Agimus Winter School
11/12/2023 - 15/12/2023
Banyuls (France)

AGIMUS
INNOVATIVE ROBOTICS FOR AGILE PRODUCTION

Model-based object pose tracking

Médéric Fourmy
Czech Technical University, Prague

Founded by the European Union under GA no 101070165

Object pose tracking

Initial pose

Converged

Object pose tracking

Initial pose

Converged

- Assumptions: object detected, matched with model, initial pose

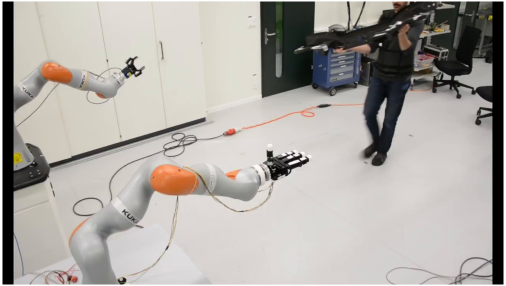
Object pose tracking

Initial pose

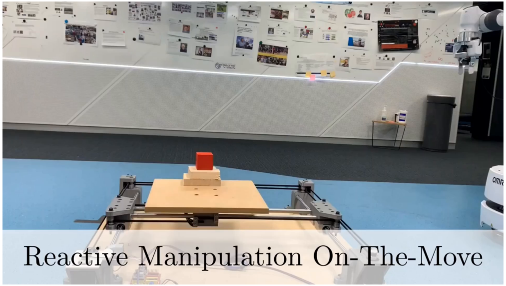
Converged

- Assumptions: object detected, matched with model, initial pose
- Local refinement of ${}^{c}\mathrm{T}_b \in SE(3)$ pose using a single RGB(-D) camera

Motivation: dynamic manipulation

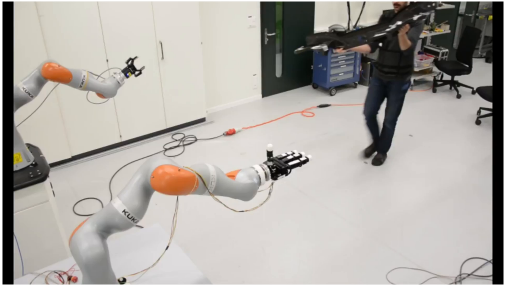


Human to robot handover [MFB18]

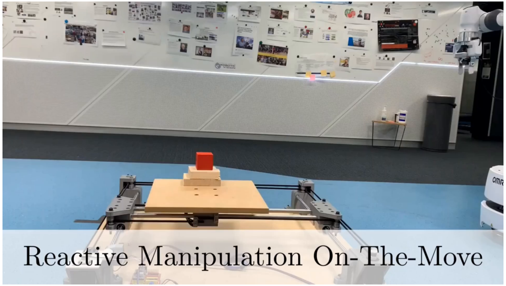


Object grasping on the move [Bur+23]

Motivation: dynamic manipulation



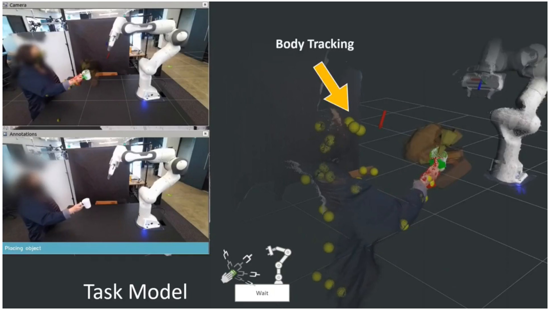
Human to robot handover [MFB18]



Object grasping on the move [Bur+23]

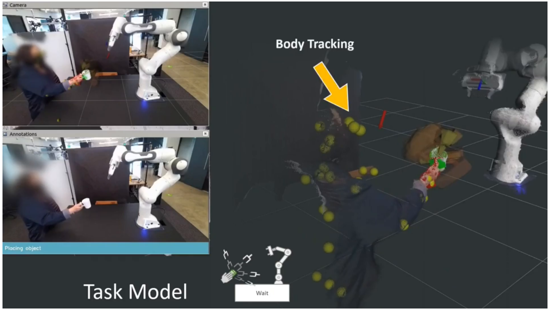
- Low latency estimation to close the loop
- Grasping level precision ($\sim$ cm)

Handover from point cloud grasp prediction



Model predictive control for fluid human-to-robot handovers [Yan+22]

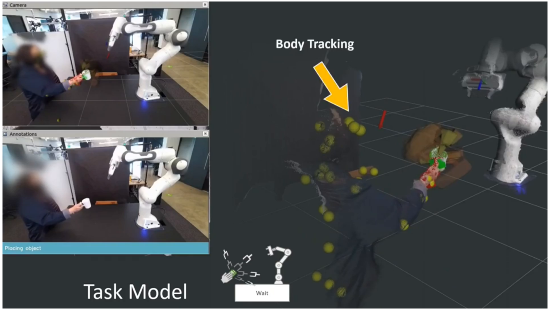
Handover from point cloud grasp prediction



Model predictive control for fluid human-to-robot handovers [Yan+22]

- Generates grasp proposals from point cloud (GraspNet)

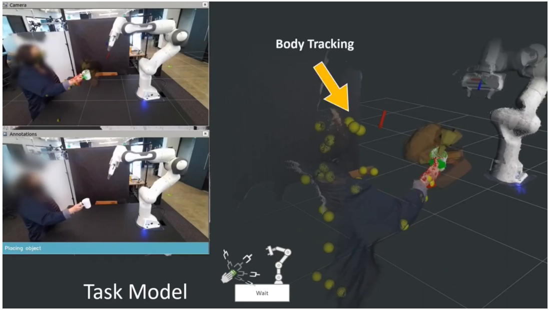
Handover from point cloud grasp prediction



Model predictive control for fluid human-to-robot handovers [Yan+22]

- Generates grasp proposals from point cloud (GraspNet)
- Runs on 6 GPUs in parallel

Handover from point cloud grasp prediction



Model predictive control for fluid human-to-robot handovers [Yan+22]

- Generates grasp proposals from point cloud (GraspNet)
- Runs on 6 GPUs in parallel
- What if we have a decent object model?

Overview

1. Model based object tracking, a short tour
2. Region based object tracking
3. Object localization and tracking for conrol

Model based object tracking

A short tour

Edges tracking

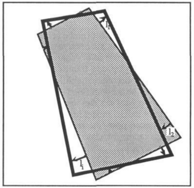


Figure 2: Diagram to show a sample of perpendicular distances, l_i

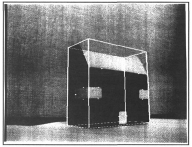


Figure 4: RAPID tracking box in a static situation

RAPID [HS90]

- **Model**: 3D geometric primitives
- **Method**: Local search for image edges from contour points, least squares

Edges tracking

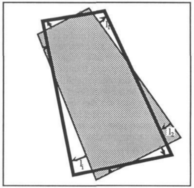


Figure 2: Diagram to show a sample of perpendicular distances, l_i

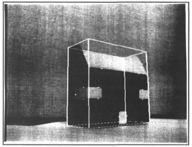
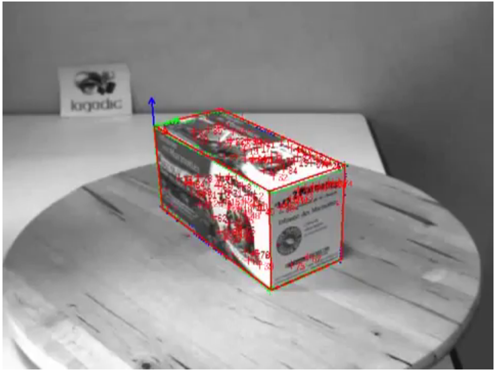


Figure 4: RAPID tracking box in a static situation

RAPID [HS90]

- **Model**: 3D geometric primitives
- **Method**: Local search for image edges from contour points, least squares
- First real time methods
- Sensitive to incorrect matches (background clutter, self occlusion), additional modelling step

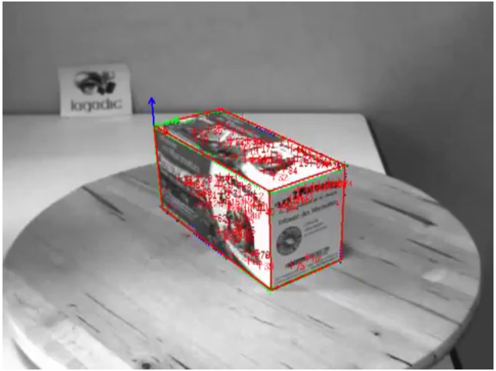
Keypoint matching



Hybrid tracking, ViSP [Com+06]

- **Model**: 3D point with descriptors
- **Method**: 3D-2D matching, minimize reprojection error (PnP problem)

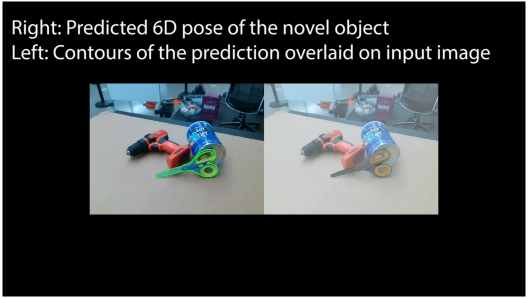
Keypoint matching



Hybrid tracking, ViSP [Com+06]

- **Model**: 3D point with descriptors
- **Method**: 3D-2D matching, minimize reprojection error (PnP problem)
- Efficient and robust if rich texture
- Fails for object with low texture

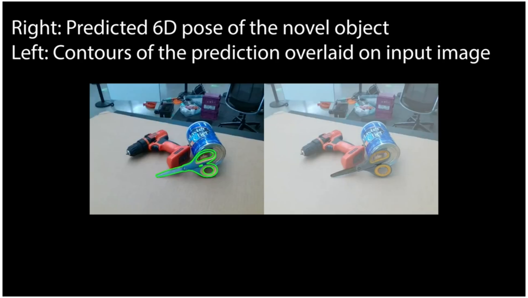
Deep learning



Megapose, tracking mode (2022) [Lab+22]
Also: PoseRBPF [Den+21],
se(3)-TrackNet [Wen+20]...

- **Model**: textured mesh
- **Method**: render and compare, regress delta pose

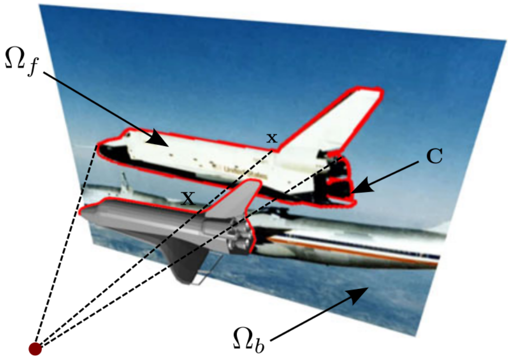
Deep learning



Megapose, tracking mode (2022) [Lab+22]
Also: PoseRBPF [Den+21],
se(3)-TrackNet [Wen+20]...

- **Model**: textured mesh
- **Method**: render and compare, regress delta pose
- Robust to occlusions, clutter, etc. Sota on standard benchmarks
- High-end GPUs at run-time, costly training, generalization ($\sim$)

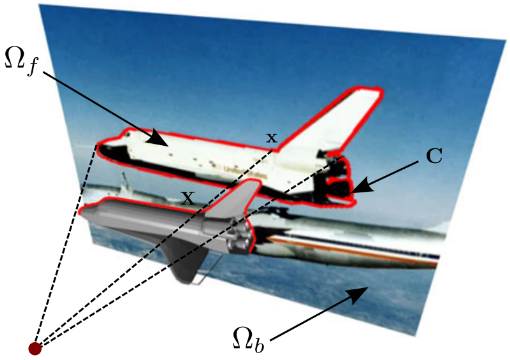
Region based tracking



PWP3D [PR12]

- **Model**: mesh (no texture)
- **Method**: probabilistic silhouette alignment, Newton's method

Region based tracking



PWP3D [PR12]

- **Model**: mesh (no texture)
- **Method**: probabilistic silhouette alignment, Newton's method
- Robust to occlusions, clutter, very efficient (1 object $\rightarrow$ $\sim$1000 FPS on CPU)
- Assumes foreground and background colors sufficiently different

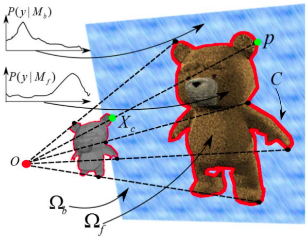
Region based tracking

SRT3D, sparse region based tracking [Sto+22]

Dense region based tracking for pose tracking

Objective: find ${}^{c}\mathrm{T}_b$ that maximizes likelihood of segmentation

$$P({}^{c}\mathrm{T}_b|\mathrm{Img}) = \prod_{x\in\Omega} \left(h_b(\phi)\cdot P_b + h_f(\phi)\cdot P_f\right)$$



Foreground/background probability distributions [Zha+14]

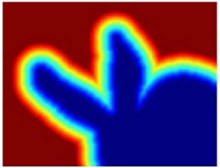
Signed Distance Function (SDF)

Objective: find ${}^{c}\mathrm{T}_b$ that maximizes likelihood of segmentation

$$P({}^{c}\mathrm{T}_b|\mathrm{Img}) = \prod_{x\in\Omega}(h_b(\phi)\cdot P_b + h_f(\phi)\cdot P_f)$$

- $\phi = f({}^{c}\mathrm{T}_b)$: SDF, from rendered contour

Contour from ${}^{c}\mathrm{T}_b$



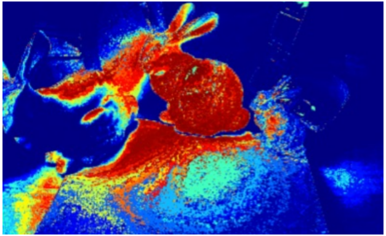
SDF

Color statistics and activation

Objective: find ${}^{c}\mathrm{T}_b$ that maximizes likelihood of segmentation

$$P({}^{c}\mathrm{T}_b|\mathrm{Img}) = \prod_{x\in\Omega}(h_b(\phi)\cdot P_b + h_f(\phi)\cdot P_f)$$

- P_b, P_f: background/foreground color distributions
- h_b, h_f: background/foreground activation functions



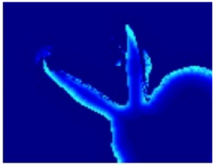
Example of P_f visualization [Keh+17]

ExecuteTrackingStep

$^{c}\mathrm{T}_b = {}^{c}\mathrm{T}_b^0$
$P_b, P_f = P_b^0, P_f^0$
for $i = 1$ to N_update_stats **do**
  for $j = 1$ to N_newton **do**
    $\mathrm{cost}({}^{c}\mathrm{T}_b) = -\log P({}^{c}\mathrm{T}_b | \mathrm{Img})$
    $\mathrm{g}, \mathrm{H} = \mathrm{ComputeCostGradientHessian}({}^{c}\mathrm{T}_b, P_b, P_f)$
    $\nu_b = \text{-}(\mathrm{H} + \lambda_{tikho} \mathrm{I}_6)^{-1} \cdot \mathrm{g}$
    $^{c}\mathrm{T}_b = \mathrm{UpdatePose}({}^{c}\mathrm{T}_b, \nu_b)$
  end for
  $P_b, P_f = \mathrm{UpdateColorStatistics}({}^{c}\mathrm{T}_b)$
end for

Are dense computations necessary?

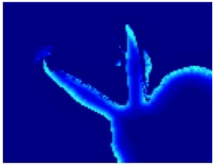
Contour prediction



Residuals $-\log P({}^{c}\mathrm{T}_{b}|\mathrm{Img})$

Are dense computations necessary?

Contour prediction



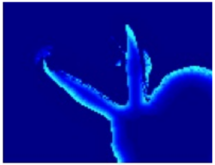
Residuals $-\log P({}^{c}\mathrm{T}_{b}|\mathrm{Img})$

Observations:

- Important residuals only close to predicted contour
- Neighbor contour points produce similar gradients
- Dense SDF computation is expensive (Repeated rendering and Direct transform)

Are dense computations necessary?

Contour prediction



Residuals $-\log P({}^{c}T_{b}|\mathrm{Img})$

Observations:

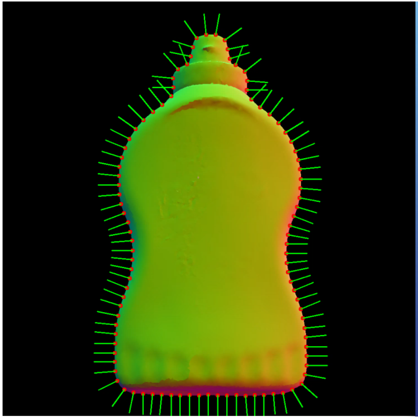
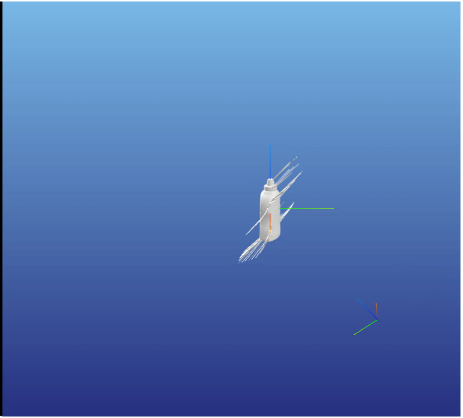
- Important residuals only close to predicted contour
- Neighbor contour points produce similar gradients
- Dense SDF computation is expensive (Repeated rendering and Direct transform)

Sparse Region based method [Keh+17]

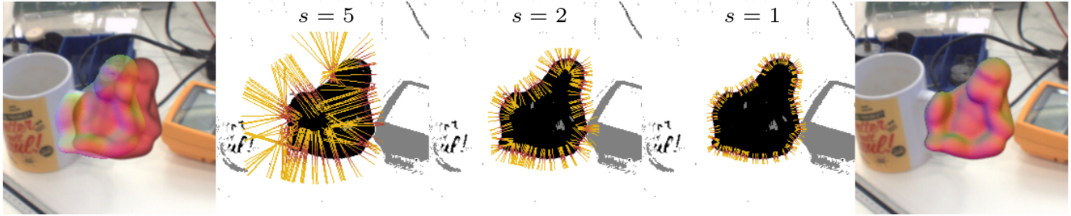
- Idea1: Sample contour control points
- Idea2: Precomputation of template views

Sparse view precomputations

Typically by using a geodesic polyhedron (e.g. 2562 views)

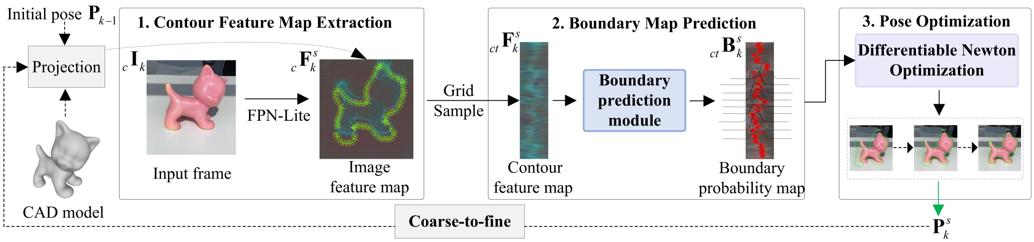



Correspondance lines reformulation



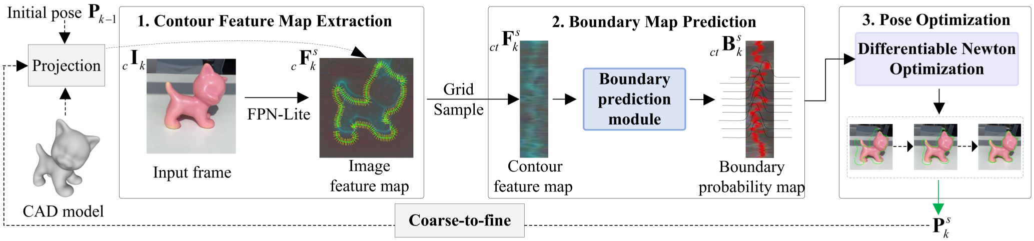
Correspondance lines, coarse to fine iterations [Sto+20]

Hybrid learning + optimization region based tracking



Deep Active Contour for Real-time 6-DoF Object Tracking [Wan+23]

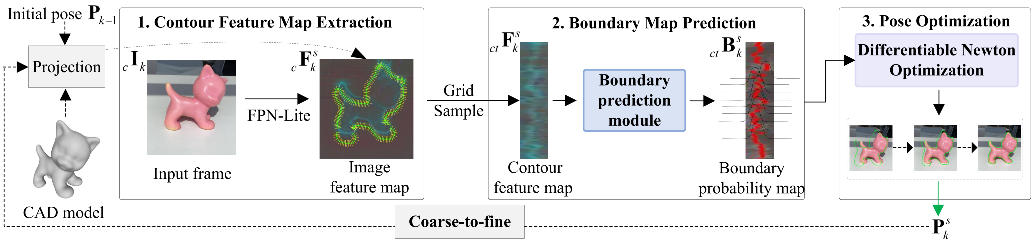
Hybrid learning + optimization region based tracking



Deep Active Contour for Real-time 6-DoF Object Tracking [Wan+23]

- Replace histograms by learning contour probability prediction

Hybrid learning + optimization region based tracking



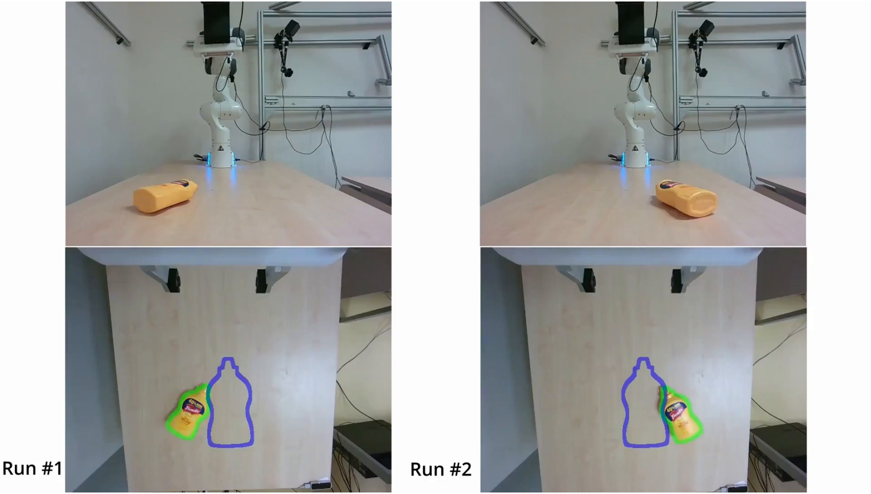
Deep Active Contour for Real-time 6-DoF Object Tracking [Wan+23]

- Replace histograms by learning contour probability prediction
- Trained end to end with differentiable optimization

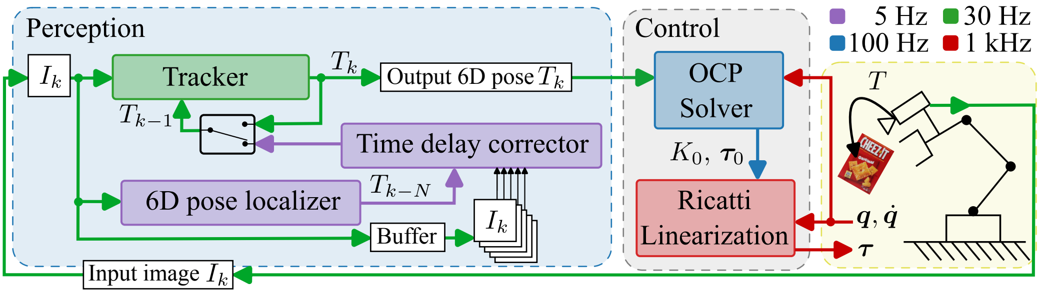
Object localization and tracking

An architecture for vision-based feedback control

Object tracking with manipulator

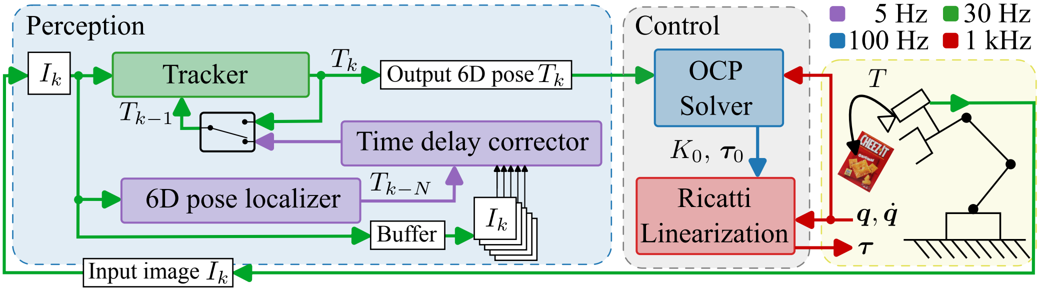


System architecture



Object localization and tracking architecture [Fou+23]

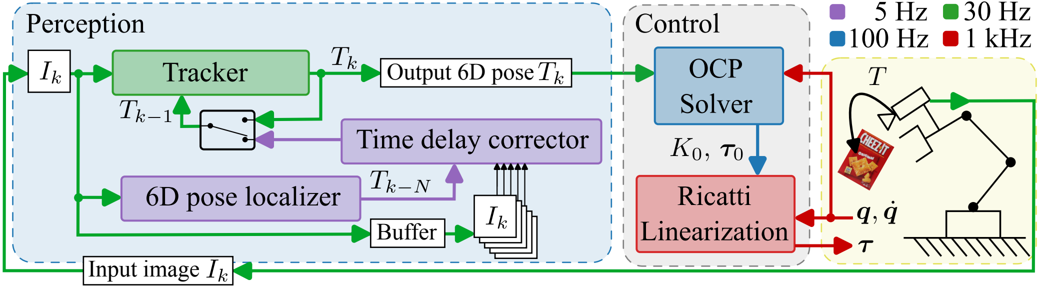
System architecture



Object localization and tracking architecture [Fou+23]

- Asynchronous object localization and tracking

System architecture



Object localization and tracking architecture [Fou+23]

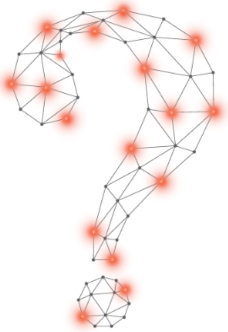
- Asynchronous object localization and tracking
- Torque level MPC (crocoddyl) with Riccatti based feedback

Practical session

Practical session

- Pose detection
  - 2D detection
  - CosyPose
  - Megapose
- Pose tracking
  - Recorded sequences
  - Webcam

Questions and Answers



Contact details

Médéric Fourmy
mederic.fourmy@cvut.cz

References

Ben Burgess-Limerick et al. "An architecture for reactive mobile manipulation on-the-move". In: *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE. 2023, pp. 1623–1629.

Andrew I Comport et al. "Real-time markerless tracking for augmented reality: the virtual visual servoing framework". In: *IEEE Transactions on visualization and computer graphics* 12.4 (2006), pp. 615–628.

Xinke Deng et al. "PoseRBPF: A Rao–Blackwellized particle filter for 6-D object pose tracking". In: *IEEE Transactions on Robotics* 37.5 (2021), pp. 1328–1342.

Mederic Fourmy et al. *Visually Guided Model Predictive Robot Control via 6D Object Pose Localization and Tracking*. 2023. arXiv: `2311.05344 [cs.RO]`.

Chris Harris and Carl Stennett. "RAPID-a video rate object tracker.". In: *BMVC*. 1990, pp. 1–6.

References (cont.)

- Wadim Kehl et al. "Real-time 3D model tracking in color and depth on a single CPU core". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017, pp. 745–753.
- Yann Labbé et al. "Megapose: 6d pose estimation of novel objects via render & compare". In: *arXiv preprint arXiv:2212.06870* (2022).
- Seyed Sina Mirrazavi Salehian, Nadia Figueroa, and Aude Billard. "A Unified Framework for Coordinated Multi-Arm Motion Planning". In: *The International Journal of Robotics Research* 37.10 (2018), pp. 1205–1232. DOI: 10.1177/0278364918765952. eprint: https://doi.org/10.1177/0278364918765952. URL: https://doi.org/10.1177/0278364918765952.
- Victor A Prisacariu and Ian D Reid. "PWP3D: Real-time segmentation and tracking of 3D objects". In: *International journal of computer vision* 98 (2012), pp. 335–354.

References (cont.)

Manuel Stoiber et al. "A sparse gaussian approach to region-based 6DoF object tracking". In: *Proceedings of the Asian Conference on Computer Vision*. 2020.

Manuel Stoiber et al. "SRT3D: A sparse region-based 3D object tracking approach for the real world". In: *International Journal of Computer Vision* 130.4 (2022), pp. 1008–1030.

Long Wang et al. "Deep Active Contours for Real-time 6-DoF Object Tracking". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2023, pp. 14034–14044.

Bowen Wen et al. "se (3)-tracknet: Data-driven 6d pose tracking by calibrating image residuals in synthetic domains". In: *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE. 2020, pp. 10367–10373.

References (cont.)

Wei Yang et al. "Model predictive control for fluid human-to-robot handovers". In: *2022 International Conference on Robotics and Automation (ICRA)*. IEEE. 2022, pp. 6956–6962.

Song Zhao et al. "3D object tracking via boundary constrained region-based model". In: *2014 IEEE International Conference on Image Processing (ICIP)*. IEEE. 2014, pp. 486–490.