# Online Interactive Perception of Articulated Objects with Multi-Level Recursive Estimation Based on Task-Specific Priors

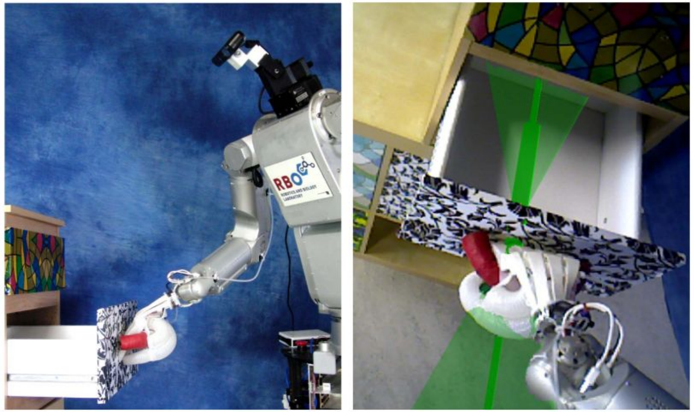
Presenter: Gu-Cheol Jeong

09/15/2022

# Motivation : manipulate in unknown environment

**To perform successful manipulation:**

1. **Perceiving degree of freedom of unknown objects in unstructured environment**
   **- Joint type, and joint axis**



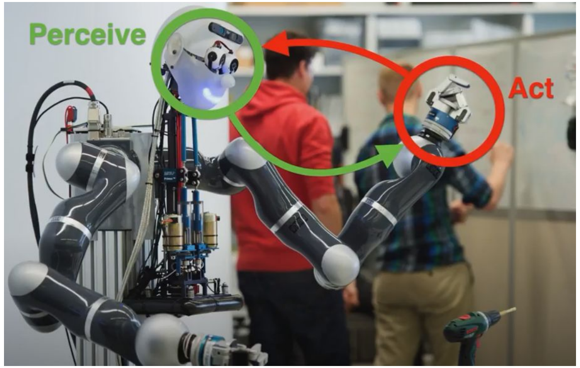
2. **Online interactive perception (IP) algorithm**

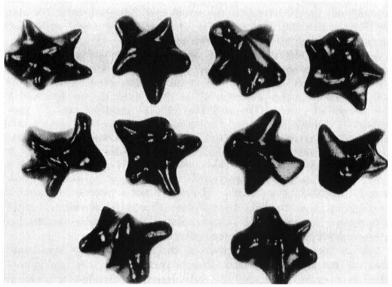
# Related work - what is interactive perception?

**Idea that perception and interaction are inextricably linked**

**= Leverage action in perception and perception in action**

- **Perception is facilitated by interaction with the environment**
- **Physical interaction -- create → rich sensory signal**
- **Regularity in combined space of sensory data and action parameters facilitates the prediction and interpretation of the signal**



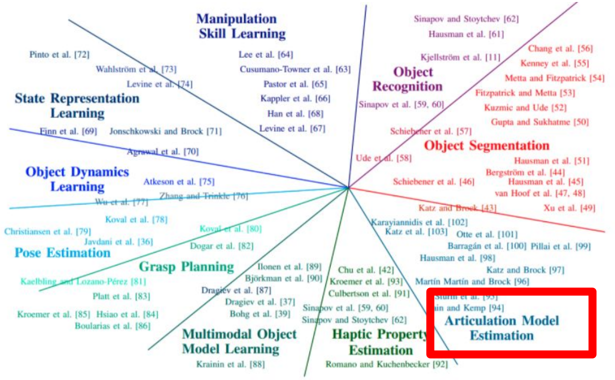



Fig. 4. Paper categorization based on application areas. Papers that address multiple application areas lie on the boundary between those applications, e.g. Kaelbling and Lozano-Pérez [81], Schiebener et al. [57]

# Motivation and Main Problem

**Online IP method is important because**

1. **Existing IP method are offline algorithm**
   **→ not inform the ongoing action**

2. **Offline setting leads to failure cases**
   **→ online method can address**

3. **Offline methods are not probabilistic**
   **→ not include an estimate of model uncertainty**

# Related work : Articulated Model Estimation

1. Yan and Pollefeys : segmentation + joint detection
   - use structure from motion -> estimate 3D feature trajectory
   - spectral clustering -> identify rigid bodies

2. Ross et al. : Yan work + maximum likelihood estimation to identify rigid bodies
   Cons : Need accumulated large motion data + offline algorithm

3. Sturm et al.: a probabilistic approach to joint classification and characterization
   Pros: uncertainty estimates about joint type + real-time.
   Cons: x detecting and tracking unknown rigid body.(need to know rigid bodies)

4. Huang et al.: offline method to extract 3-D models of articulated rigid object using IP
   Cons: require multiple object views to generate first full point cloud of an objects

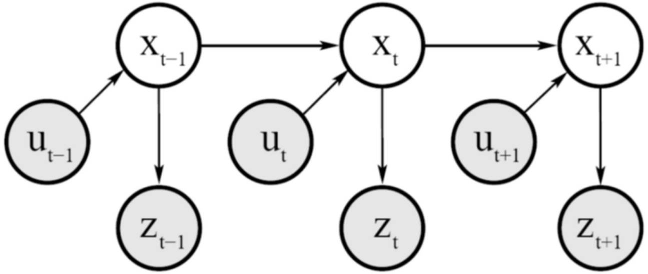
5. Katz et al. [14]: RGB-based, offline solution for the perception of three-dimensional, rigid kinematic structure.
   Pros: cost-effective + accurate + computationally efficient
   Cons: offline. ← limitation(newly appeared objects)

# Recursive Estimation - Bayesian Filter

When state and observation are stochastic processes, recursive estimation can be solved by using recursive Bayesian filtering.

Filter estimate : posterior $p(\underbrace{\boxed{x_t}}_{\text{State}} | \underbrace{\boxed{z_{1:t}}}_{\text{Measurement}}, \underbrace{\boxed{u_{1:t}}}_{\text{Action}})$

based on : $p(x_{t-1} | z_{1:t-1}, u_{1:t-1})$



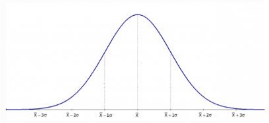

Integration for the prediction

=> increase the complexity and computational load

Solution

1. Monte Carlo integration = Particle Filter
2. Gaussian distribution = Kalman Filter

**Algorithm Bayes_filter**($bel(x_{t-1}), u_t, z_t$):
  for all $x_t$ do
    $\overline{bel}(x_t) = \int p(x_t \mid u_t, x_{t-1})\, bel(x_{t-1})\, dx$
    $bel(x_t) = \eta\, p(z_t \mid x_t)\, \overline{bel}(x_t)$
  endfor
  return $bel(x_t)$



$$p(x) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{(x-\bar{X})^2}{2\sigma^2}\right)$$

# Recursive State Estimation - Kalman Filter

- Process and measurement model : linear functions with Gaussian noise
- Initial State is normally distributed,

  ---> Kalman Filter

$$\boxed{x_t} = g(x_{t-1}, \boxed{u_t}) + \boxed{\varepsilon_t}$$

State

Action

Gaussian Noise

Measurement model

$$\boxed{z_t} = h(x_t) + \delta_t$$

- Process and measurement model : Nonlinear function but linearizable (Taylor series or Jacobian)

  ---> Extended Kalman Filter(EKF)

$$f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \frac{f'''(a)}{3!}(x-a)^3 + \cdots.$$

Figure 5. Taylor Series

**EKF is used in this paper for estimation**

# Recursive Estimation of Feature Motion

first level : motion of a set of point features in an RGB-D sensor

state : $x_t^{fm} \in \mathbb{R}^{3N}$ $(fm = \text{feature motion})$

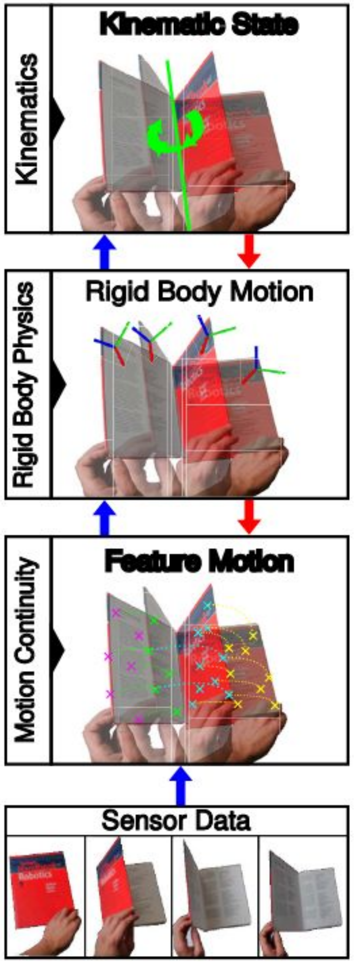
N : tracked point feature

3N = 3D coordinate * tracked point feature

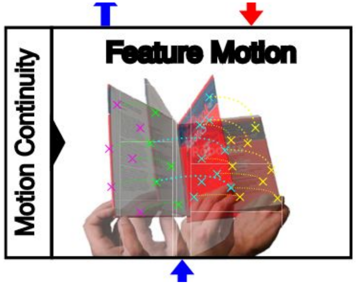
measurement: $z_t^{fm} \in \mathbb{R}^{2N}$

measurements are obtained by tracking the features in the RGB sensor

Two priors: 1) motion continuity, 2) physics of rigid bodies



# Recursive Estimation of Feature Motion




1) **Prediction in feature motion estimation**

   **current location --- predict → next feature location**

   **current location --- predict → estimated rigid body velocities**

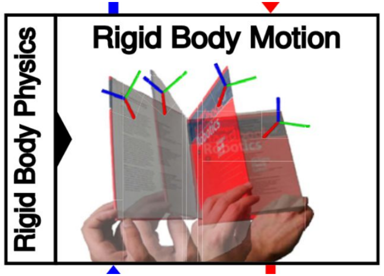
2) **Measurement update in feature motion estimation**

   **predicted 3D feature location --- project ---> image plane(2D) ---> Kanade-Lucas_Tomasi tracking algorithm(refine 2D locations, update the filter state) ---> estimates of the next feature location**

3) **Feature initialization and maintenance**

   **reject the features a) features over the course of 15 frames, b) features liying close to depth continuities in the RGB-D image, c) surface of the robot arm.**

# Recursive Bayesian Estimation of Rigid Body Motion




1) **Prediction in Single Rigid Body Motion Estimation**

   **Use three different process models in parallel**

   **a) current pose + velocity -- predict → next pose**

   **b) special case : stop movings -- set → current vel = 0**

   **c) kinematic model -- predict → alternative next pose & velocity**

2) **Measurement Update in Single Rigid Body Motion Estimation**

   **input : 3D feature location(lower level), output : future locations of feature**

$$z_t^{rbm} = h(\boldsymbol{x}_t^{rbm}) + \delta_t^{rbm} = \begin{pmatrix} T(\boldsymbol{p})\boldsymbol{f}_{init}^{1} \\ T(\boldsymbol{p})\boldsymbol{f}_{init}^{2} \\ \vdots \\ T(\boldsymbol{p})\boldsymbol{f}_{init}^{M} \end{pmatrix} + \delta_t^{rbm}, \qquad (1)$$

3) **Recursive Bayesian Estimation of Multi-Body Motion**

   **group features,**

   **Cannot predict features(error<=2cm) ---RANSAC → find a rigid body transformation**

   **f_min > 15 features, create a new RBF**

# Recursive Estimation of Kinematic Model

**motion of rigid body -- estimate and tracks → the kinematic model of the scene -- prediction → next state of the next-lower level(feedback)**

**Assume four states**

**1) Prismatic joint**

$$z_t^{joint,p} = \begin{pmatrix} q^p \cdot \hat{\boldsymbol{o}}^p \\ \mathbf{0}_3 \end{pmatrix} + \delta_t^{joint}$$

**2) Revolute joint**

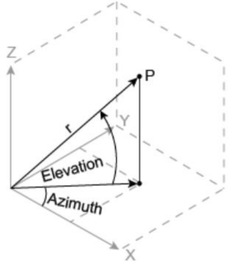
$$z_t^{joint,r} = \begin{pmatrix} \boldsymbol{t}^r \\ q^r \cdot \hat{\boldsymbol{o}}^r \end{pmatrix} + \delta_t^{joint}$$

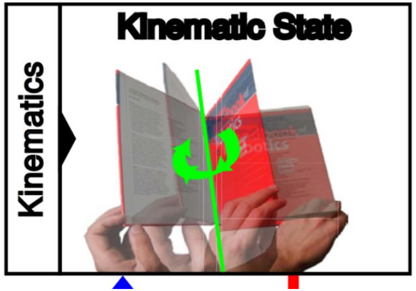
**3) Rigid connection**

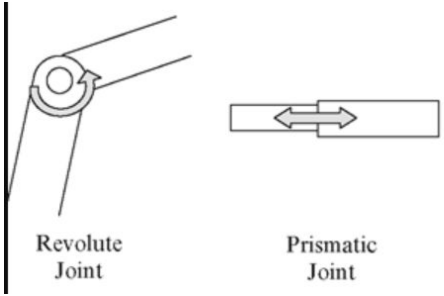
**: No relative motion**

$$z_t^{joint,rigid} = \mathbf{0}_6 + \delta_t^{joint}$$

**4) Disconnected**



# Experimental Setup

- Evaluate the performance of the online IP algorithm.

  : Place artificial markers for the ground truth of the joint

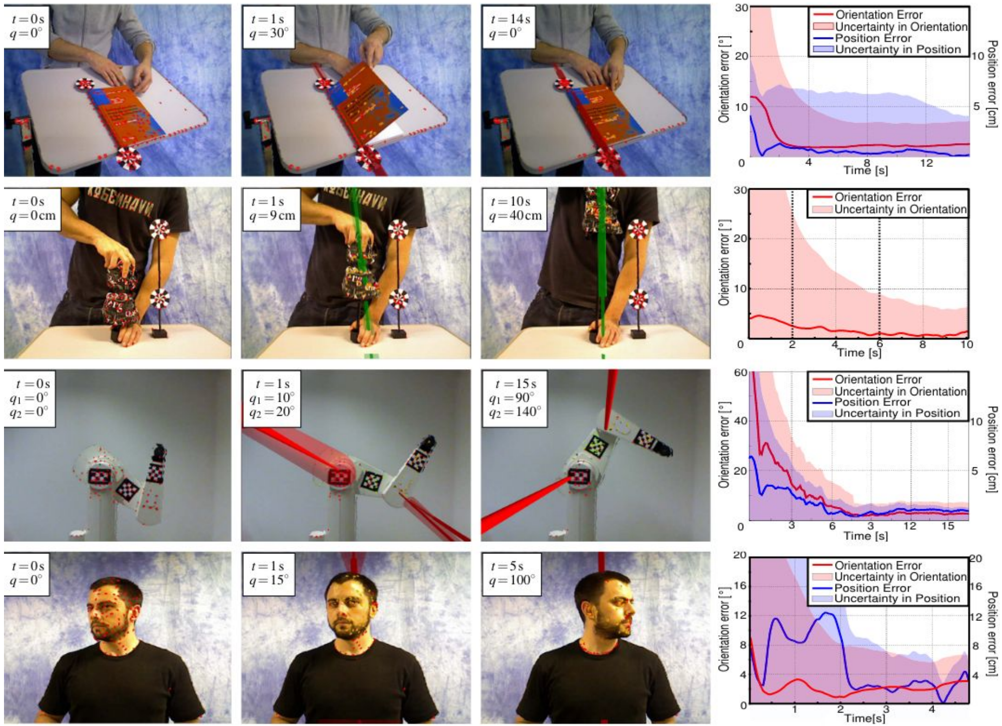
- Test the algorithm in scenarios were offline algorithms fail

- Use of the online abilities of the algorithm to control the motion of a robot

- Algorithm tracks N = 150 features at a frame rate of 15 frames / sec

# Experimental Evaluation

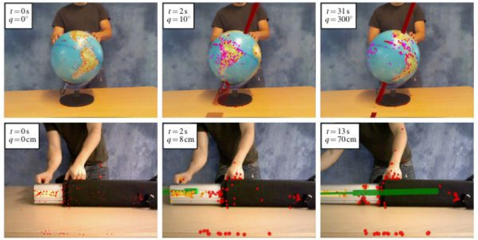
- ❖ Book experiment
- ❖ Umbrella experiment
- ❖ PUMA 560 experiment
- ❖ Human Head experiment



# Failure case of Previous Offline Algorithms Solved with Online IP

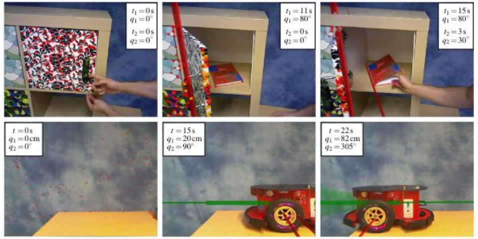
❖ Disappearing Features

- detects moving bodies
- incrementally assigns new features



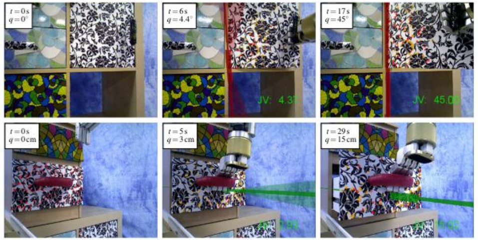
❖ Appearing Objects

- not be shown at the beginning.
- once object is visible, new features are detected



❖ Identical initial and final configuration

- model remains converged after the object returns



# Controlling Interaction with Online IP

- Two experiments with two objects each, Repeated ten times
- Goal:

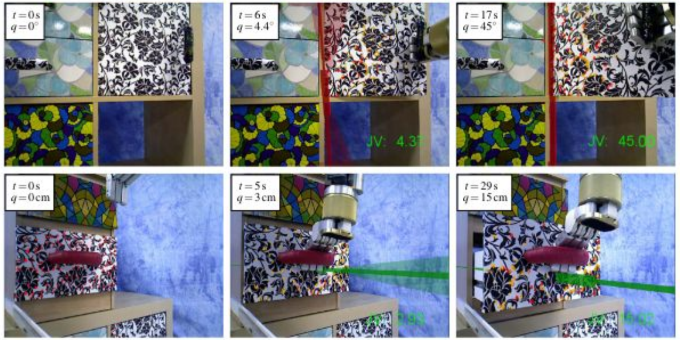
  task 1) obtain a kinematic model with a specified uncertainty

  task 2) move one of the joints to a specific configuration
- Result

  1) online IP can be used to monitor and control interactions

  2) possible to adjust the robot's action based on a desired uncertainty bound for the accuracy during the estimation of kinematic model.

  3) Online estimation of joint values can be used to monitor and attain manipulation goal



# Critique / Limitations / Open Issues

1. **There must be motion for the method to work**
2. **the limitations of feature tracking relatively stable lighting conditions and bounded object acceleration.**
3. **Computational limitation**

   **- algorithm must perform at reasonably high frame rates**

   **- track 150 features at 15 Hz**
4. **Only handles four kinematic relationships**

   **- revolute, prismatic, rigidly connected, and disconnected.**

# Extended Readings

[1] H. Li, “A Brief Tutorial On Recursive Estimation With Examples From Intelligent Vehicle Applications (Part I): Basic Spirit And Utilities,” p. 20.

[2]J. Bohg *et al.*, “Interactive Perception: Leveraging Action in Perception and Perception in Action,” *IEEE Trans. Robot.*, vol. 33, no. 6, pp. 1273–1291, Dec. 2017, doi: 10.1109/TRO.2017.2721939.

# Summary

- ❖ Present an online algorithm for the interactive perception of articulated bodies
- ❖ Input : RGB-D stream

  Output : kinematic model of observed scene
- ❖ Three interconnected recursive estimation filters to solve the perceptual problem

  1) Feature motion

  2) Rigid motion

  3) Kinematic model of moving objects
- ❖ Recursive estimation filters are bidirectional
- ❖ Result : Highly robust algorithm