Manipulator and Object Tracking for In Hand Model Acquisition

Michael Krainin, Peter Henry, Xiaofeng Ren, and Dieter Fox

Click to edit Master subtitle style

University of Washington and Intel Labs Seattle

May 7, 2010

Problem Statement

- Robots can never be trained for all possible objects – let them learn new ones
- Target Scenario:

1) Robot picks up unknown object using heuristic or feature-based techniques (e.g. [Saxena ‘08])

2) Robot examines object from multiple perspectives, much as a human might

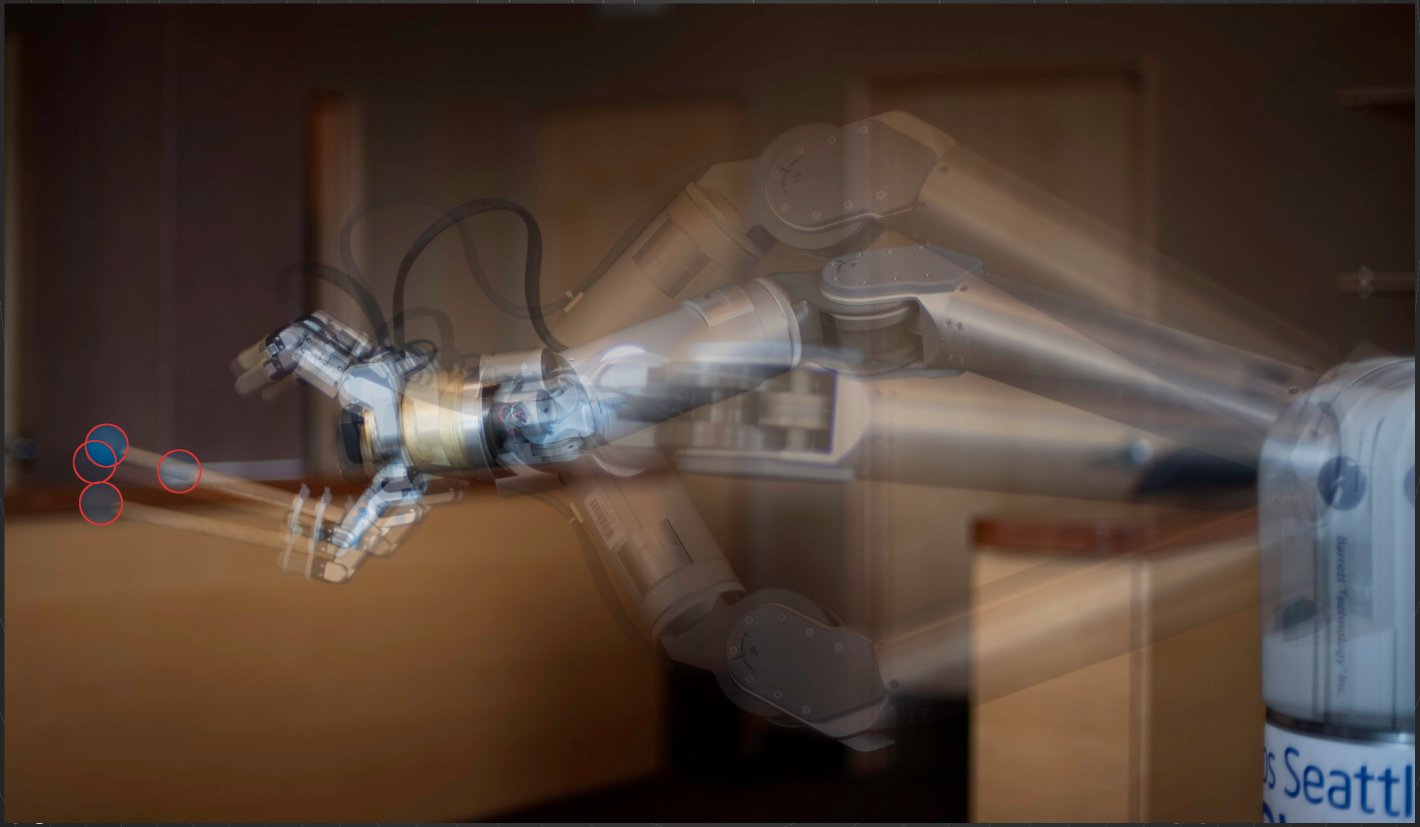
3) Robot can use learned model for later tasks (e.g. detection, pose estimation, future grasping)

In-Hand Modeling

- Visual feature matching [Pan '09]
- ICP-based surface matching [Weise '09]
- No explicit tracking of hand
  - Hand subtraction not straightforward
  - Cannot register textureless or symmetric objects

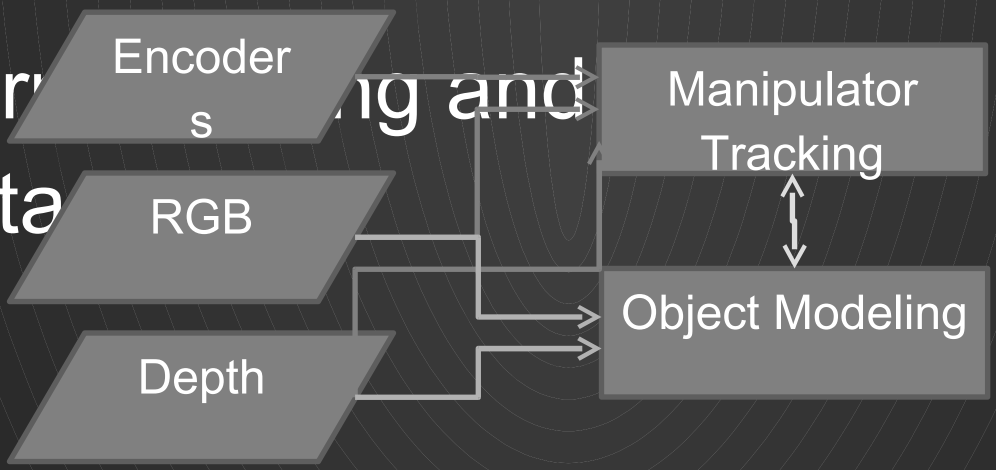
Encoders for Object Modeling

- Commonly used but requires high accuracy e.g. [Sato '97], [Kraft '08]



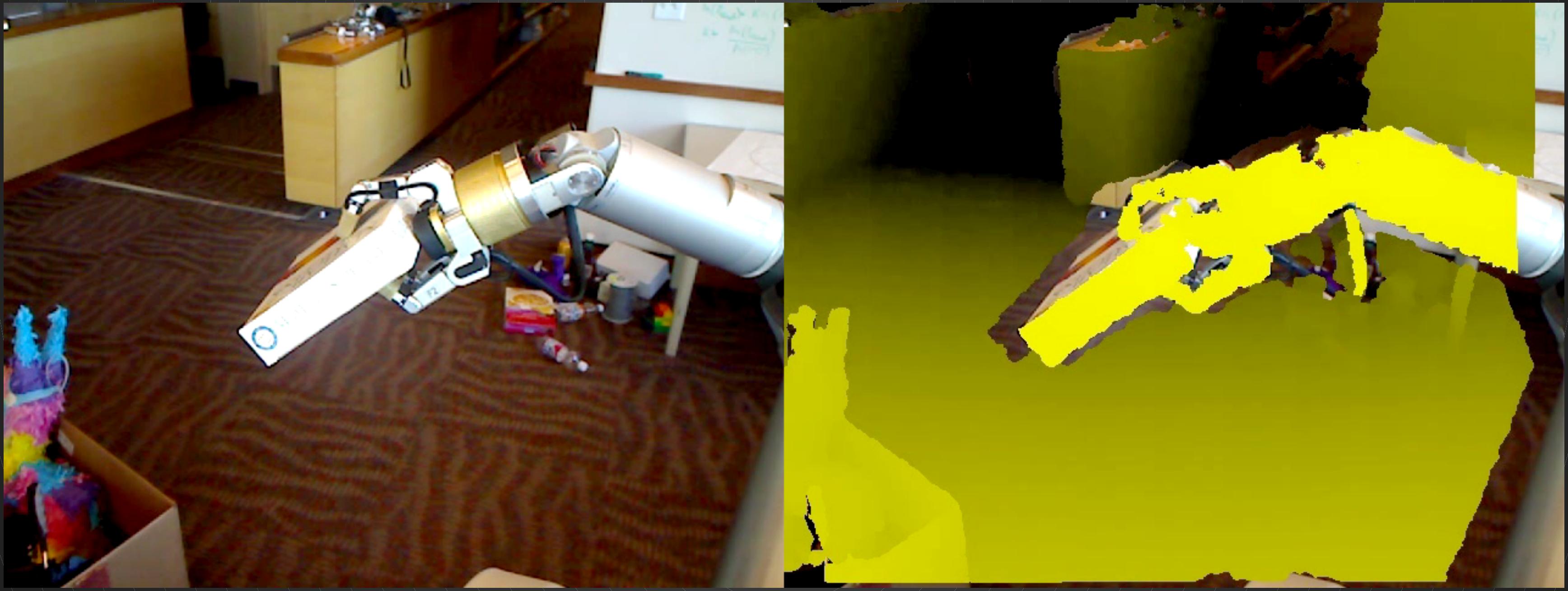
Proposed Framework

- Use encoders only to the extent to which they are reliable
- Use vision and shape to improve alignment
- Perfor ng and simulta



Depth Sensor

- 640x480 RGB + Depth at 30 Hz
- Developed for gaming applications



Articulated ICP for Robot Arm and Hand Tracking

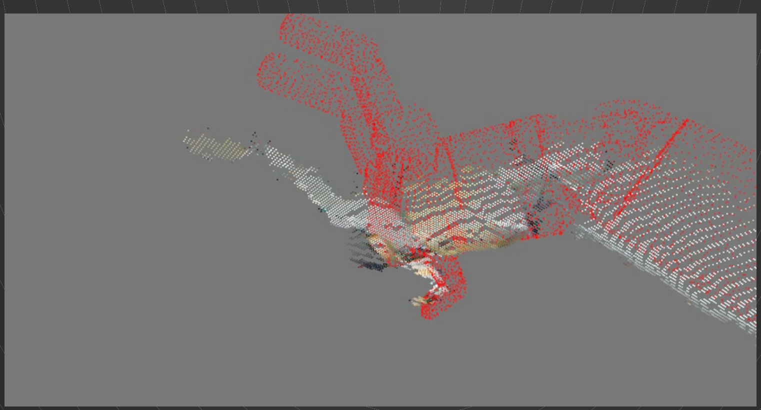
Joint Hand and Object Tracking

Object Modeling

Articulated ICP

- Goal: Find joint angles (and global transform) best aligning observed and expected clouds
- Iterates between correspondence selection and parameter estimation

$$E_{pt-to-plane}(T) = \sum_{i=1}^{M} \omega_i [(T(x_s^i) - x_d^i) \cdot n_d^i]^2$$



Integration Into Kalman Filter

- State: Joint angles + global transformation

- Encoders for motion update

$$\bar{\mu}_k = m_{k-1} + \mathbf{B}(\tilde{q}_k - \tilde{q}_{k-1})$$

- Kalman covariance fed back into ICP to give preference to adjusting less certain DoFs

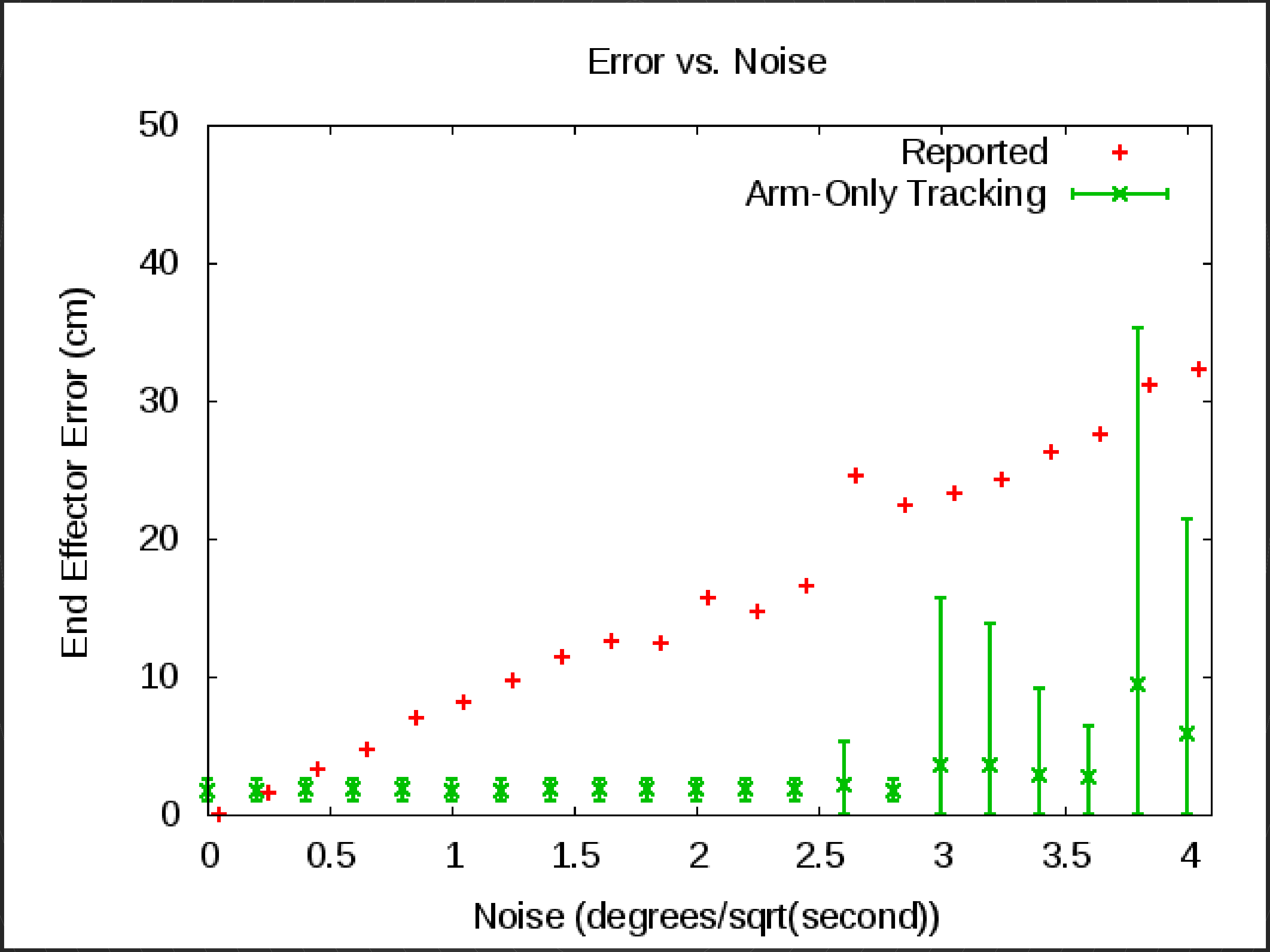
$$E_{prior}(T) = (T - \bar{\mu}_k)\bar{\Sigma}_k^{-1}(T - \bar{\mu}_k)$$

$$T^* = \arg\min_T [E_{pt-to-plane}(T) + E_{prior}(T)]$$

$$\mu_k$$

Tracking with Added Noise

Tracking Results



Articulated ICP for Robot Arm and Hand Tracking

Joint Hand and Object Tracking

Object Modeling

Joint Tracking

- Motivation
  - Need accurate pose of object within hand
  - Object may occlude hand, making tracking more difficult

- Add pose of object within hand to Kalman state
- Add object matching into ICP

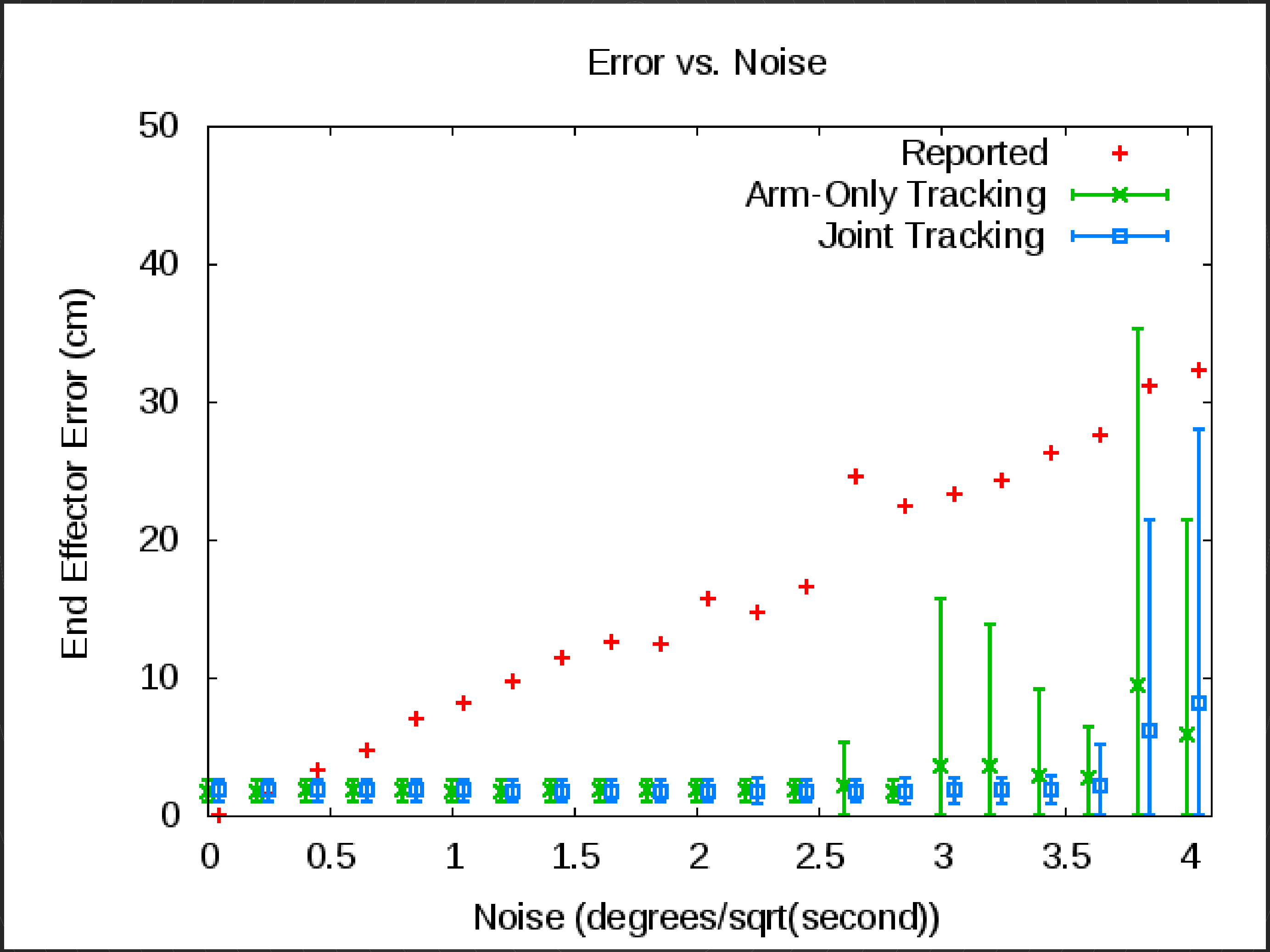
SIFT matching

- Use object texture for additional constraints
- Keep estimates of 3D locations of SIFT keypoints within object model
- Add geometrically consistent frame-to-model matches as point-to-point correspondences

$$E_{sift}(T) = \underset{(\kappa_\sigma, \kappa_\tau) \quad \Sigma}{} \; | T(\kappa_\sigma) - \kappa_\tau |^2$$

$$T^* = \underset{T}{\alpha\rho\gamma\,\mu\,\iota\nu}[E_{\pi\tau-\tau o-\pi\lambda\alpha\nu\varepsilon}(T) + E_{\pi\rho\iota o\rho}(T) + E_{\sigma\iota\phi\tau}(T)]$$

Joint Tracking Results



(Horizontal offset for readability only)

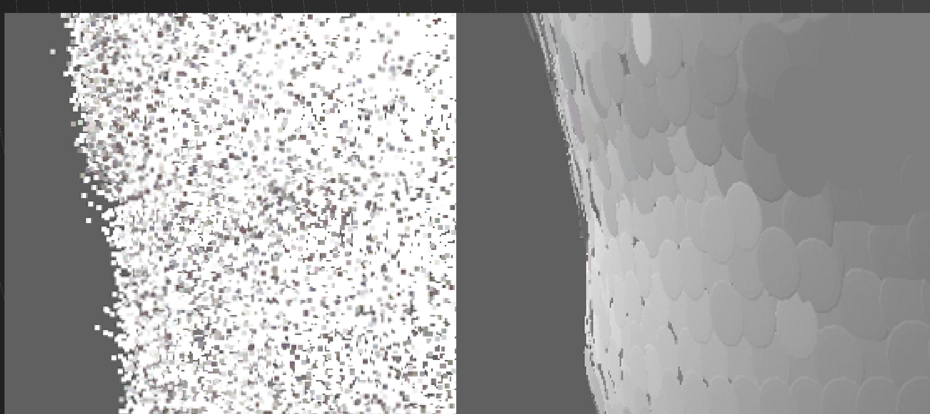
Articulated ICP for Robot Arm and Hand Tracking

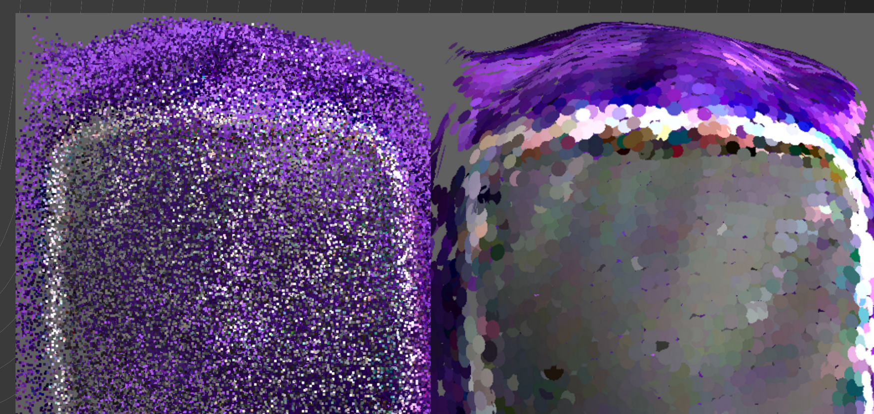
Joint Hand and Object Tracking

Object Modeling

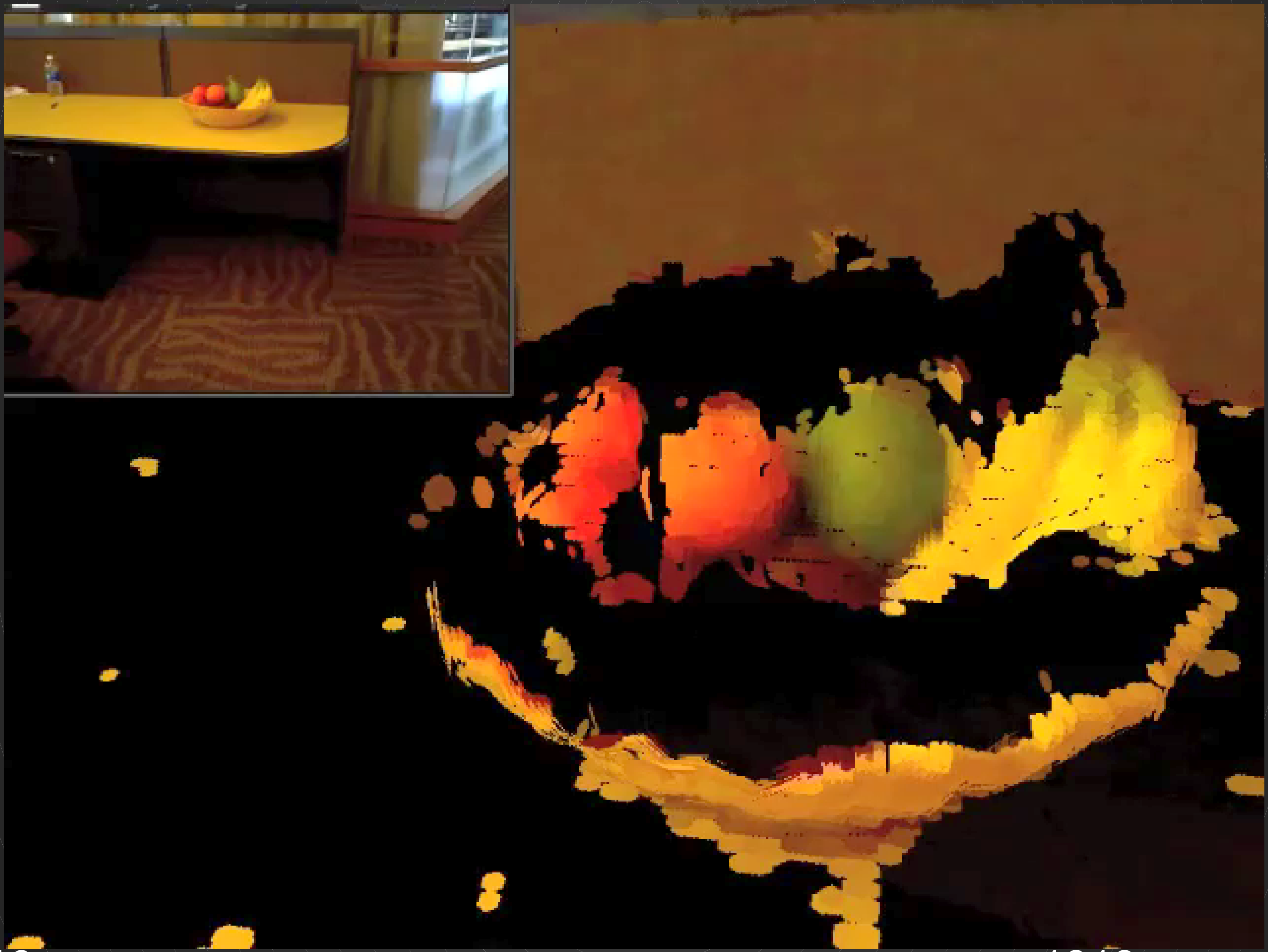
Surfel Representation

- "Surface Elements" – circular disks representing local surface patches
- Existing representation from the literature [Pfister '00], [Habbecke '07], [Weise '09]





Surfel Demonstration

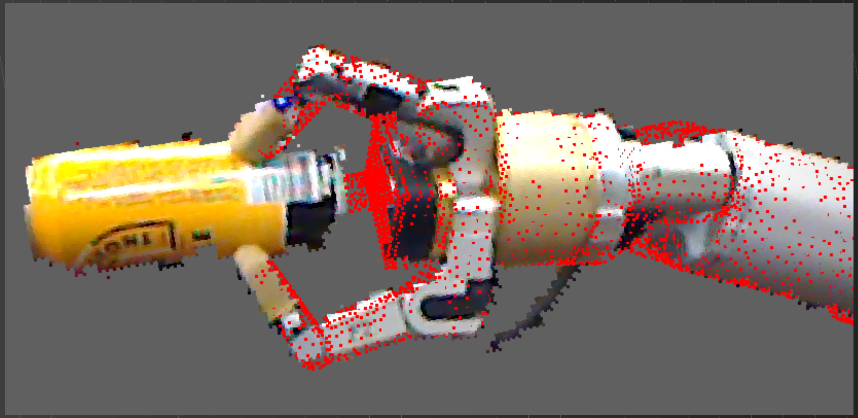


Advantages of Surfels

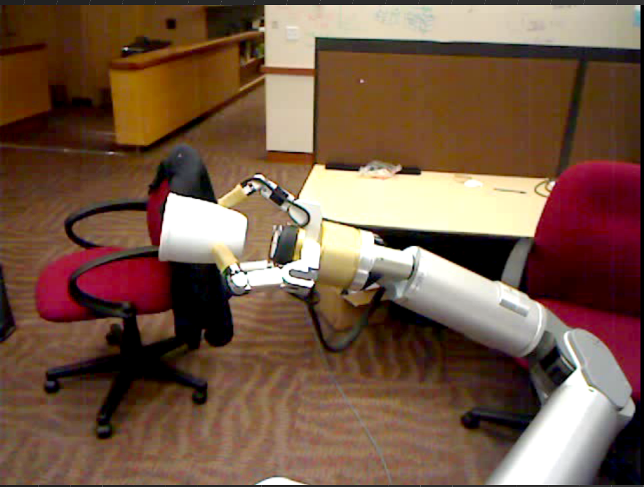
- Simple update rules
- Incremental updating
- Occlusion checking
- Efficient representation (scalable)
- Automatic resizing

Modeling

- Object detected based on tracking result
- Surfel model transformed to align with sensor data
- Surfels modified according to update rules



Tracking + Modeling



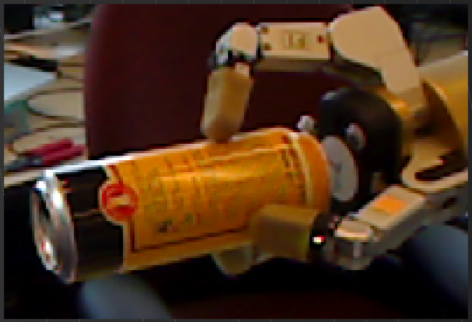
Modeling Properties

- Averaging of noise
- Can handle symmetric and textureless objects
- Limitations
  - Holes where gripper occludes object
  - Rigid objects only
  - Motion speed limited by motion blur and ICP

Modeling Results

Handling Multiple Grasps

- Switching Kalman filter
  - Examining object
  - Moving to or from table
  - Grasping or releasing
  - Between grasps
- Second grasp should be computed from



Summary

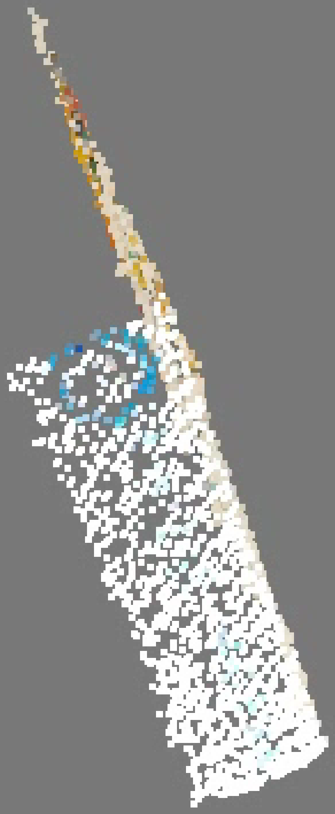
- Framework for simultaneously tracking robotic manipulator and modeling grasped objects
  - Incorporates information from encoders, vision, and depth
  - Joint estimation benefits tracking and modeling
  - Accurate models of symmetric and featureless objects

Future Work

- Automatic generation of motion sequences
- Self-supervised learning of objects and grasping strategies
- Development of a shared object database

Questions

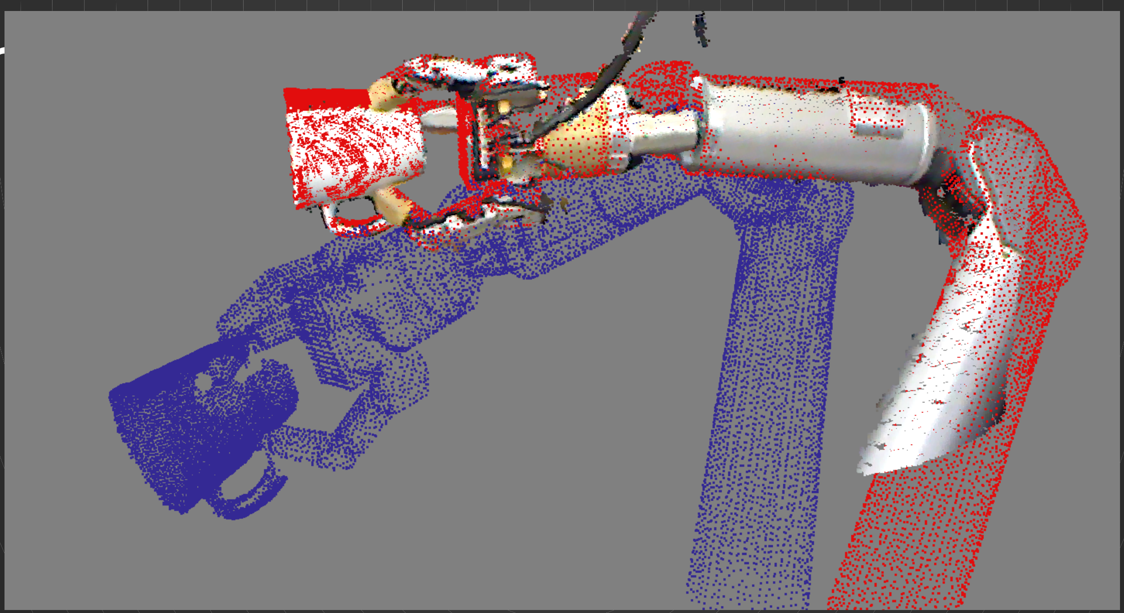
?



Approach Outline

Articulated Tracking

- Goal: Match a model of the arm (and manipulated object) into the sensor data
- Use shape, visual, and encoder inform



ICP Details

$$E_{pt-\varpi-\pi\lambda\alpha\nu\varepsilon}(T) = \mathop{}_{\iota=1}^{M} \omega_\iota \, \mu \, \iota\nu_{\varphi}((T(\pi_\sigma^\iota) - \pi_\tau^\varphi) \;\; \nu_\tau^\varphi)^2$$