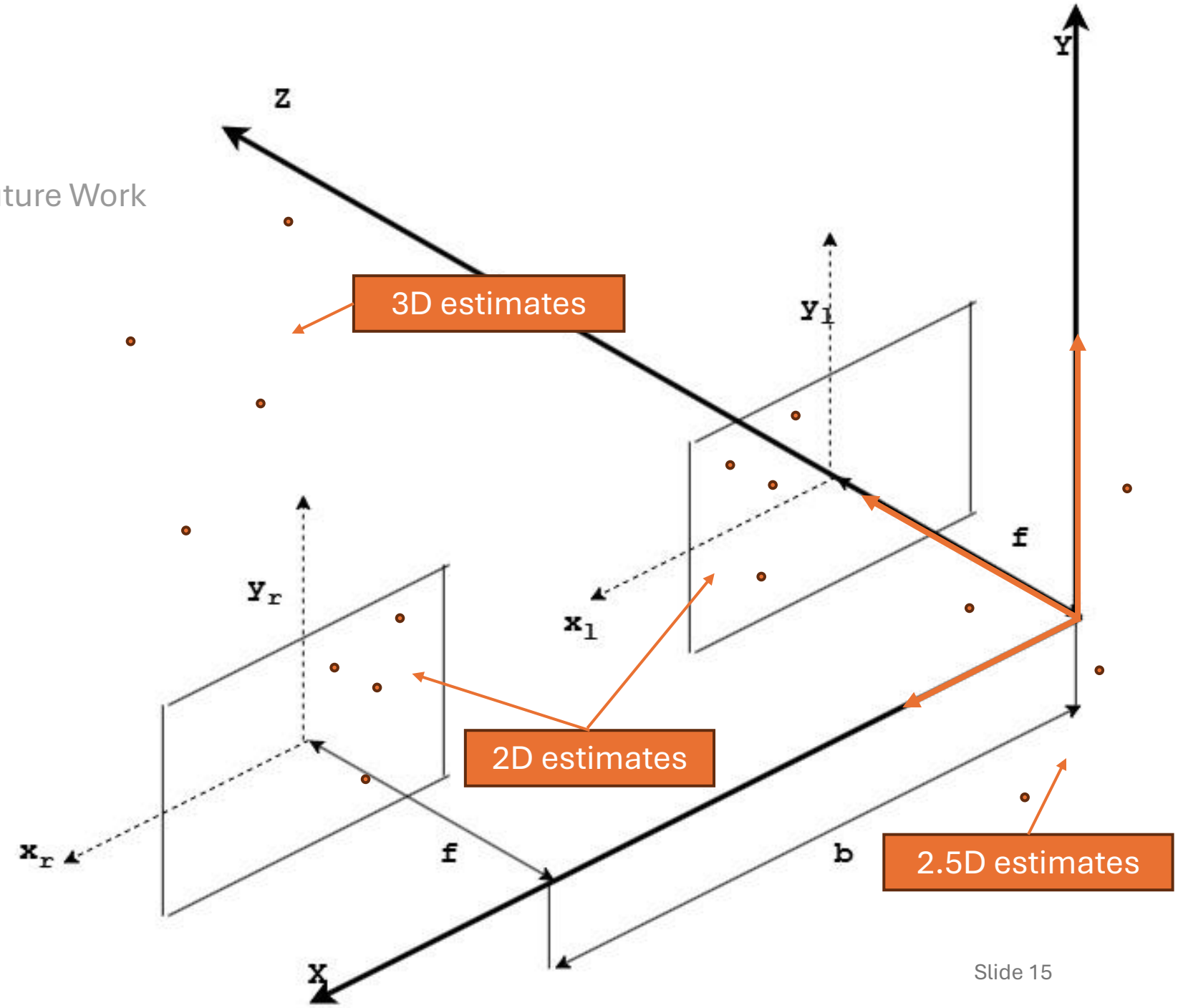
Robust to Occlusions caused by Surgical Tools

Predicts parts of the hand occluded by the surgical tool

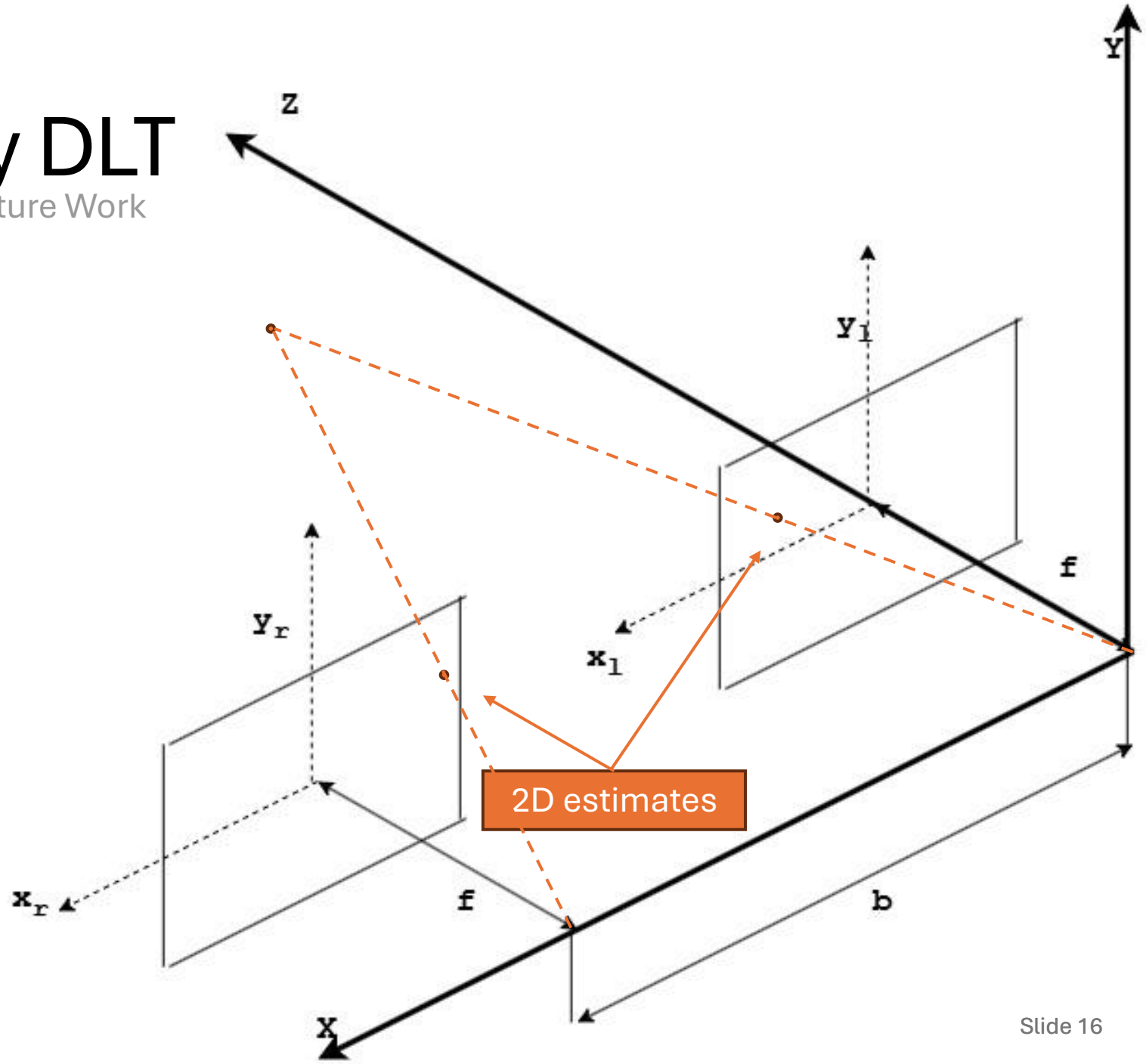
Portable and Real Time Capable

Validate Results

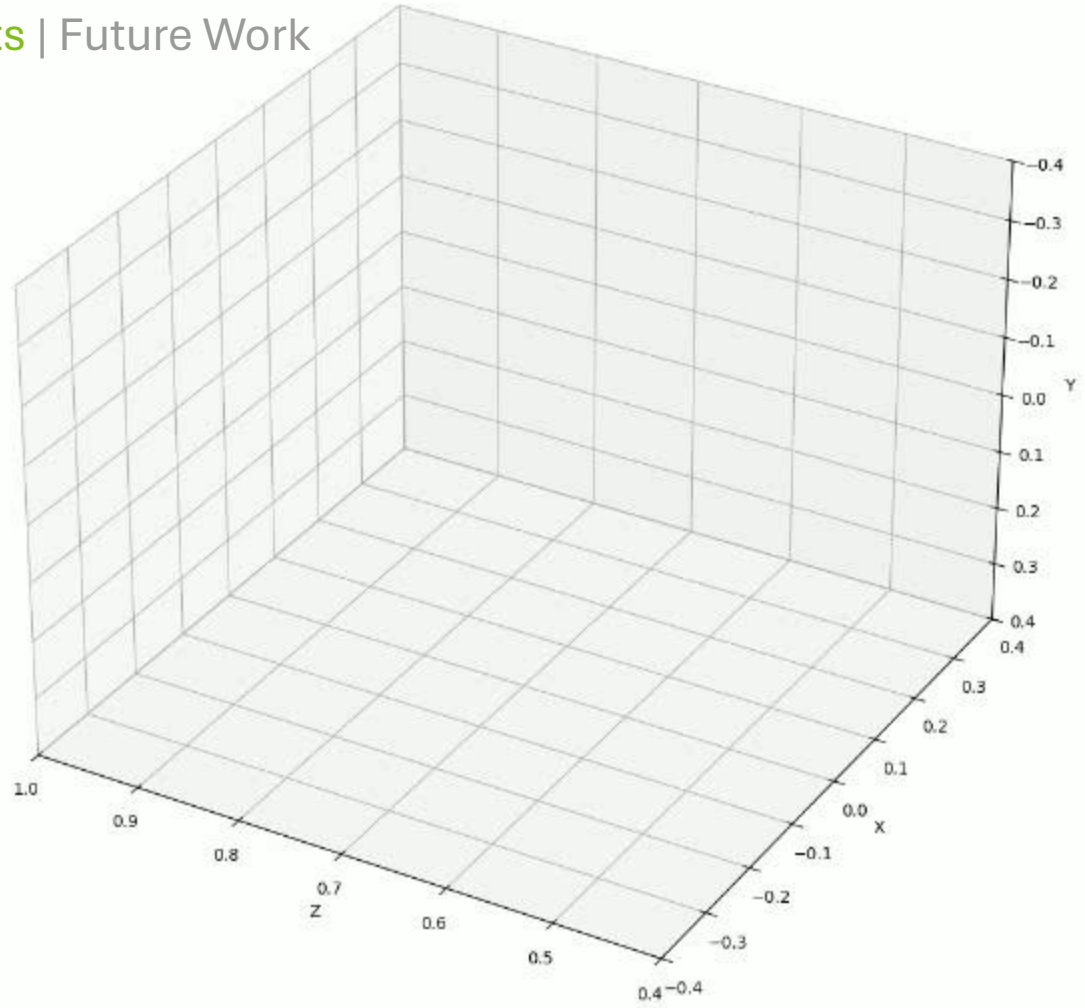
Triangulation



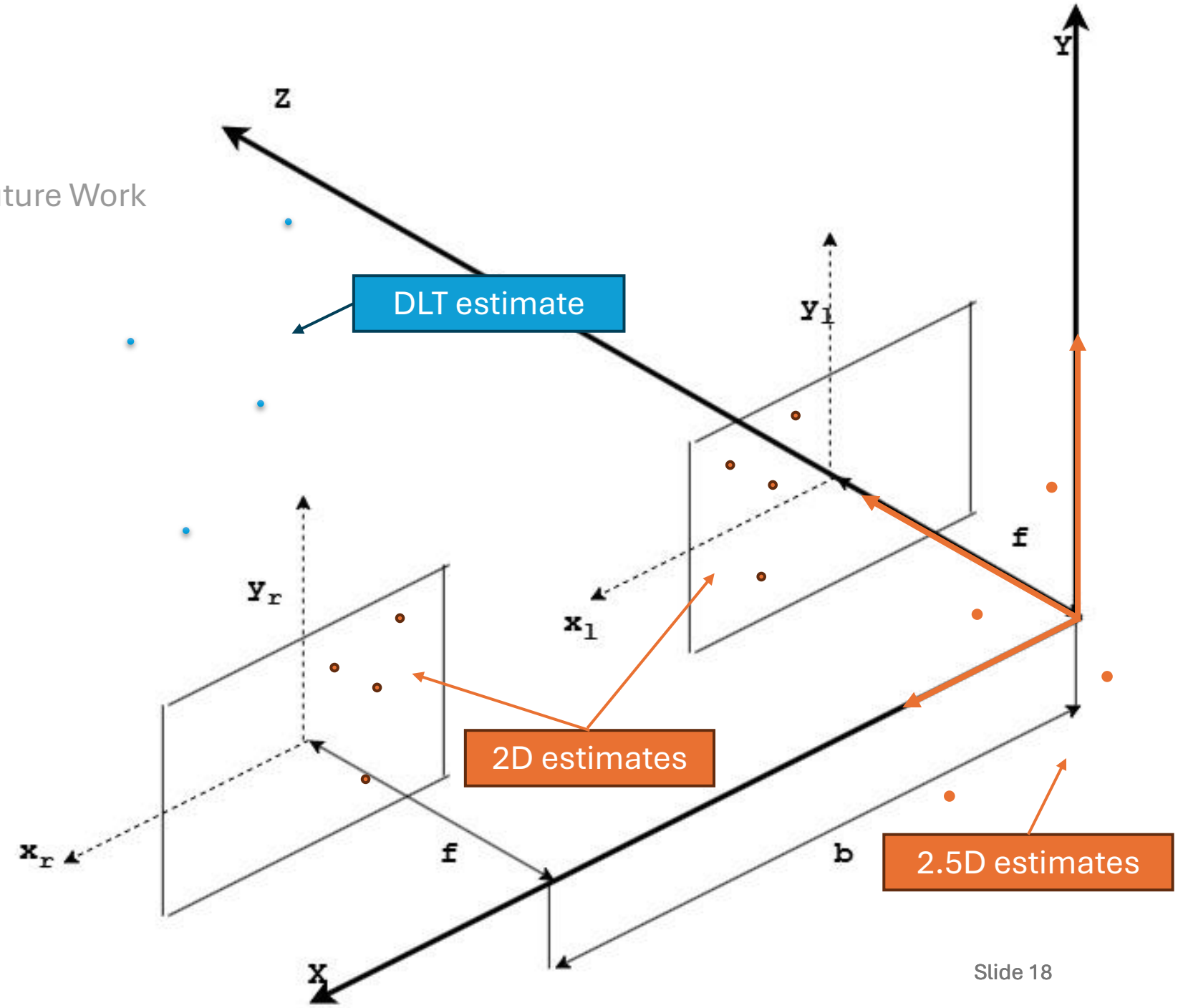
Triangulation by DLT



Results: DLT

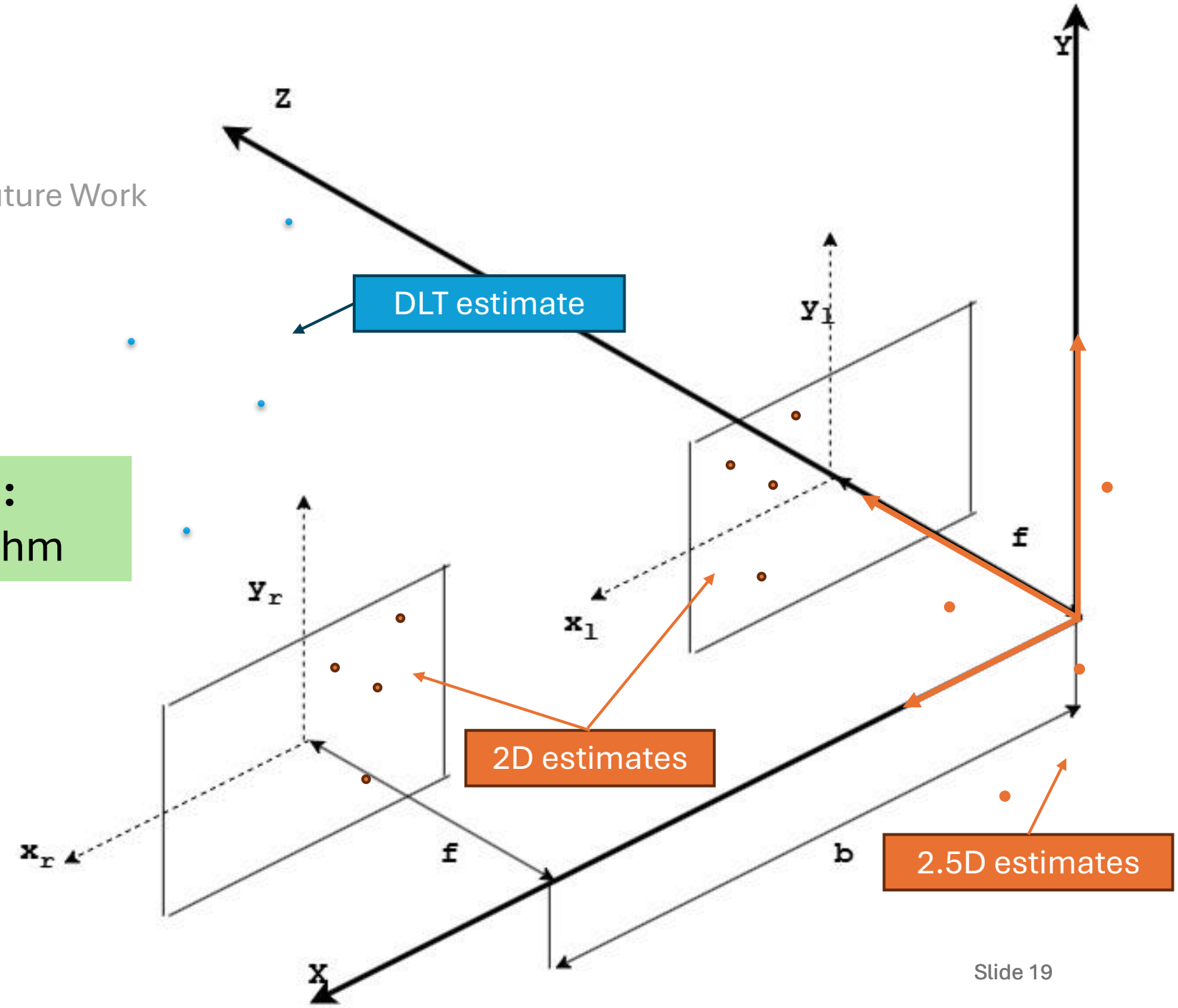

1.0
0.9
0.8
0.7
0.6
0.5
0.4
Z
−0.4
−0.3
−0.2
−0.1
0.0
0.1
0.2
0.3
0.4
X
−0.4
−0.3
−0.2
−0.1
0.0
0.1
0.2
0.3
0.4
Y

Optimization

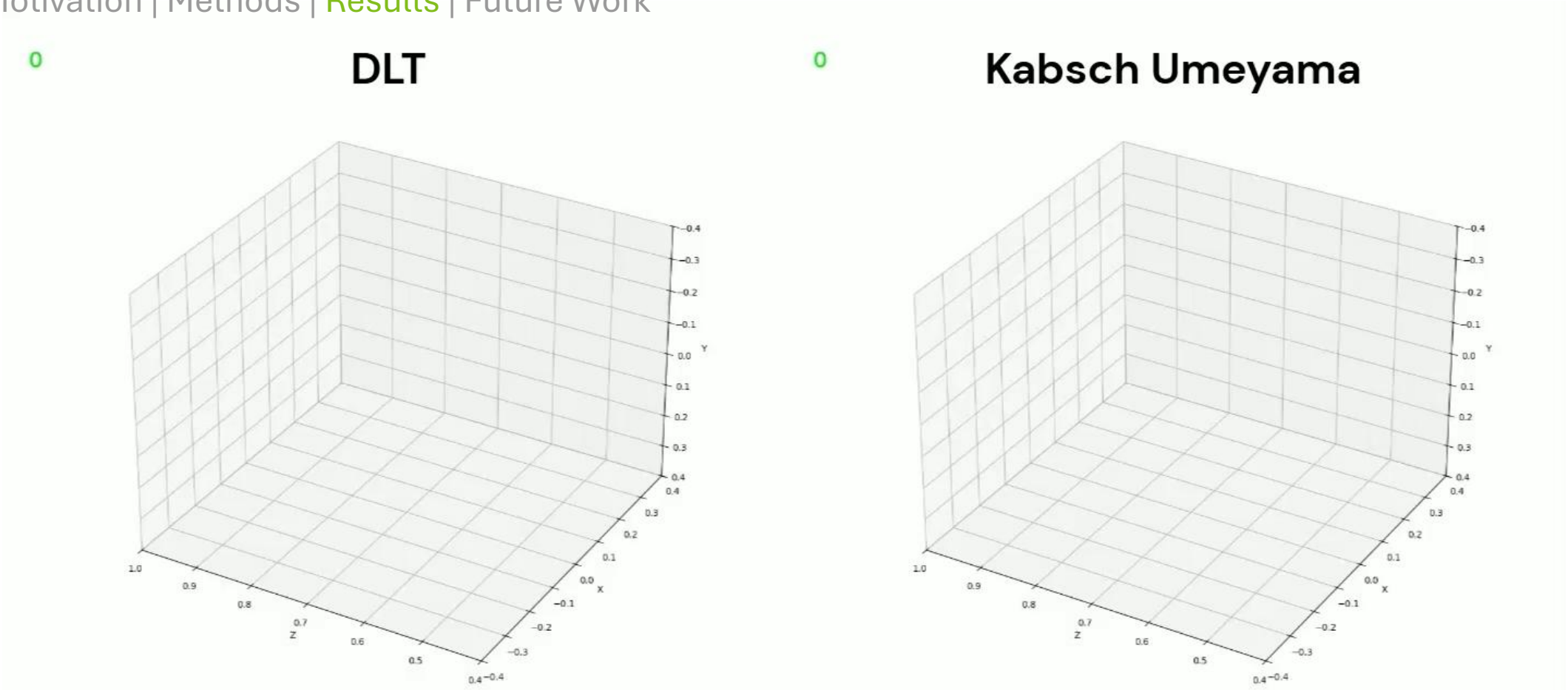


Optimization

Closed Form Solution:
Kabsch Umeyama Algorithm



Results: Optimization

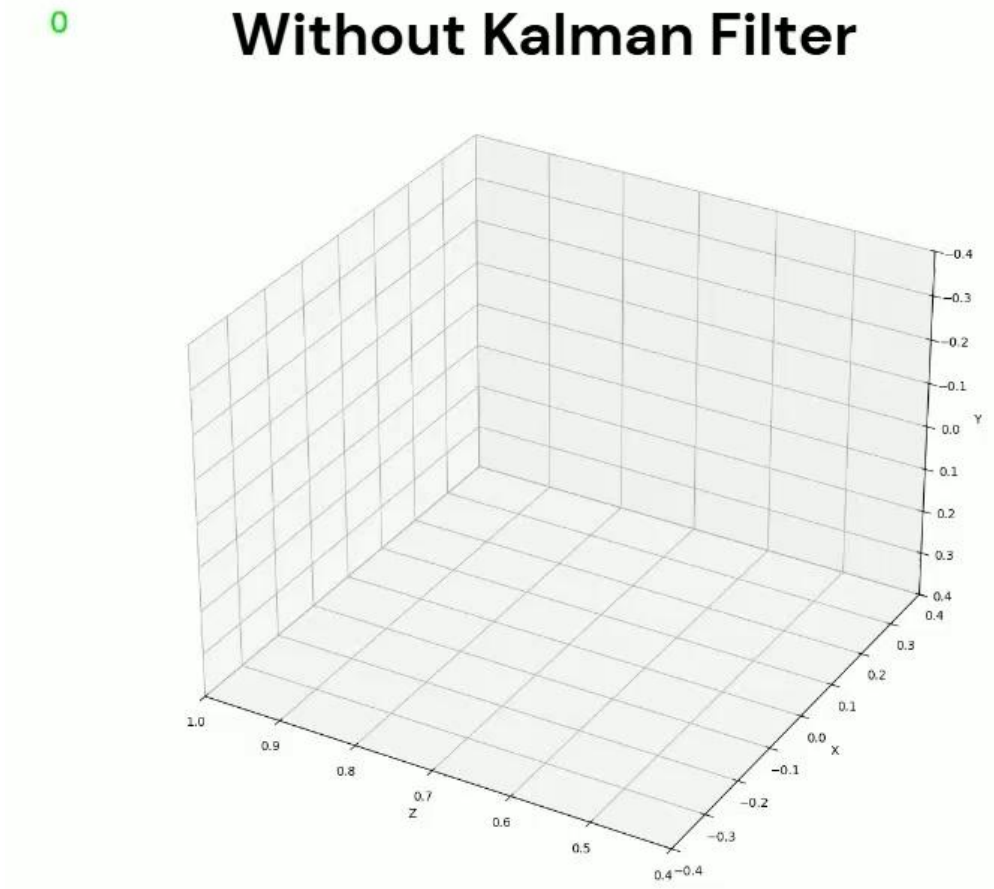


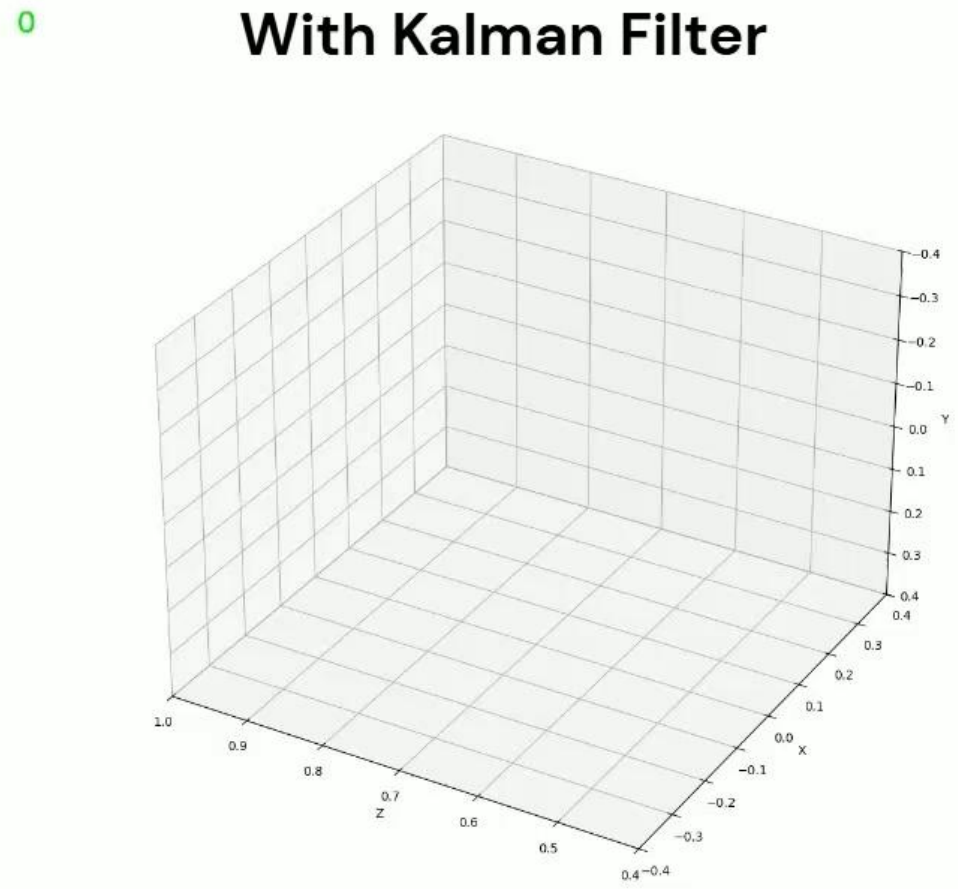
Temporal Noise

Temporal inconsistencies between frames

Kalman Filtering: smoothen the detections

Results: Kalman Filtering





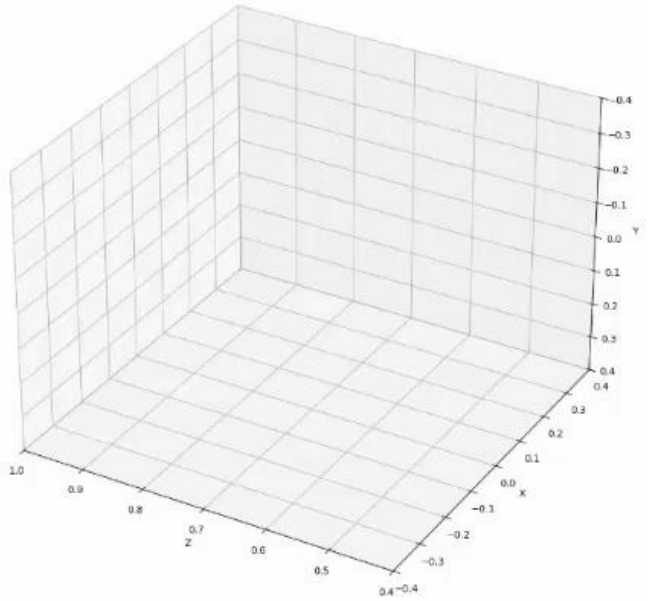
Spatial Noise

Subtle differences in left and right detections

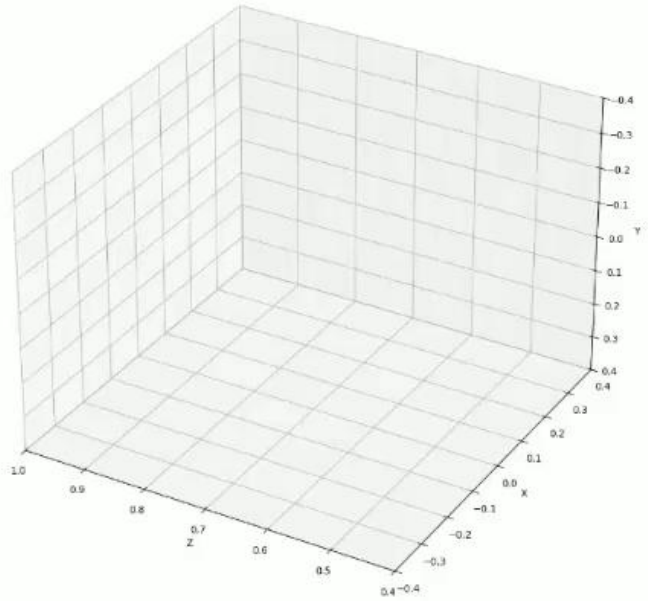
Remove Outliers

Results: Outlier Removal

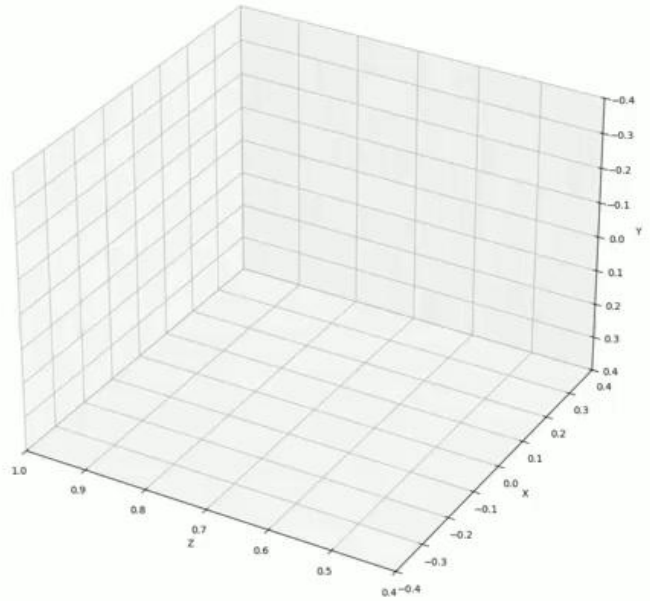
Kabsch Umeyama



With Kalman Filter



Outlier Removal



Goals

Estimate Hand or Body Pose in 3D up to a 10 mm accuracy

Only 2.5D coordinates

Robust to Occlusions caused by Surgical Tools

Predicts parts of the hand occluded by the surgical tool

Portable and Real Time Capable

Validate Results

A Fully Integrated System

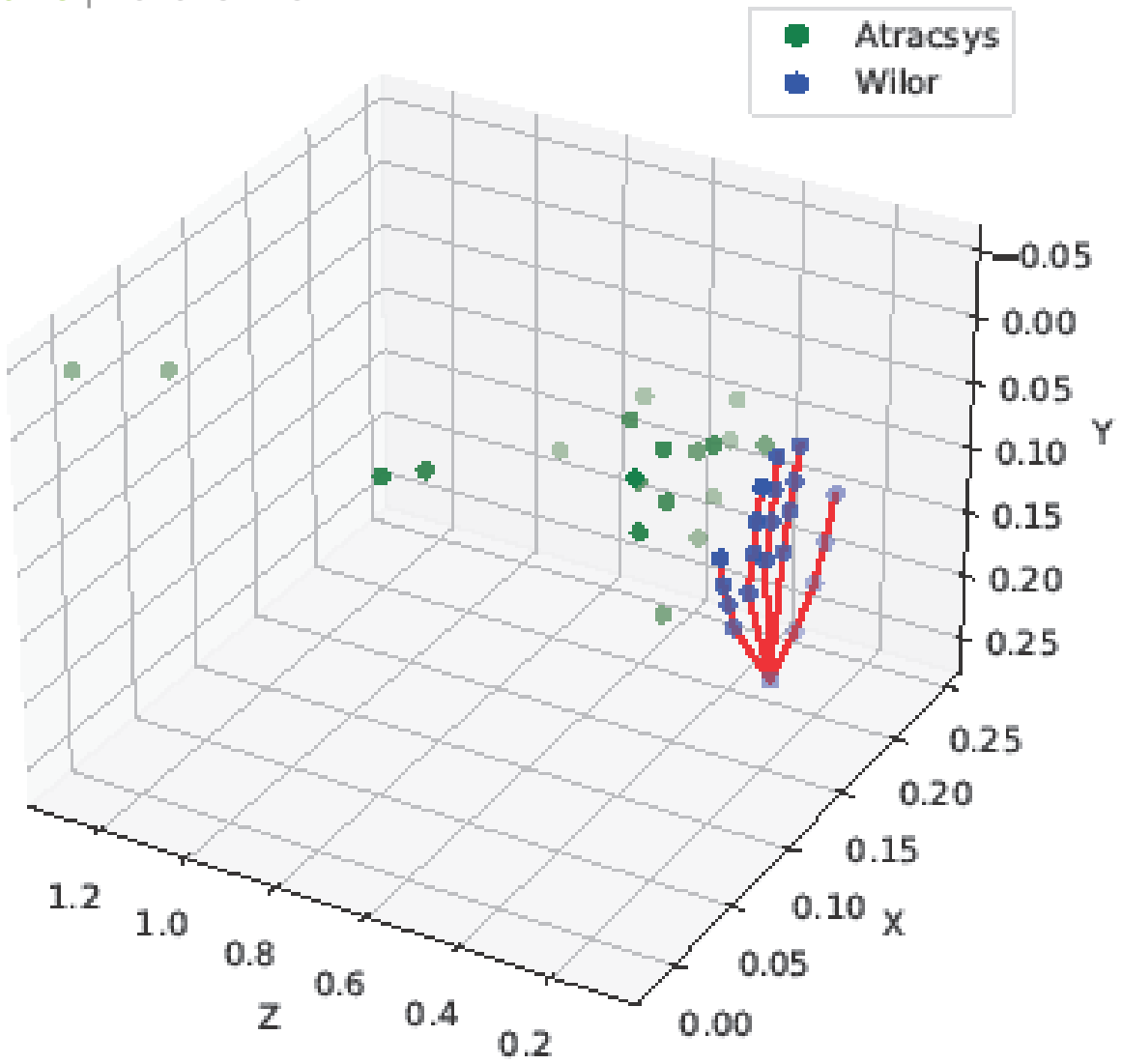
ZED Box NX 16GB

ZED-X

Validation - Atracsys

- Use infrared markers to get 3D coordinates from Atracsys
- Compare with the WiLoR estimates

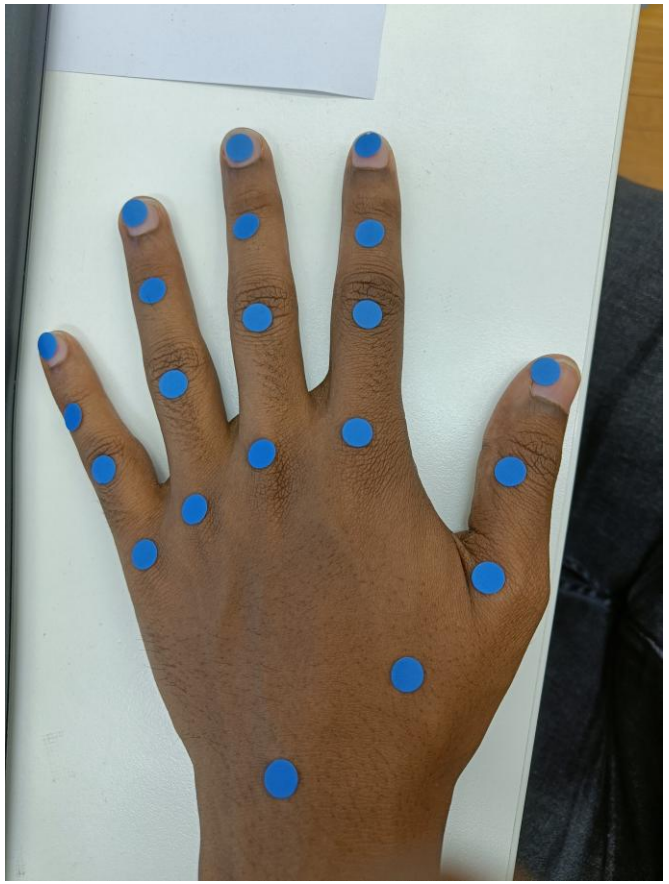
Results: Atracsys



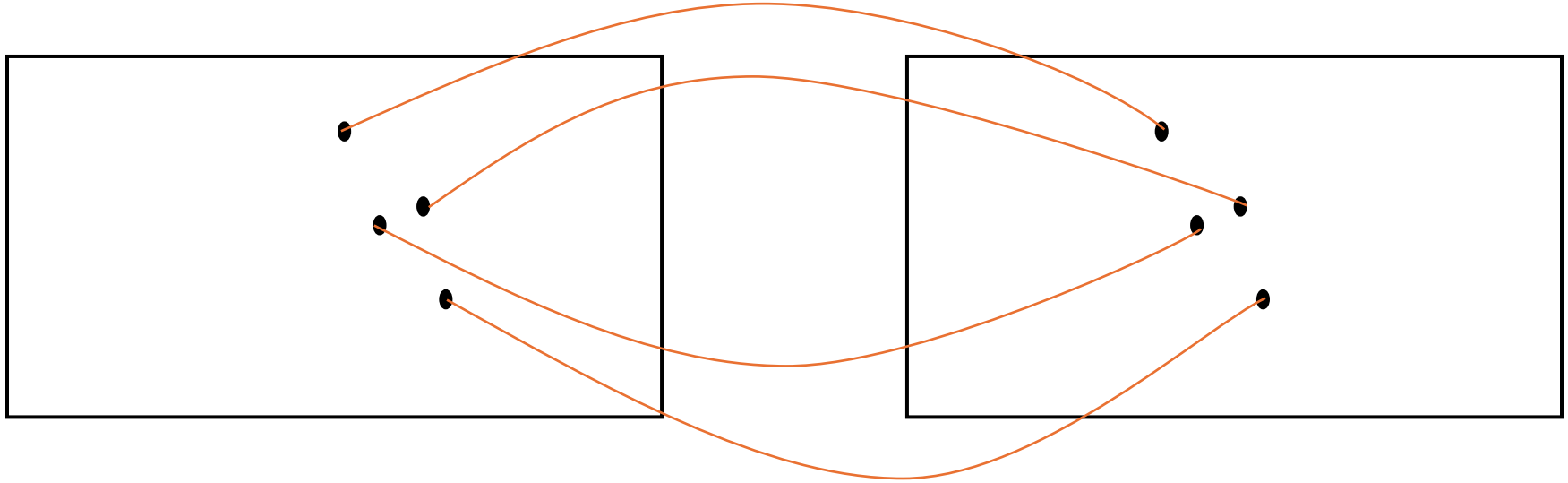
Validation – Blob Markers

- Detect circular markers using Blob Detection
- Compare with the WiLoR estimates

No correspondences



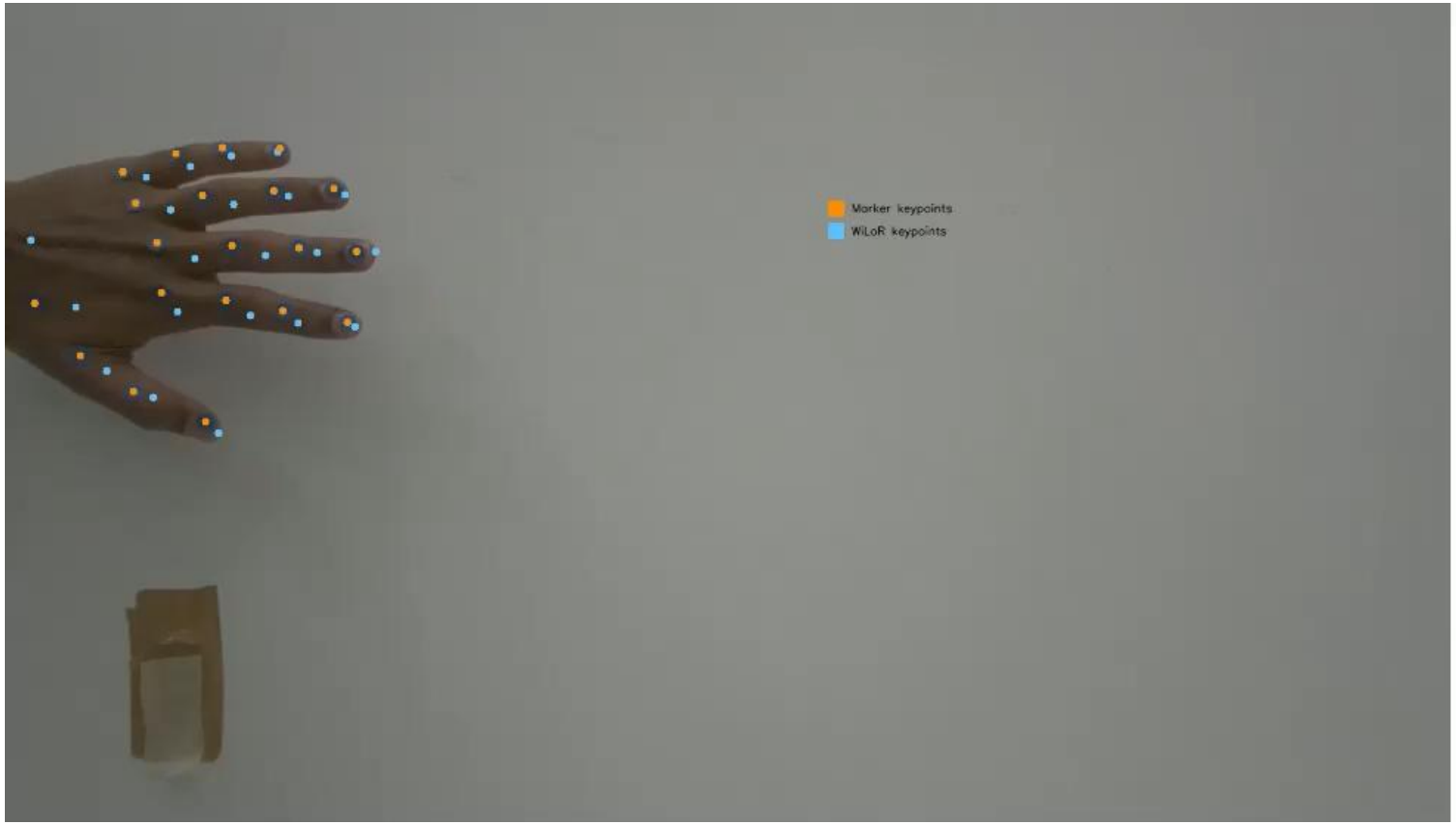
Matching Detected Keypoints

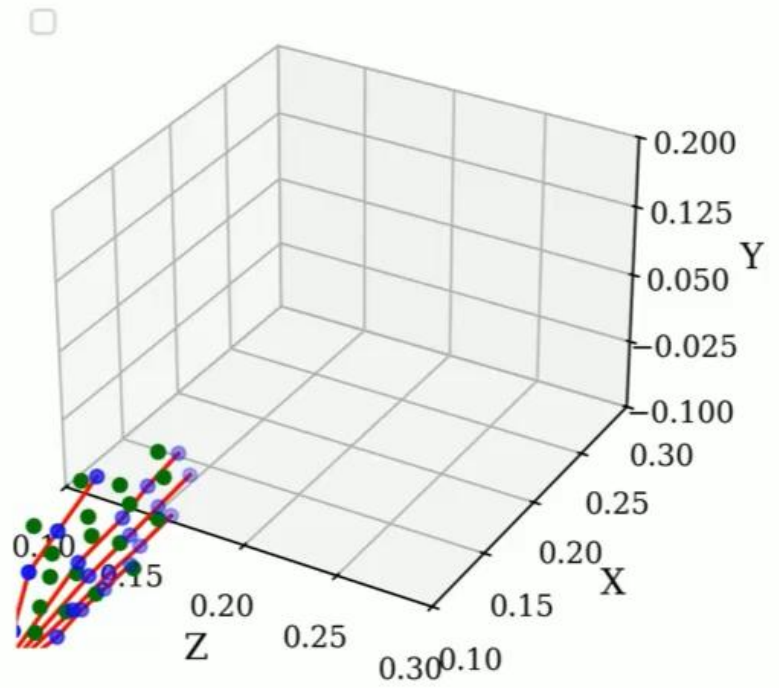


- Group into pairs
- Same y coordinate
- Similar Δx for all pairs

Assignment Problem – Use the **Hungarian Algorithm**

Results: Blob Markers





Goals

Estimate Hand or Body Pose in 3D up to a 10 mm accuracy

3D coordinates

Robust to Occlusions caused by Surgical Tools

Predicts parts of the hand occluded by the surgical tool

Portable and Real Time Capable

Max FPS achieved was 8 FPS

Validate Results

Validated up to ~10 mm accuracy

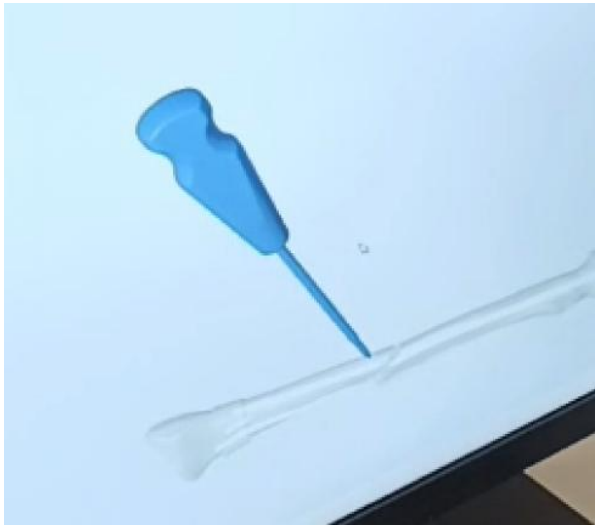
Limitations

- FPS ~ 8 is too low
- Validation test case is too simple

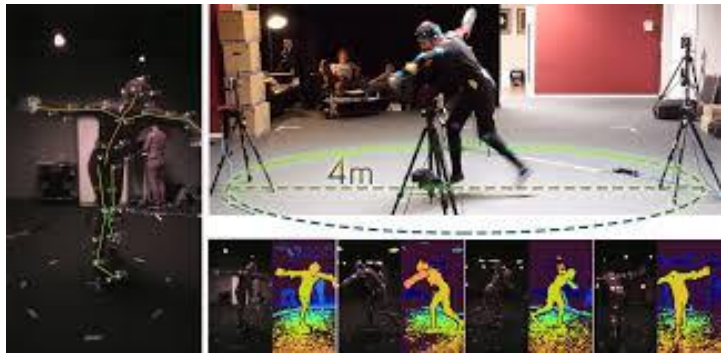
Future Work

Fine tune WiLoR on custom dataset

Use tool pose to refine estimates



More robust validation method



THANK YOU!

Optimization

$$X(p) = Translate \cdot R_z \cdot R_y \cdot R_x \cdot Scale \cdot MANO$$

$$\underset{p}{\text{minimize}} \quad ||X(p) - DLT||_2$$

Closed Form Solution:
Kabsch Umeyama Algorithm