Efficient Model-based 3D Tracking of Hand Articulations using Kinect

Iason Oikonomidis, Nikolaos Kyriazis, Antonis A. Argyros

Institute of Computer Science, FORTH, Greece **AND** Department of Computer Science, University of Crete, Greece

PROBLEM STATEMENT

Given a sequence of RGB-D images of a human hand, recover its 3D position, orientation and full articulation (26 DoFs).

MOTIVATION

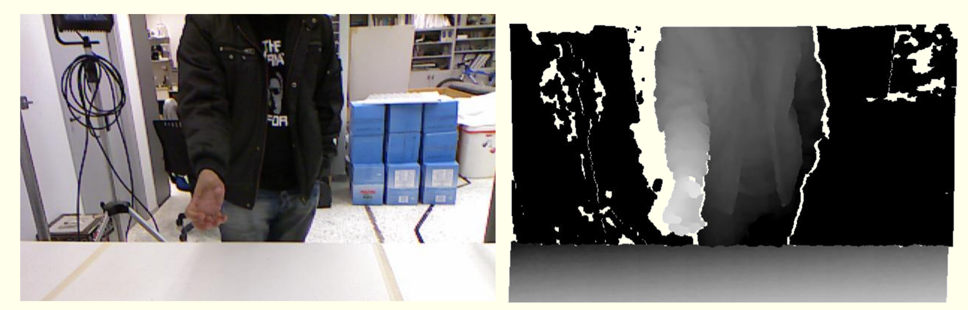
The markerless tracking of hand articulations is a challenging problem with diverse applications including H.C.I., understading human grasping, robot learning by demonstration, etc.

MAIN IDEA

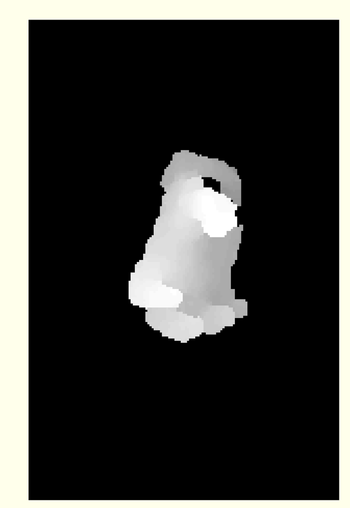
Exploit temporal continuity to fit a 3D hand model on the RGB-D data provided by the Kinect sensor, through Particle Swarm Optimization (**PSO**) [4] in a GPU powered framework [3].

PROPOSED METHOD

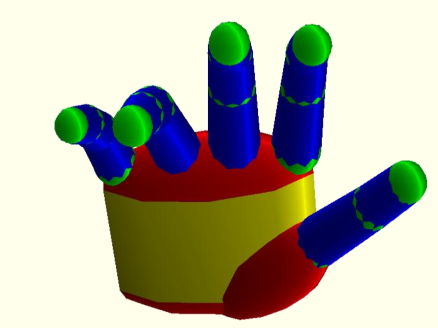
Observation



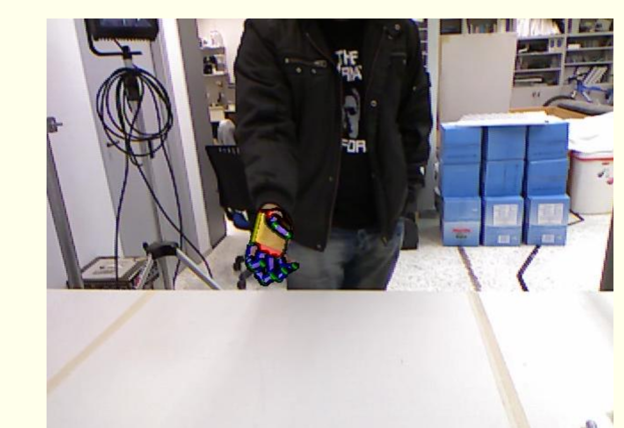
- RGB-D images are acquired using a Kinect sensor.
- Skin color and depth cues, along with the temporal continuity assumption are used to segment the hand.



Fit model to data

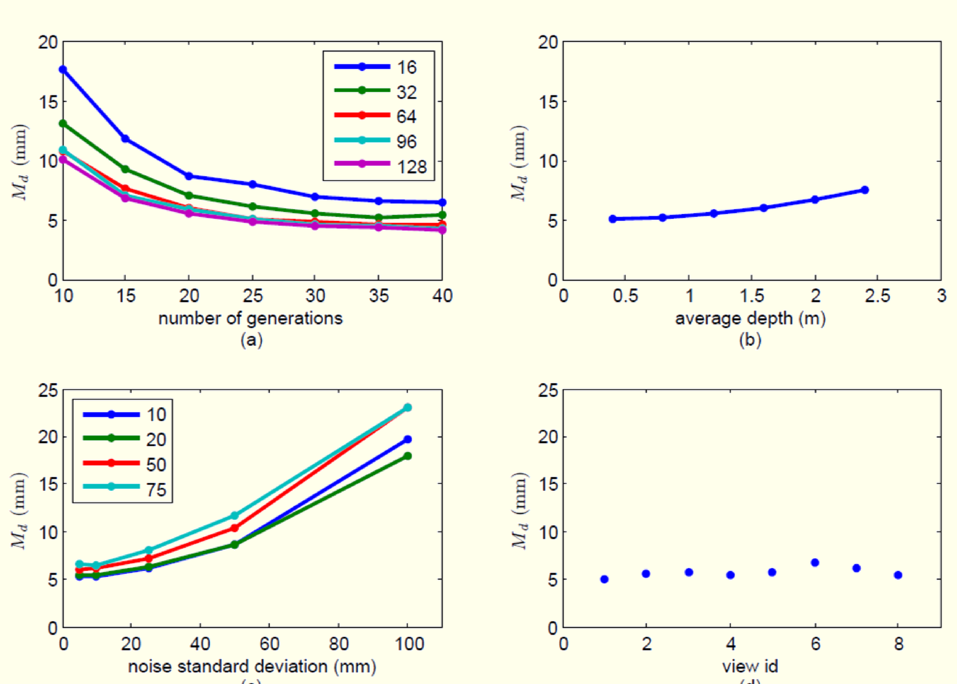


- A parametric hand model is employed [1].
  - Comprised of 37 geometric primitives: 15 cylinders and 22 spheres.
  - 26 DoFs: 6 for global pose and 20 kinematics angles.
- Given a hand configuration, a depth map and a skin occupancy map are synthesized by means of rendering.
- The maps are used to quantify the discrepancy between observation and hypothesis (objective function).
- A variant of the PSO method [4] searches in the model parameter space for the best scoring configuration.
  - Efficient evaluation of multiple hypotheses on the GPU [3].
- Candidate poses for the next frame are obtained by perturbing the solution of the previous frame.



EXPERIMENTAL RESULTS

Quantitative evaluation on synthetic data



64 particles and 40 generations yield 15fps on a modern PC

Sample results obtained on a real-world sequence

STRENGTHS OF THE APPROACH

- Joint optimization: no simplifying assumptions over the problem structure, simultaneous consideration of all parameters.
- Clear and effortless treatment of self-occlusions.
- Careful design and efficient GPU implementation [3] lead to near real-time performance (15fps).
- Extensibility [2] due to the model-based nature of the method.
- Minimally invasive markerless approach.

KEY REFERENCES

1. Oikonomidis, I., Kyriazis, N., Argyros, A. A. "Markerless and Efficient 26-DOF Hand Pose Recovery". *ACCV*, 2010.
2. Oikonomidis, I., Kyriazis, N., Argyros, A. A. "Full DOF tracking of a hand interacting with an object by modeling occlusions and physical constraints". **To appear** in *ICCV*, 2011.
3. Kyriazis, N., Oikonomidis, I., Argyros., A. A. "A GPU-powered Computational Framework for Efficient 3D Model-based Vision". *Technical Report TR420, ICS-FORTH*, 2011.
4. Kennedy, J., Eberhart, R. "Particle swarm optimization". *International Conference on Neural Networks*, 1995.

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

This work was partially supported by the IST-FP7-IP-215821 project GRASP