3D Tracking of Hands Interacting with Several Objects

Nikolaos Kyriazis and Antonis Argyros
Institute of Computer Science,
FORTH, Greece

PROBLEM

Track, in 3D and in detail, the interaction in large scenes, comprising hands and many objects.

MOTIVATION

- State of the art 3D tracking methods can handle a single hand (27-D) [4], a hand and an object (36-D) [5] or two hands (54-D) [3].
- **What about several hands and many objects?**
  - Interactions in the form of occlusions and collisions.
  - Potential interactions increase quadratically with the number of entities.
  - The dimensionality of the problem increases linearly with the number of entities.

MAIN IDEA

- **Hypothesize-and-test** using **PSO**:
  - Data compatibility: **Computer Graphics**.
  - Physical plausibility: **Physics Simulation**.
- **Scalable optimization [1]**: **Exploit** the **structure** of the optimization problem to efficiently guide search.
- **Scalable modelling [2]**: **Exploit** the **causality** inherent to everyday hand(s)-object(s) interactions to effectively reduce the degrees of freedom of the problem..

REFERENCES

[1] N. Kyriazis and A. Argyros. *"Scalable 3d tracking of multiple interacting objects."* Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on. IEEE, 2014.
[2] N. Kyriazis and A. Argyros, *"Physically Plausible 3D Scene Tracking: The Single Actor Hypothesis,"* in CVPR, 2013.
[3] I. Oikonomidis, N. Kyriazis, and A.A. Argyros. *Tracking the articulated motion of two strongly interacting hands*. In CVPR, 2012.
[4] I. Oikonomidis, N. Kyriazis, and A. Argyros. *Efficient model-based 3d tracking of hand articulations using Kinect*. In BMVC, 2011.
[5] I. Oikonomidis, N. Kyriazis, and A. Argyros, *"Full DOF tracking of a hand interacting with an object by modeling occlusions and physical constraints,"* in ICCV, 2011.

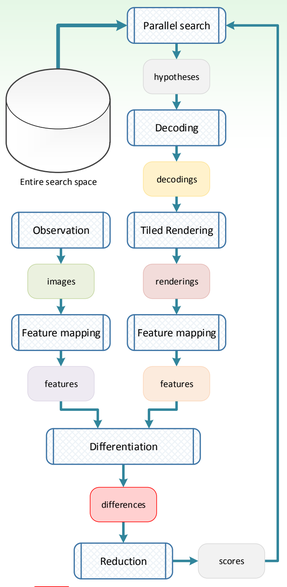
SCALABLE MODEL-BASED 3D TRACKING: MULTIPLE HANDS AND OBJECTS

Tracking framework

Objective function: $\mathbf{E}(x, o, h) = \|\mathbf{M}(x, h) - \mathbf{P}(o)\| + \lambda \mathbf{L}(x, h)$ **Optimization**: $s \triangleq \underset{x}{\operatorname{argmin}} \mathbf{E}(x, o, h)$

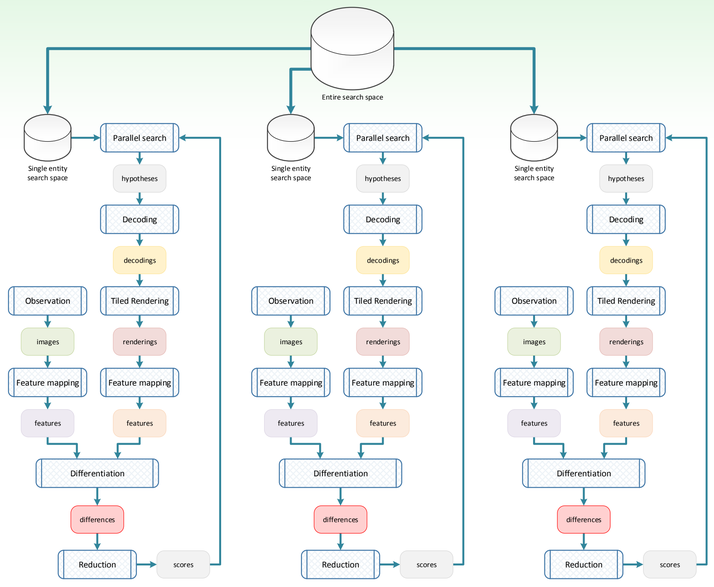
Scalable Optimization [1]

Joint Tracking



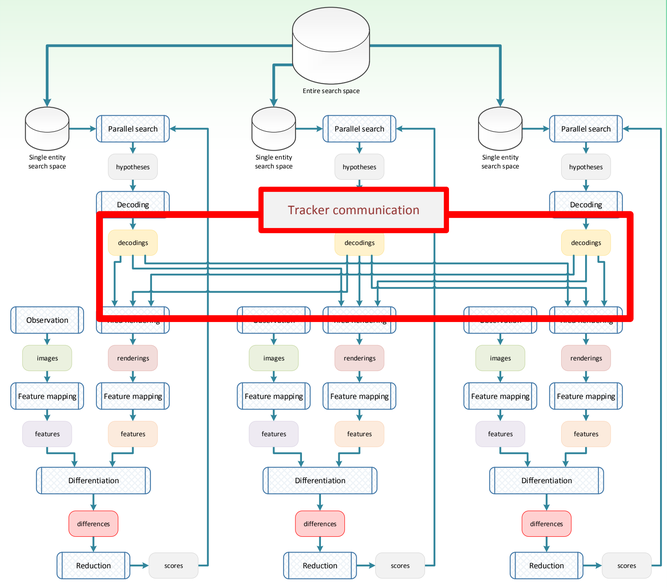
Scalability
Interaction

Set of Independent Trackers



Scalability
Interaction

Ensemble of Collaborative Trackers

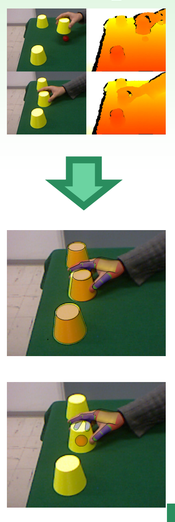


Scalability
Interaction

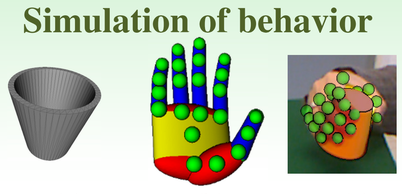
The implementations will be made available soon!

Scalable Modelling [2]

Concept



Simulation of behavior



(a) (b) (c)

For rigid objects the 3D model that approximates their appearance can also be used as a collision model (a).

The human hand is a complicated structure. Its collision model is abstracted towards a simple yet dexterous model (b).

Hypothesized hand motion is simulated to obtain interaction consequences on the state of the passive objects (c).

Simulation of appearance

(a) (b) (c)

The acquisition process yields RGB image sequences with corresponding depth maps (a).

3D rendering (b) is used to simulate the acquisition process, provided 3D models and the already simulated physical state.

The simulation can be of high degree of fidelity, as the synthetic result may be slightly different from its actual counterpart (c).

Simulation-based inference

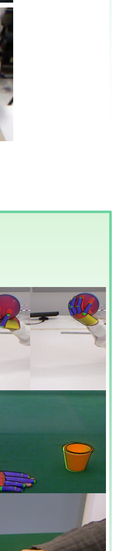
- Acquire input
- Define penalty function over hand motion

$$\mathbf{M}' = \mathbf{M}(\mathbf{S}(\mathbf{x}))$$

$$\mathbf{E}'(x, o, h) = \|\mathbf{M}'(x, h) - \mathbf{P}(o)\| + \lambda \mathbf{L}(x, h)$$

- Minimize through Particle Swarm Optimization
  - Until convergence
    - Hypothesize
    - Simulate
    - Compare
- Dub minimizer hand motion as the physically plausible explanation of input

$$s' \triangleq \underset{x}{\operatorname{argmin}} \mathbf{E}'(x, o, h)$$

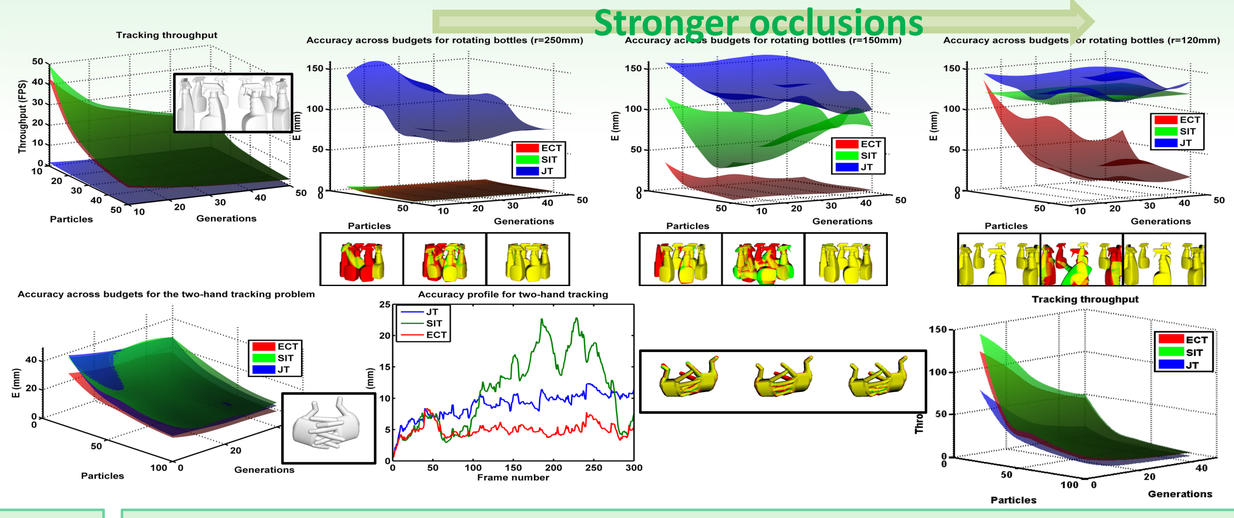


EXPERIMENTAL RESULTS

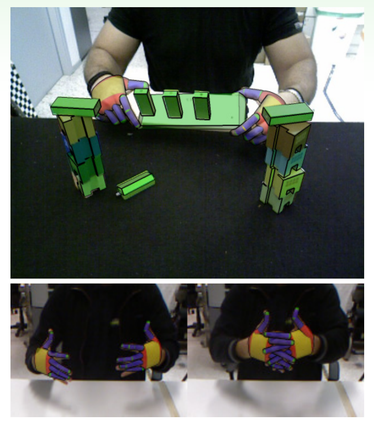
Quantitative results

- 8 rotating bottles
  - 56-D problem
  - Various interaction intensities
  - Budgets $[0,60]p \times [0,50]g$
- Two hands
  - 54-D problem
  - Strong interaction
  - Budgets $[0,100]p \times [0,50]g$
- **ECT** performs **best** overall
  - Accuracy
  - Speed

Stronger occlusions

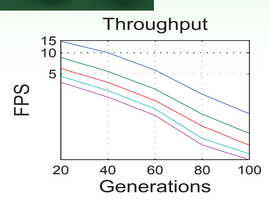


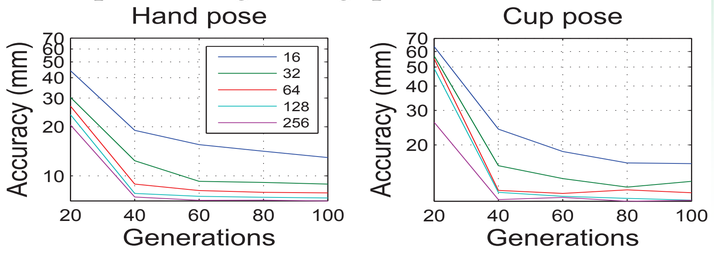
Qualitative results



Quantitative results

- Synthetic dataset of firm manipulation
- Method's explanation of firmness: 3.7mm, 1.3°
- 2fps tracking throughput





- Budget:
  - 64 PSO particles
  - 100 PSO generations
- Tracking throughput: 0.5fps
- Processed at 2X speed
- Challenges include:
  - Multiple objects
  - Elaborate interactions
  - Total occlusions
  - Cascaded occlusions

Qualitative results

For more information, visit http://www.ics.forth.gr/~kyriazis/?id=oui2015kyriazis, or contact {kyriazis,argyros}@ics.forth.gr

This work was partially supported by the IST-FP7-IP-288533 project RoboHow.cog and the FP7-ICT-2011-9-601165 project WEARHAP