Giving a Hand to Kinect

Nikolaos Kyriazis, Iason Oikonomidis, Antonis A. Argyros

Institute of Computer Science, FORTH, Greece

AND

Computer Science Department, University of Crete, Greece

PROBLEM STATEMENT

Track the 3D position, 3D orientation and full articulation (26 DoFs) of one (or more) human hand(s) given a sequence of RGB-D frames of the scene.

MOTIVATION

The markerless tracking of hand articulations is a **challenging problem** with **diverse applications** such as gesture recognition, H.C.I., H.R.I., understanding human grasping, robot learning by demonstration, etc.

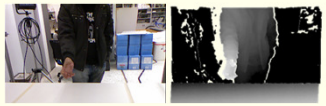
MAIN IDEA

Employ a hand model. Produce full hypotheses of it and extract full-image features. Quantify the similarity of synthesized and observed features in parallel [5]. Use these scores to drive an iterative optimization process using Particle Swarm Optimization (PSO).

PROPOSED METHOD

Observation

- RGB-D frames are acquired by a **Kinect** using **Microsoft SDK**.



- **Skin color** and **depth** cues, along with the **temporal continuity** assumption are used to **segment** the hand.

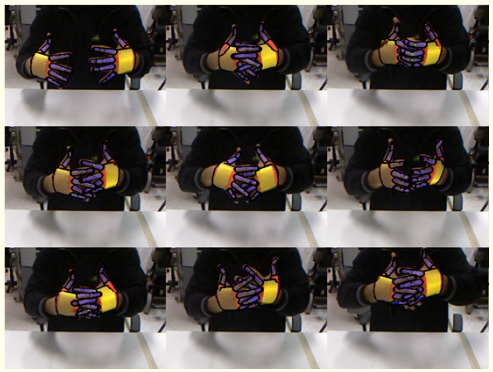
Fit model to data

- Employ a parametric hand model [1] comprised of 15 cylinders and 22 spheres.
  - ❑ For a single hand: 26 DoFs, 6D global pose, 20 joint angles.
  - ❑ For two hands: 2x26 = 52 DoFs
- From a full configuration (all 26 or 52 DoFs of the hand(s) model), a **skin occupancy map** and a **depth map** can be **synthesized** by means of rendering.
- These **maps** are used to quantify the **discrepancy** between **observation** and **hypothesis** (objective function).
- The objective function also **penalizes physically implausible** configurations (inter- and intra-hand collision checking).
- A **variant** of the **Particle Swarm Optimization** method searches in the model parameter space for the best scoring configuration.
  - ❑ **Efficient** evaluation of multiple hypotheses on the **GPU** [5].
- Candidate poses for the **next frame** are obtained by **perturbing** the solution of the **previous frame**.

EXPERIMENTAL RESULTS

A single hand [3]

Two hands [4]



Single hand: 64 particles and 30 generations yield 20fps on a modern PC
Two hands: 64 particles and 40 generations yield 4fps on a modern PC
For the quantitative evaluation w.r.t. accuracy, tolerance to noise, computational budget, viewpoint invariance, e.t.c, check [3], [4]

3D hand tracking in CHALEARN:
Use the developed 3D hand tracking approach to recognize hand postures defined by the user on the fly!

STRENGTHS OF THE APPROACH

- Fundamental building block for countless applications.
- Tracking hand(s) with unconstrained motion.
- Non invasive markerless approach.
- Safeguards against physically implausible hand configurations.
- Careful design and exploitation of parallelism in a GPU implementation [5] lead to a computationally efficient system [1-4].

KEY REFERENCES

1. Oikonomidis, I., Kyriazis, N., Argyros, AA, "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". *ICCV*, 2011.
3. Oikonomidis, I., Kyriazis, N., Argyros, A. A. "Efficient Model-based 3D Tracking of Hand Articulations using Kinect". *BMVC*, 2011.
4. Oikonomidis, I., Kyriazis, N., Argyros, A. A. "Tracking the articulated motion of two strongly interacting hands". *CVPR*, 2012.
5. Kyriazis, N., Oikonomidis, I., Argyros., A. A. "A GPU-powered Computational Framework for Efficient 3D Model-based Vision". *Technical Report TR420, ICS-FORTH*, 2011.

For more information, visit http://cvrlcode.ics.forth.gr/handtracking

This work was partially supported by the IST-FP7-IP-215821 project GRASP and the IST-FP7-IP-288533 project RoboHow.Cog