Pose Tracking: Articulated Object Tracking

CS 6334 Virtual Reality
Professor Yu Xiang
The University of Texas at Dallas

Tracking in VR

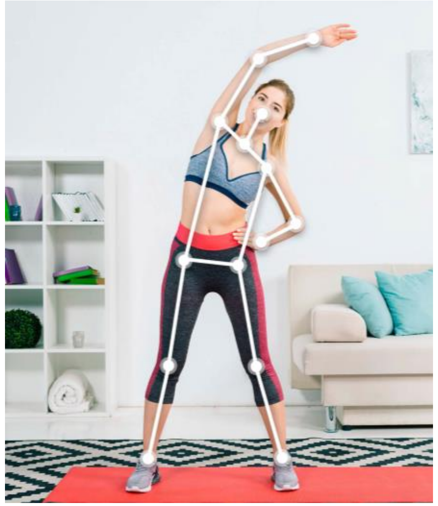
- Tracking the user's sense organs
  - E.g., Head and eye
  - Render stimulus accordingly

- Tracking user's other body parts
  - E.g., human body and hands
  - Locomotion and manipulation

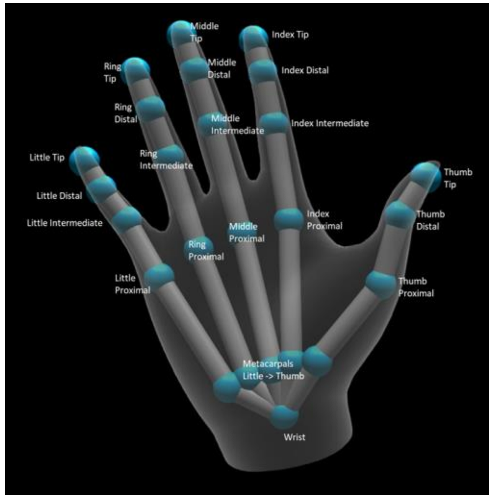
- Tracking the rest of the environment
  - Augmented reality
  - Obstacle avoidance in the real world

Articulated Objects

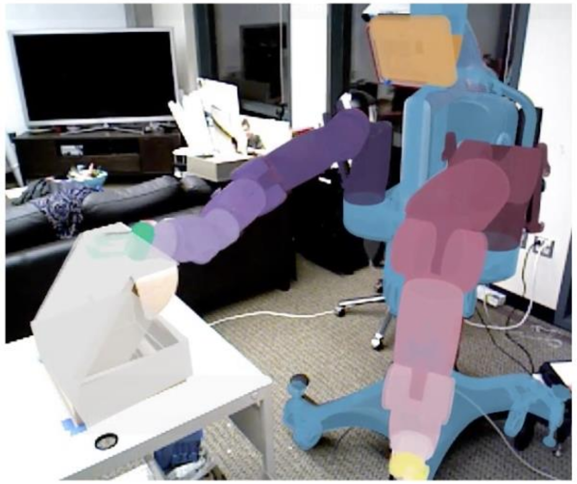
- Objects with joints or articulations
- Links or parts of the object can move relative to each other



Human body



Human hand



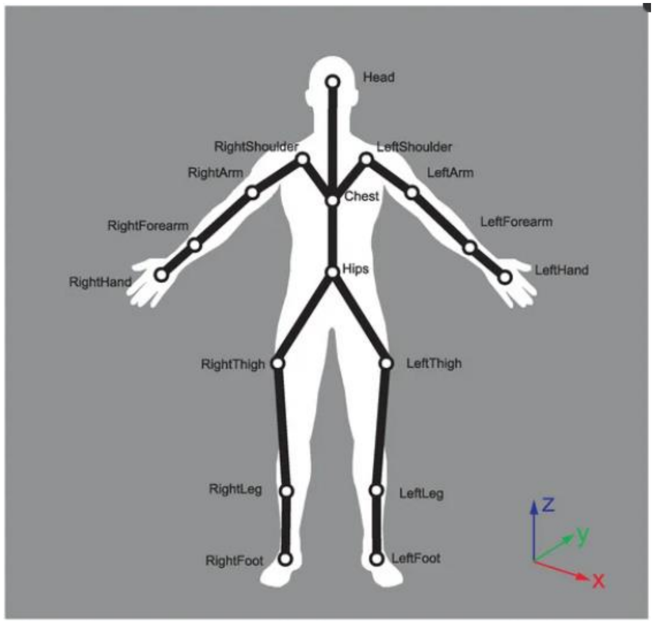
Robot

Drawer

Cabinet

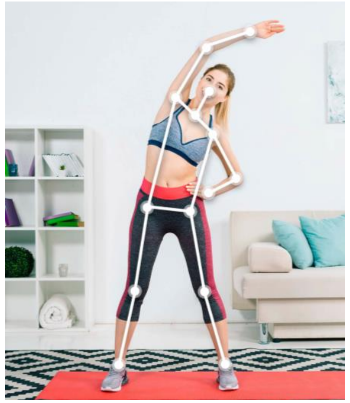
Kinematics

- The study of motion, without regard for the cause of the motion
  - Articulated objects
  - Do not consider the forces that cause the motion

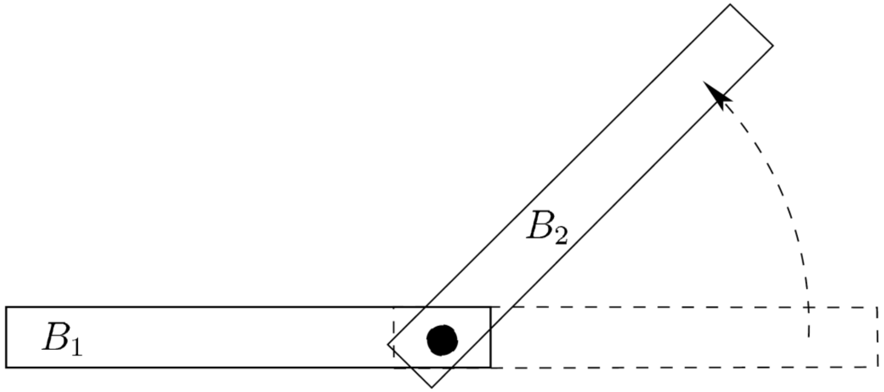


https://www.mdpi.com/1424-8220/17/11/2590/htm

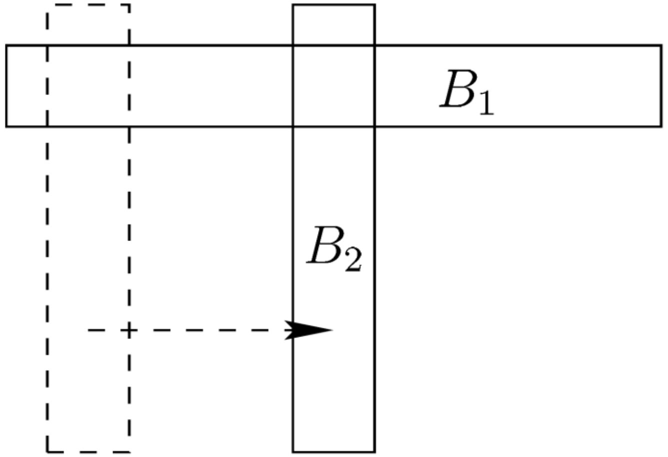
Links and Joints



- Each body of a multibody system is called a link
- A pair of bodies are attached at a joint
  - Revolute joint
  - Prismatic joint



Revolute

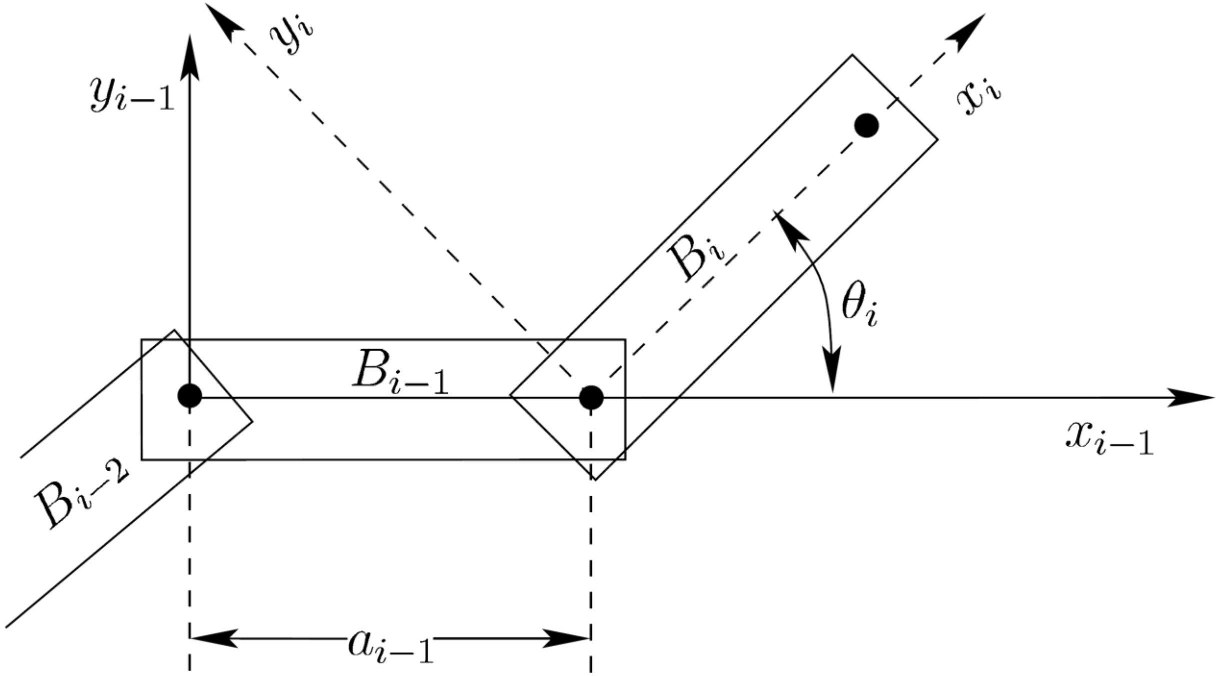


Prismatic

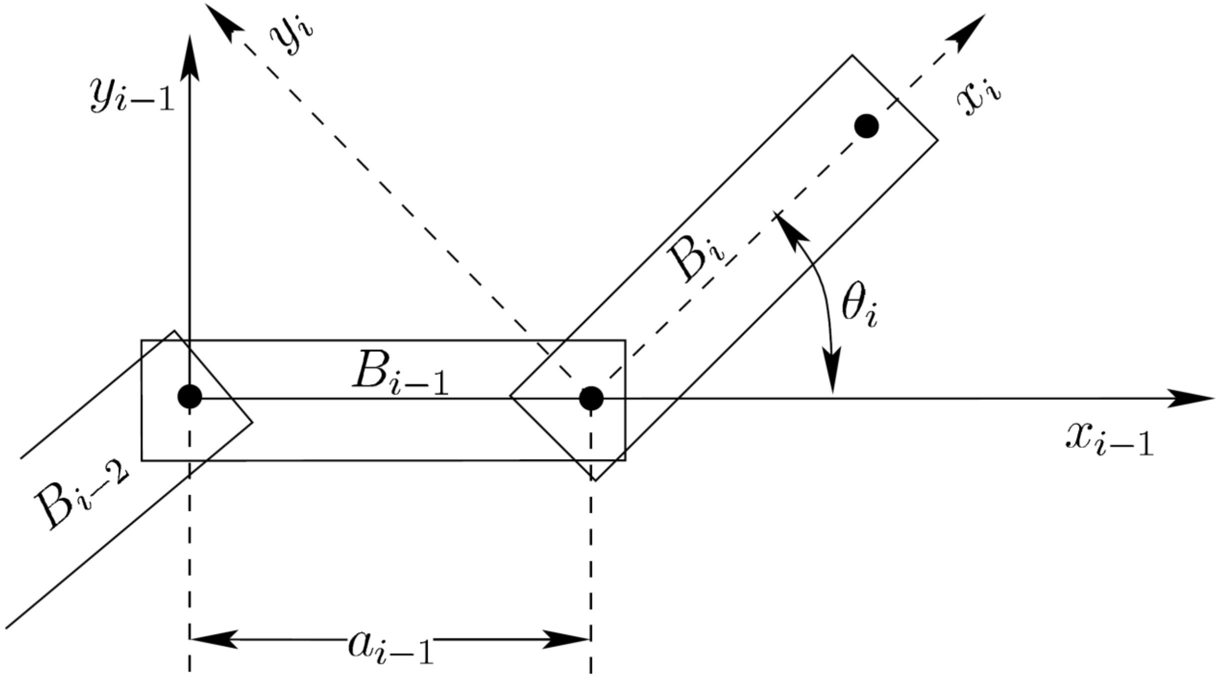
Forward Kinematics

- Given joint parameters, compute the position of a point on the last link in the coordinate frame of the first link (world frame)

- Body frame of each link
  - Origin defined on the joint



Forward Kinematics



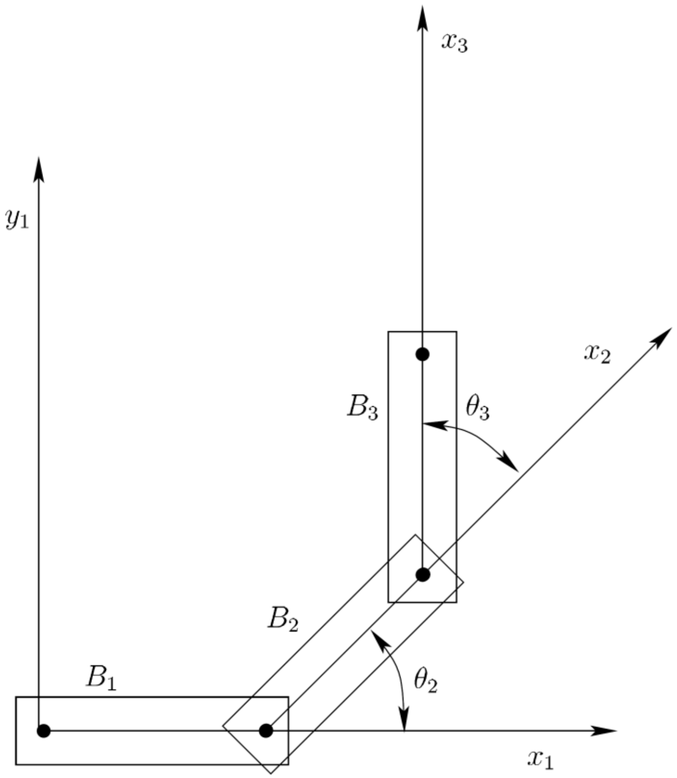
$$T_i = \begin{pmatrix} \cos\theta_i & -\sin\theta_i & a_{i-1} \\ \sin\theta_i & \cos\theta_i & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Transform a point in B_i to the body frame of B_{i-1}

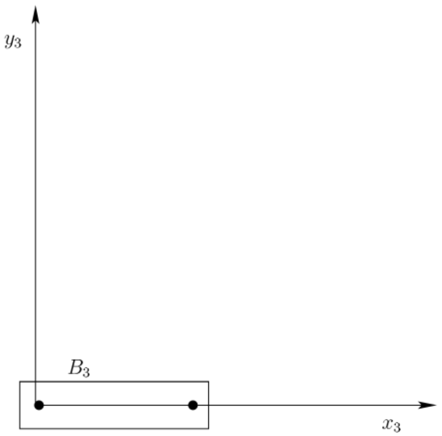
For a revolute joint, θ_i is a variable, a_{i-1} is a constant

For a prismatic joint, θ_i is a constant, a_{i-1} is a variable

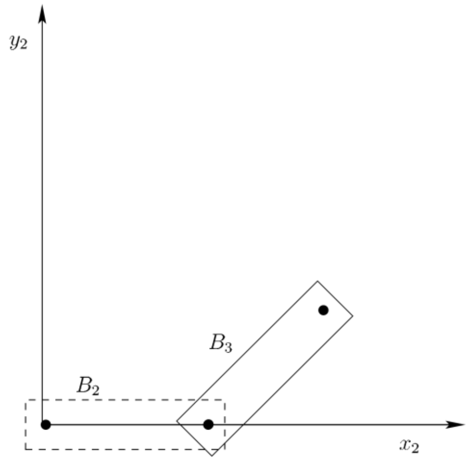
Forward Kinematics



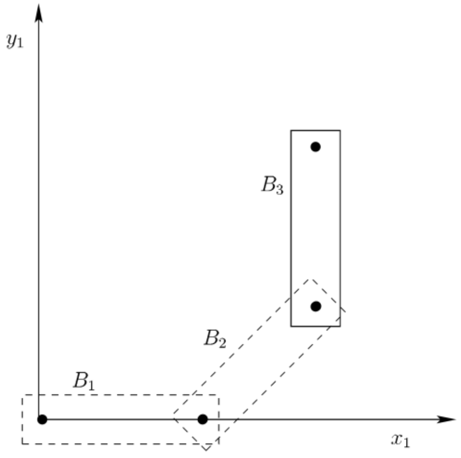
(a) A three-link chain



(b) B_3 in its body frame



(c) T_3 puts B_3 in B_2's body frame



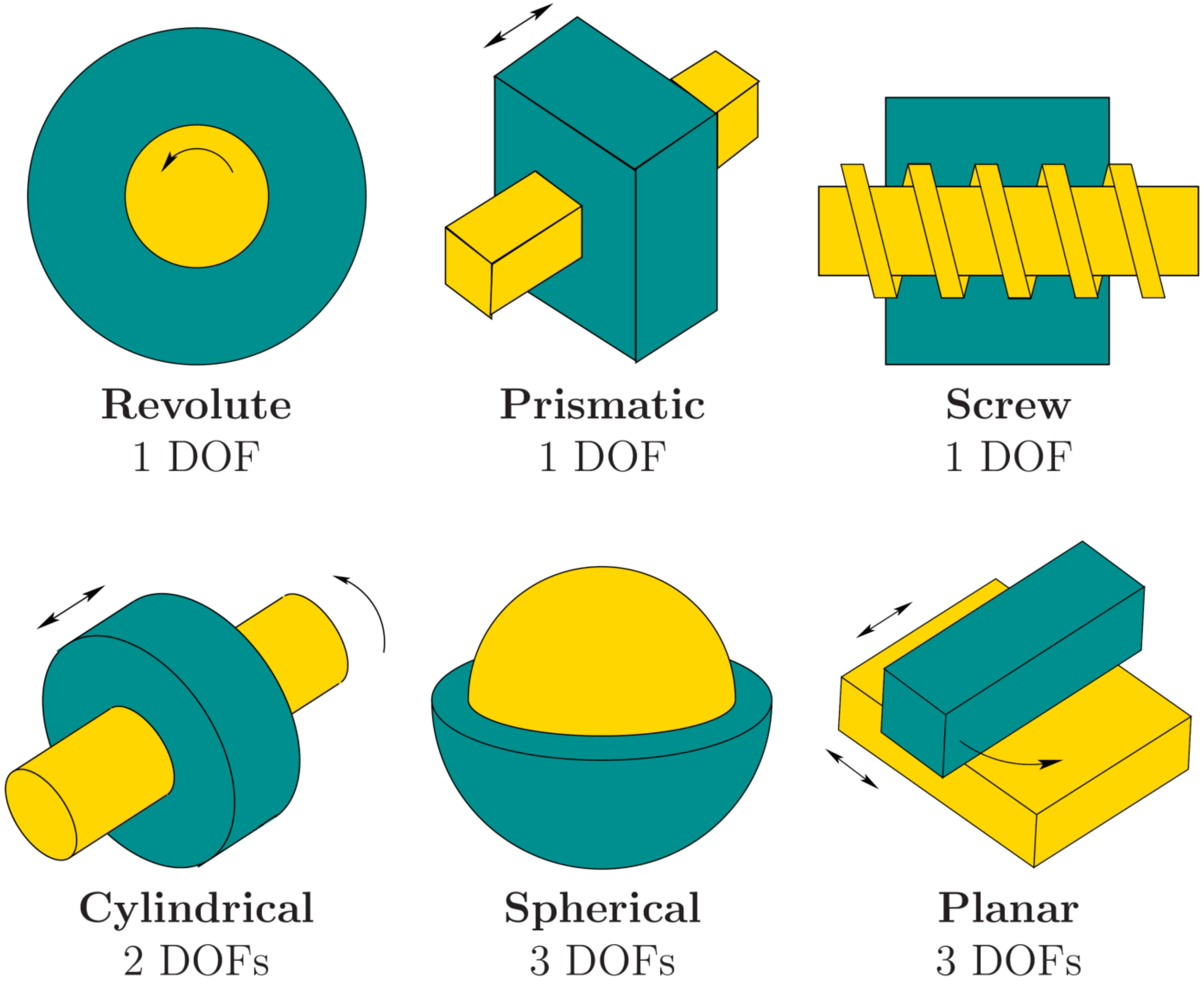
(d) T_2T_3 puts B_3 in B_1's body frame

Kinematic chain

$$T_1 T_2 \cdots T_i$$

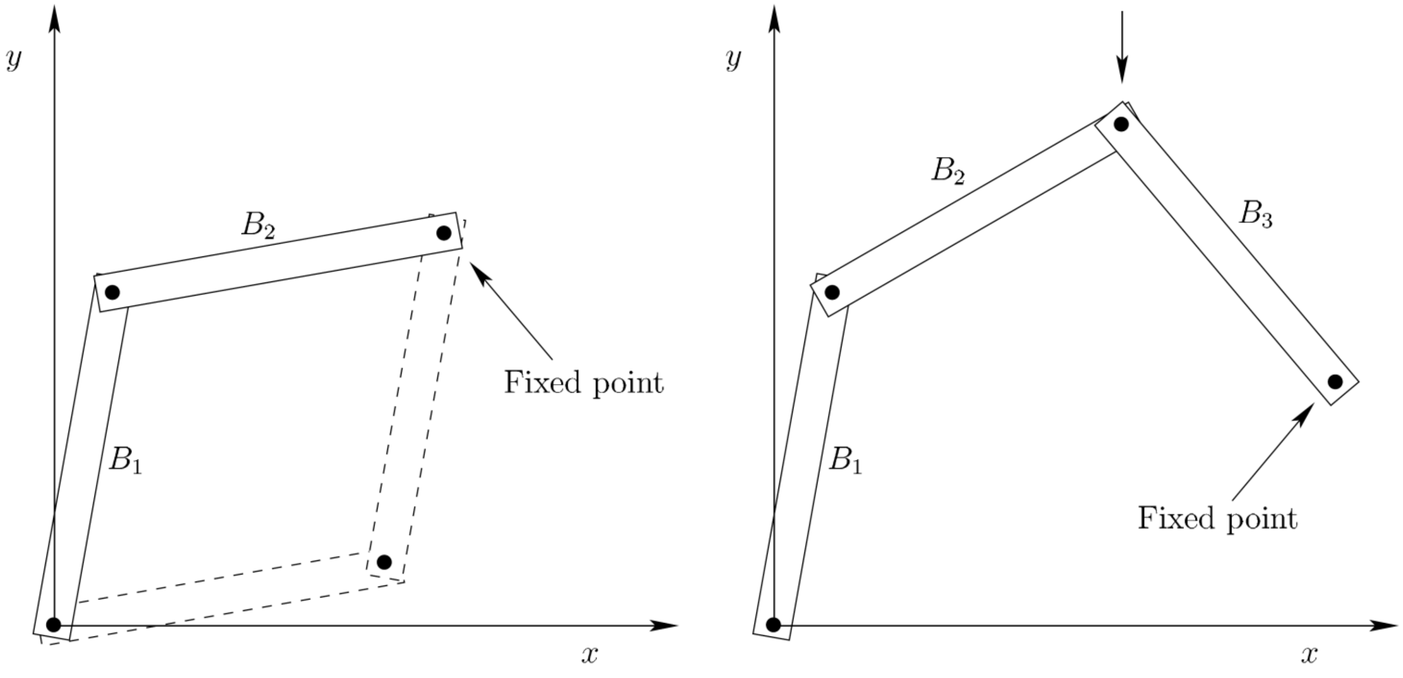
Generalize to 3D

3D Joints


Revolute
1 DOF
Prismatic
1 DOF
Screw
1 DOF
Cylindrical
2 DOFs
Spherical
3 DOFs
Planar
3 DOFs

Inverse Kinematics

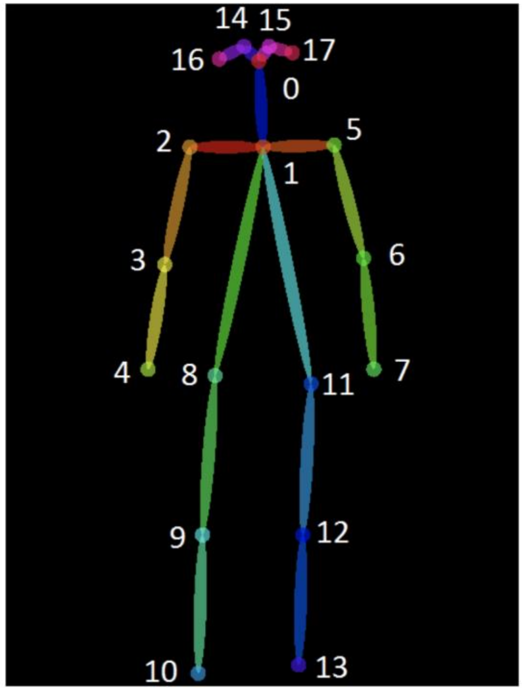
- Compute the joint parameters given the pose of the last link
  - The number of solutions can be infinite, finite, one or zero



(a) (b)

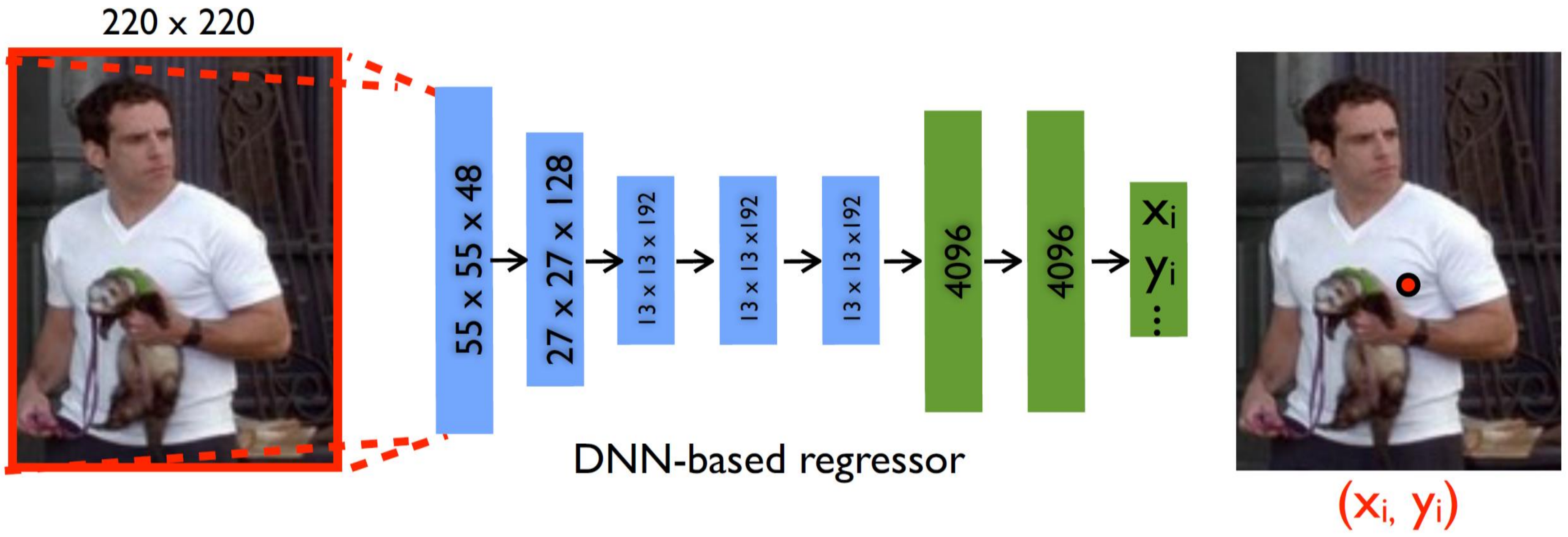
Human Pose Estimation

- Localizing human joints in images or videos
- 2D human pose estimation
  - Detect human joints in images (x, y)
- 3D human pose estimation
  - Detect human joints in 3D (x, y, z)



Human Pose Estimation

- Body joint detection/regression

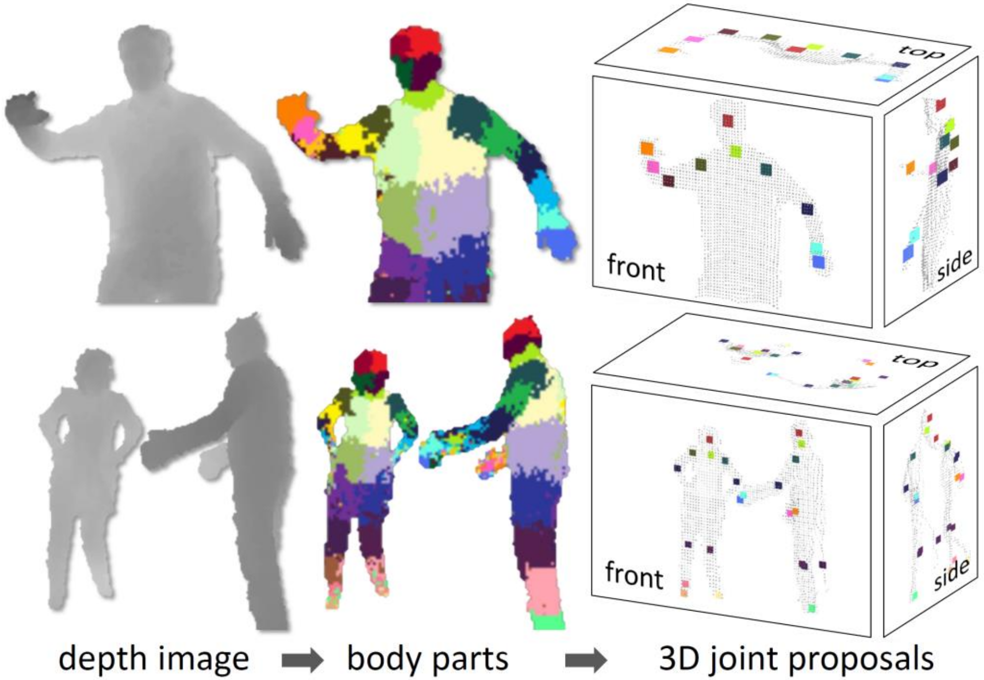


K joint locations

DeepPose: Human Pose Estimation via Deep Neural Networks. Toshev and Szegedy, CVPR'14

Human Pose Estimation

- Kinect: 3D human pose estimation from depth images



Real-Time Human Pose Recognition in Parts from Single Depth Images. Shotton et al, CVPR'11

- Randomized decision forests
- Mean shift to find the modes

Human Pose Estimation

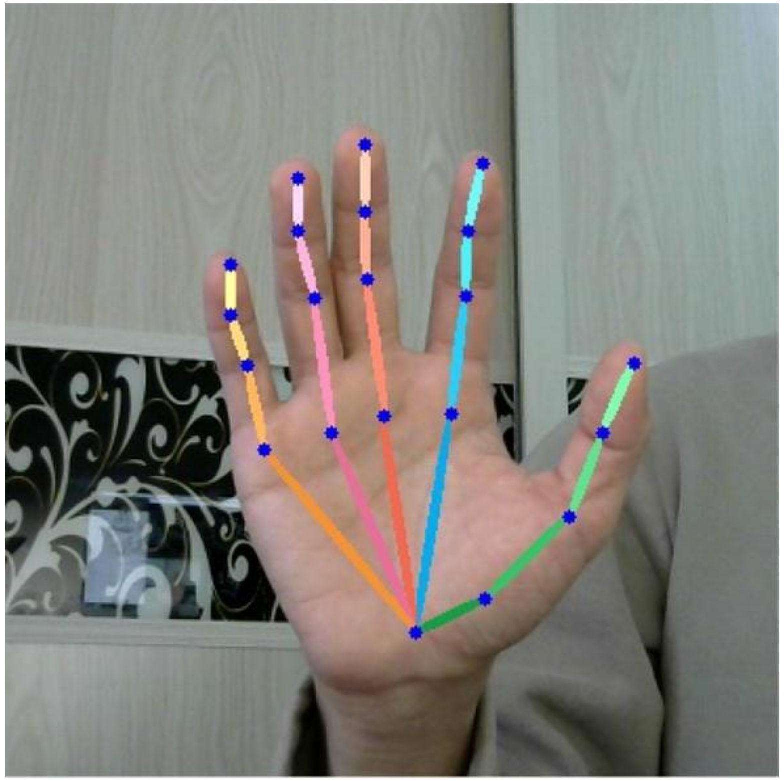
Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields. Cao et al, CVPR'17.

Human Pose Estimation

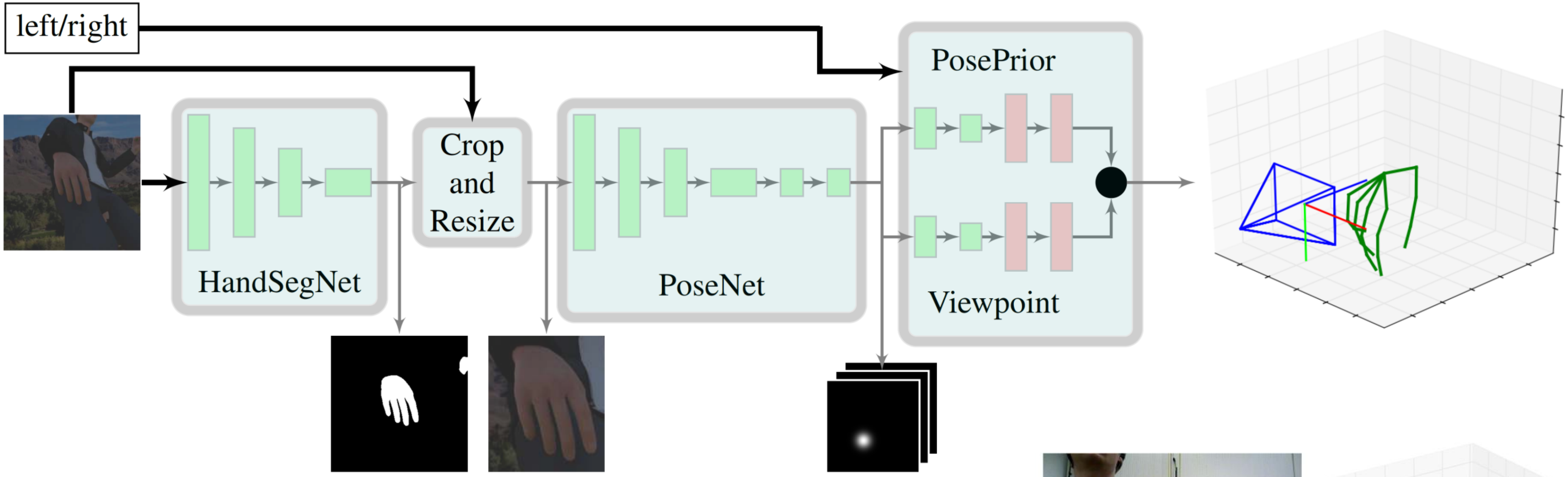
OpenPose: https://github.com/CMU-Perceptual-Computing-Lab/openpose

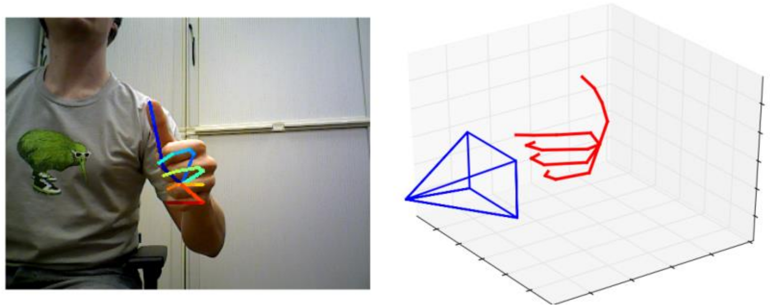
Hand Pose Estimation

- Localizing hand joints in images or videos
- 2D hand pose estimation
  - Detect hand joints in images (x, y)
- 3D hand pose estimation
  - Detect hand joints in 3D (x, y, z)



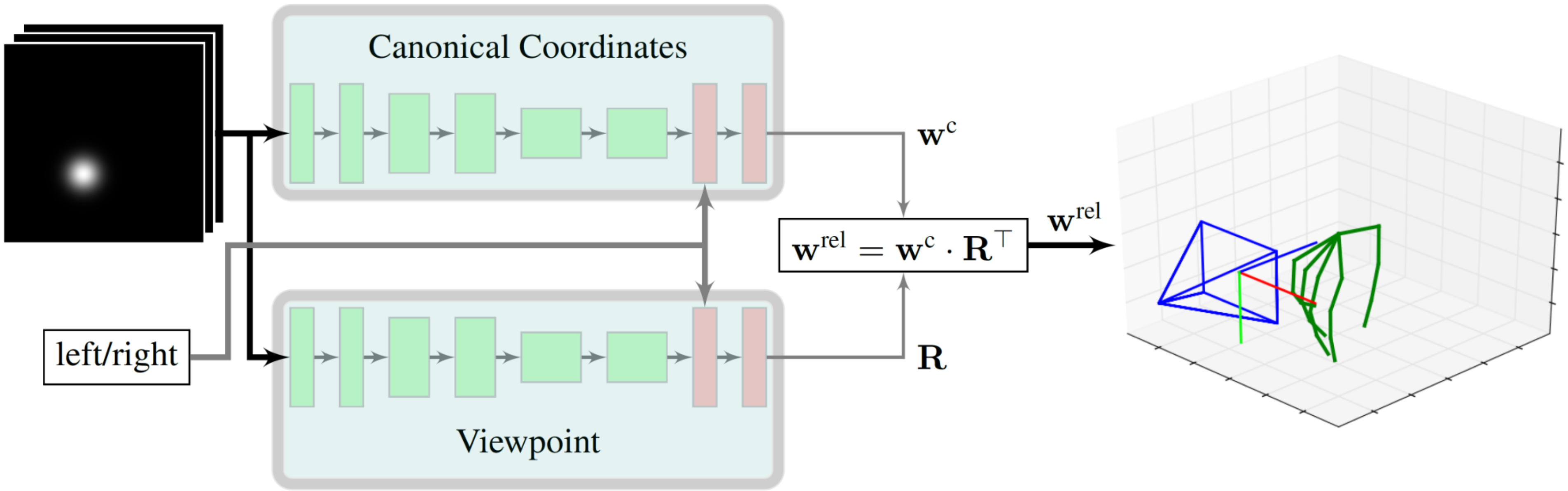
Hand Pose Estimation



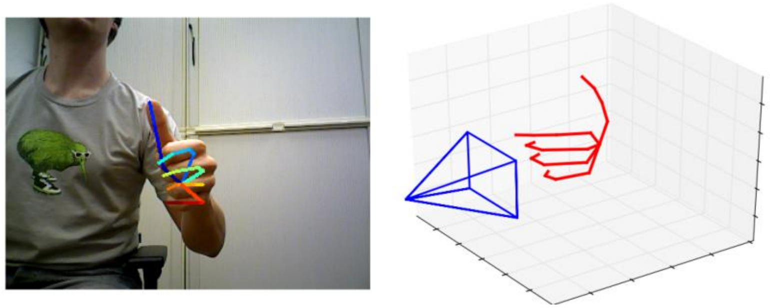


Learning to Estimate 3D Hand Pose from Single RGB Images.
Zimmermann and Brox. ICCV'17.

Hand Pose Estimation

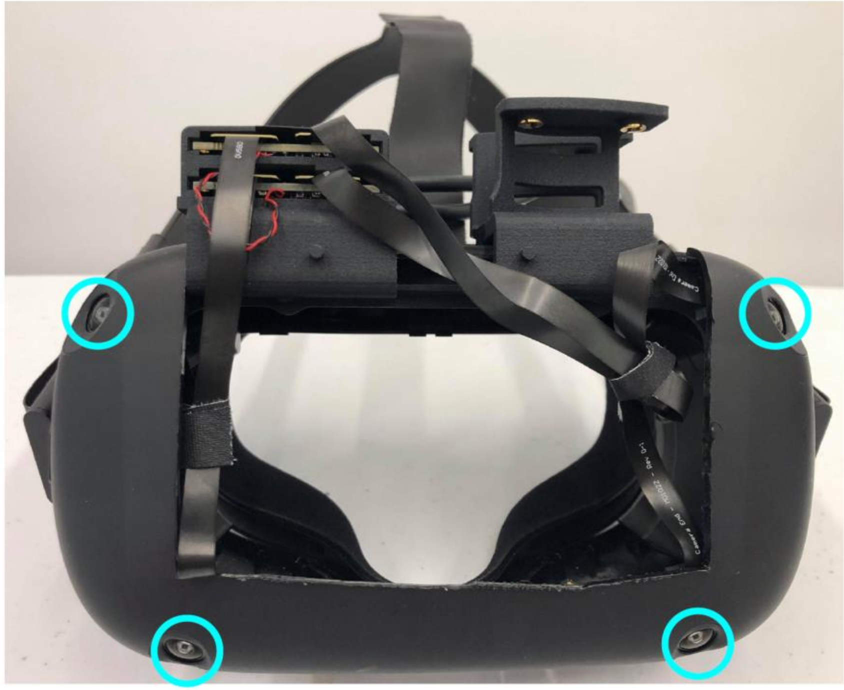


the PosePrior network

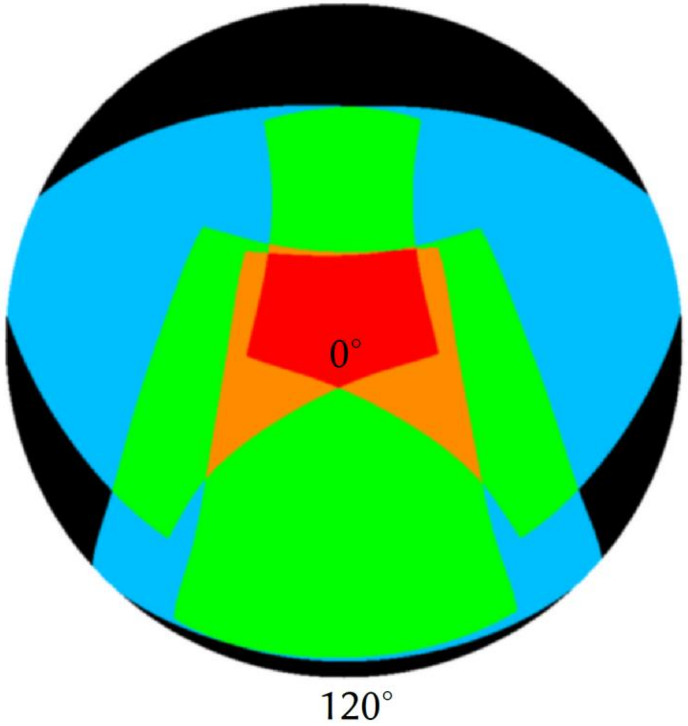


Learning to Estimate 3D Hand Pose from Single RGB Images.
Zimmermann and Brox. ICCV'17.

Hand Pose Estimation for VR



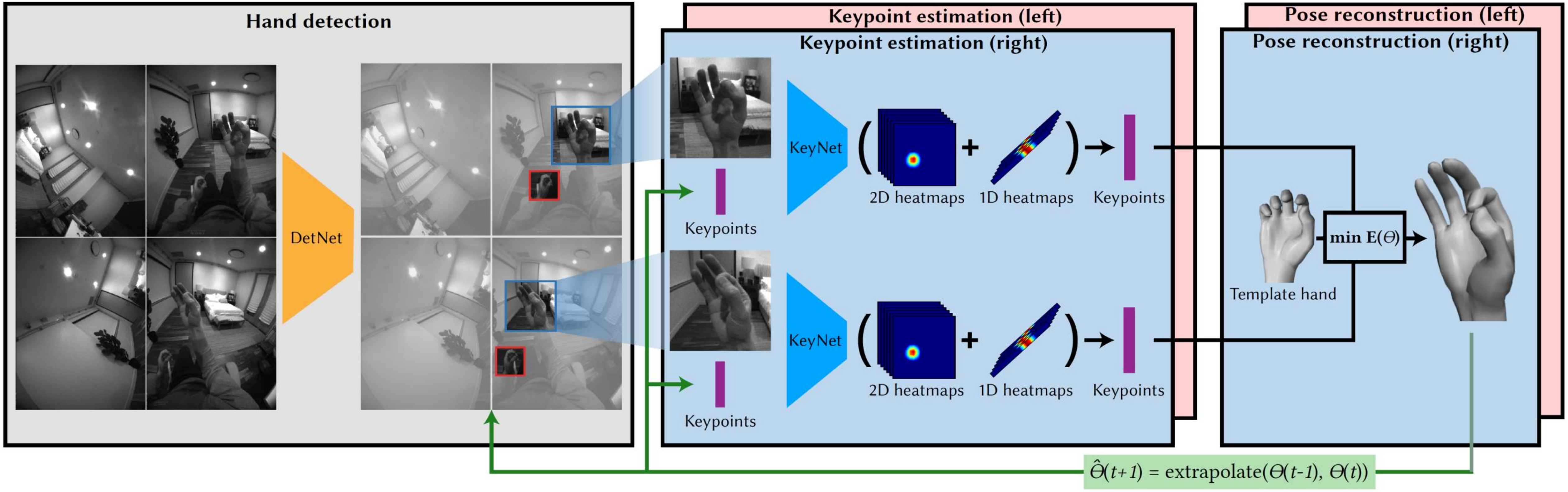
4 monochrome VGA fisheye cameras



field of view at 50cm. 4(red), 3(orange), 2(green), 1(blue), 0(black).

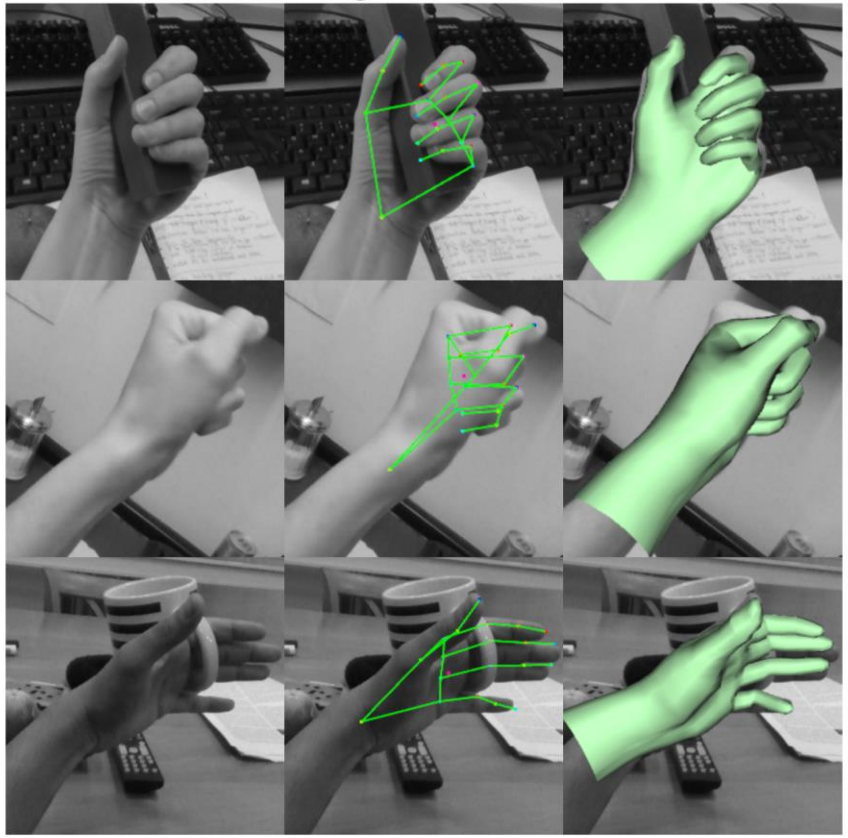
MEgATrack: Monochrome Egocentric Articulated Hand-Tracking for Virtual Reality. Han et al., SIGGRAPH’20.

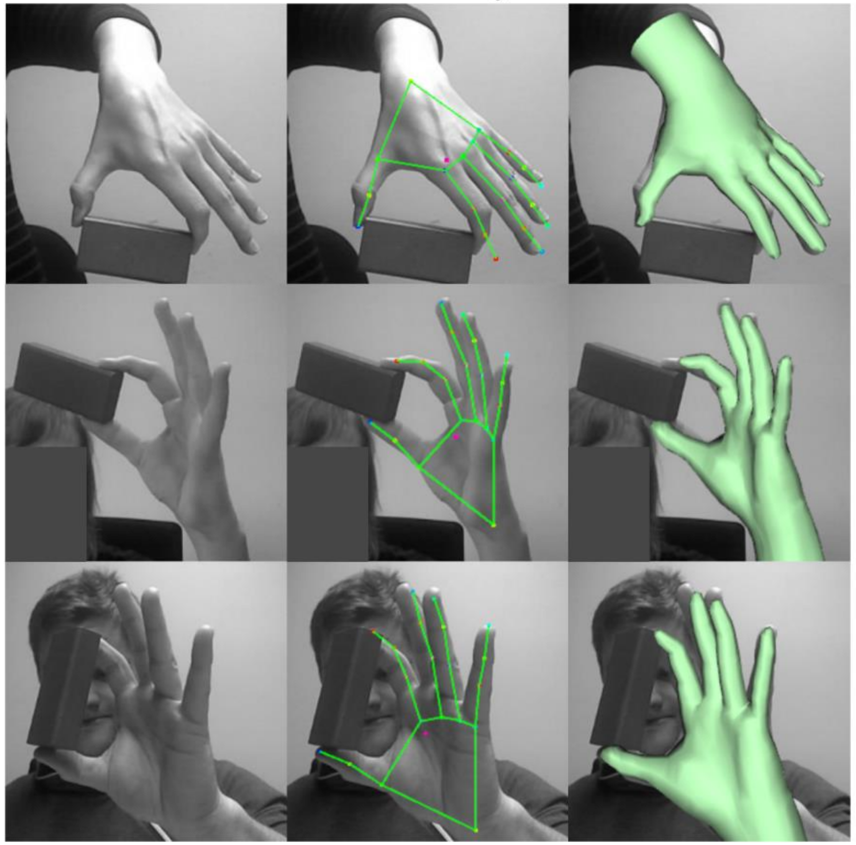
Hand Pose Estimation for VR



MEgATrack: Monochrome Egocentric Articulated Hand-Tracking for Virtual Reality. Han et al., SIGGRAPH'20.

Hand Pose Estimation for VR





MEgATrack: Monochrome Egocentric Articulated Hand-Tracking for Virtual Reality. Han et al., SIGGRAPH'20.

Hand Pose Estimation for VR

https://research.fb.com/publications/megatrack-monochrome-egocentric-articulated-hand-tracking-for-virtual-reality/

MEgATrack: Monochrome Egocentric Articulated Hand-Tracking for Virtual Reality. Han et al., SIGGRAPH’20.

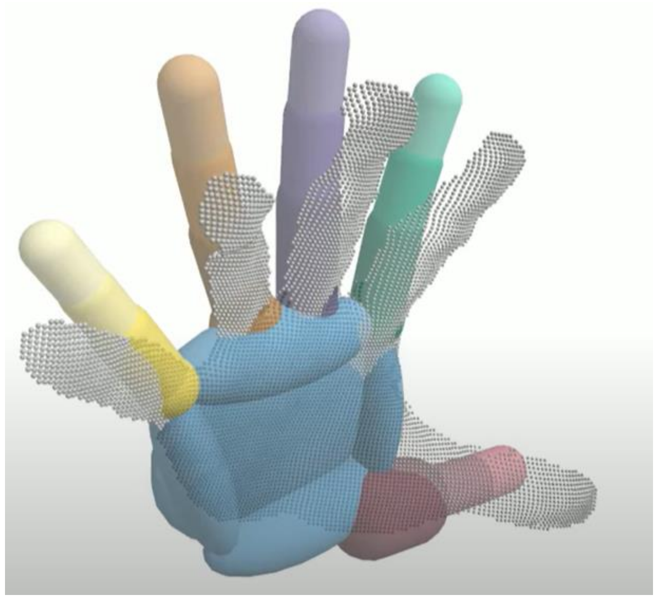
Hand Pose Estimation for VR

- Leap Motion
  - 3 infrared LEDs
  - A stereo IR camera
  - No details of the tracking algorithm available

https://www.youtube.com/watch?v=NzAaMbEHrpg

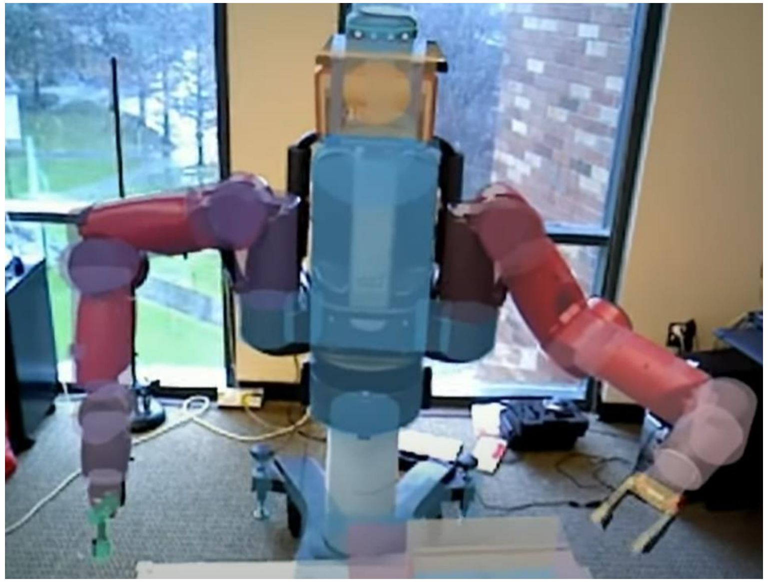
Model-based Articulated Object Tracking

- Given a 3D model of an articulated object, match the 3D model to the input image (RGB or depth)



Human hand

Human body



Robot

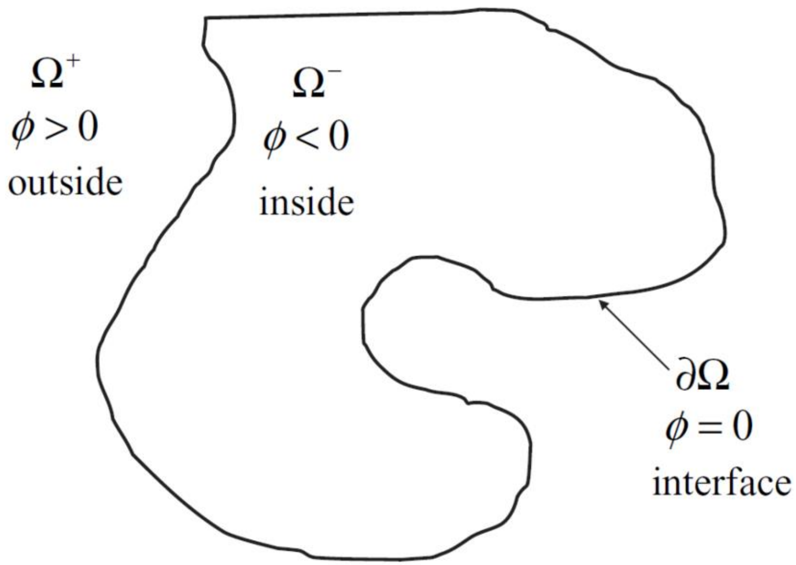
DART: Dense Articulated Real-Time Tracking. Schmidt, Newcombe and Fox, RSS’14.

Model-based Articulated Object Tracking

- Signed Distance Function (SDF)

$$\phi : \Omega \subseteq \mathbb{R}^3 \rightarrow \mathbb{R}$$

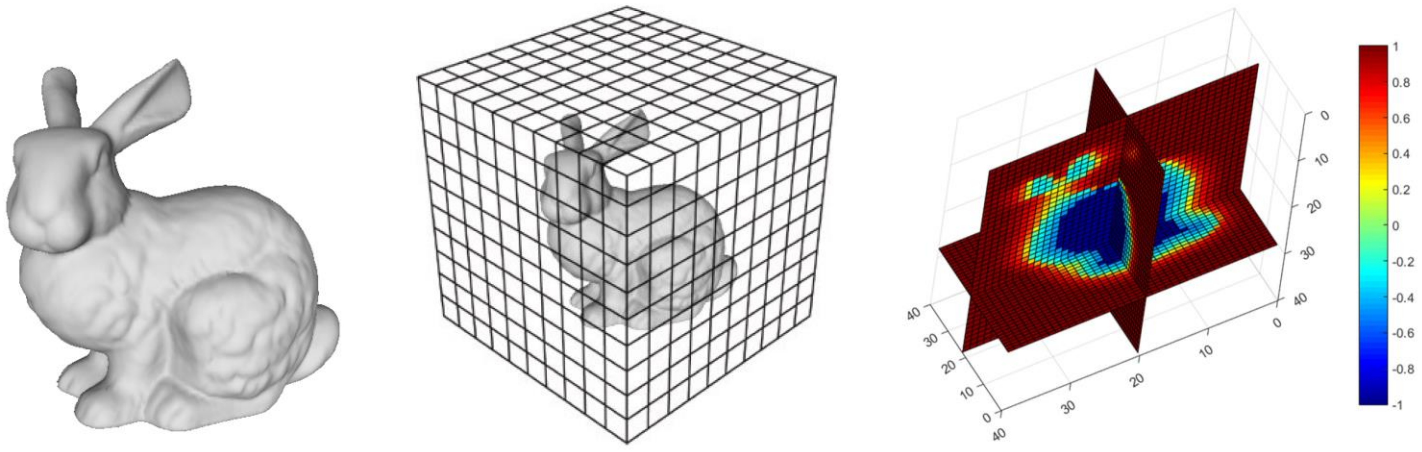
Signed distance to the closest object boundary



DART: Dense Articulated Real-Time Tracking. Schmidt, Newcombe and Fox, RSS'14.

Model-based Articulated Object Tracking

- Signed Distance Function (SDF)



(a) Surface view. (b) Bounding volume. (c) Generated SDF.

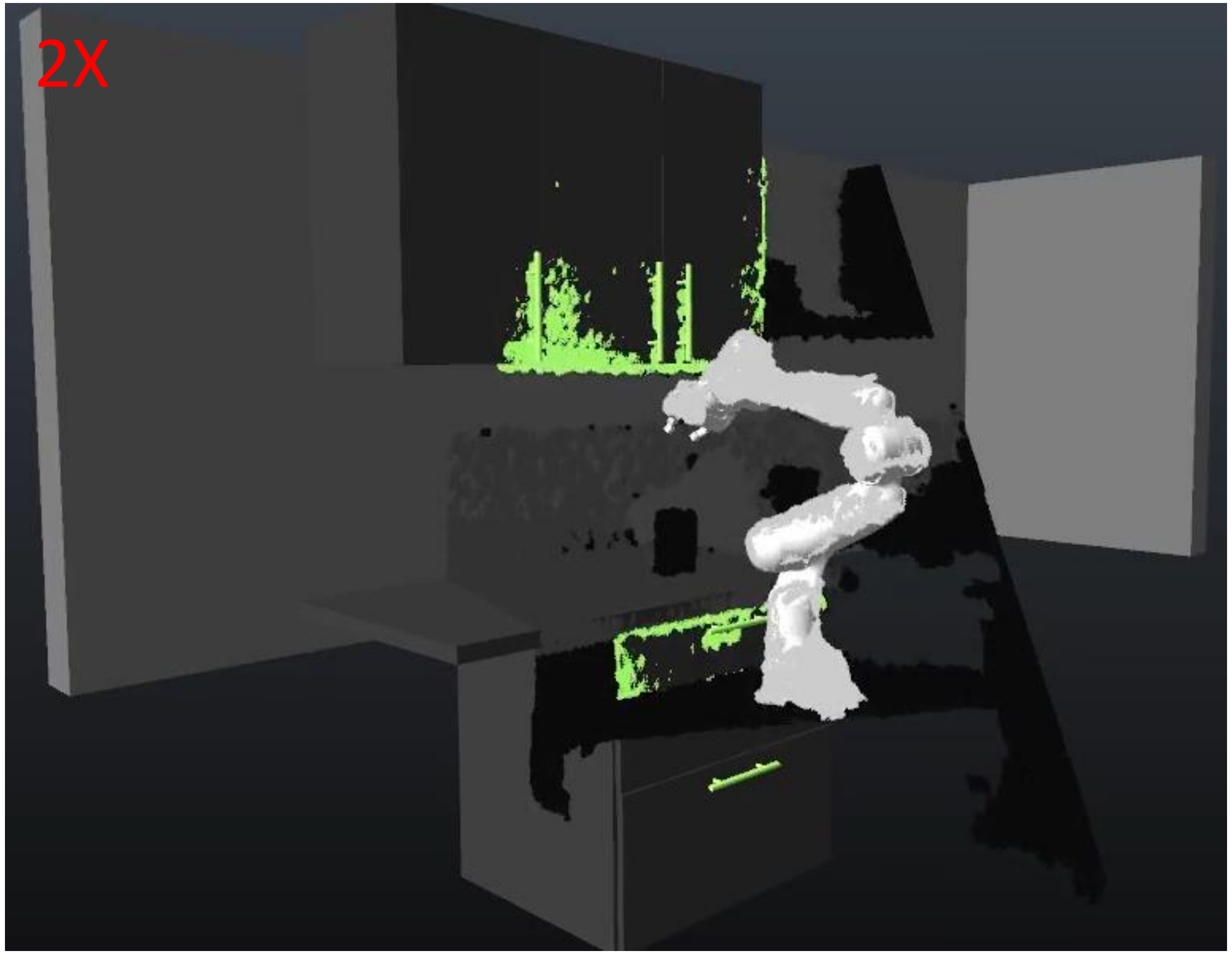
Signed Distance Fields for Rigid and Deformable 3D Reconstruction. Miroslava Slavcheva.

DART idea

- Compute SDF from the 3D model of each link
- Estimate link poses that can match point cloud from depth image to the SDFs

DART: Dense Articulated Real-Time Tracking. Schmidt, Newcombe and Fox, RSS'14.

Model-based Articulated Object Tracking



DART: Dense Articulated Real-Time Tracking Schmidt, Newcombe and Fox, RSS'14.

Further Reading

- Section 9.4, Virtual Reality, Steven LaValle
- MEgATrack: Monochrome Egocentric Articulated Hand-Tracking for Virtual Reality. Han et al., SIGGRAPH'20.
- DART: Dense Articulated Real-Time Tracking. Schmidt, Newcombe and Fox, RSS'14.