Dynamics Based 3D Skeletal Hand Tracking

Stan Melax, Leonid Keselman, Sterling Orsten
Intel Corporation

Abstract
This research explores a new approach to tracking hands, or a rigidly articulated model, by using rigid body simulation based on a depth sensor’s samples. The system generates constraints that limit motion orthogonal to the rigid body model’s surface. These constraints, along with prior motion, collision/contact constraints, and joint mechanics, are resolved with a projected Gauss-Seidel solver.

Hand Model
We construct a model of the human hand using seventeen convex polyhedral parts (bones). These bones overlap slightly at joints, and are constrained to limit motion to physically plausible states. We use a reference hand model 20 cm long, authored in 3D Studio Max, and can parameterically scale different bones to track users with different sized hands. As our model is specified entirely in data, our generic tracking system is able to track other articulated objects without retraining.

Rigid Body Dynamics
Our approach to hand tracking draws its technical core from the world of rigid body physics. We phrase pose tracking as a linear complementarity problem, and use a projected Gauss-Seidel solver to find a solution.

Each frame, we generate surface constraints based on our depth camera’s samples, in addition to mechanical constraints from the hand model. We then run a small, fixed number of solver iterations. This flexible approach enables us to easily incorporate additional information and heuristics, as elaborated below.

Multiple Simulations
Our efficient simulation framework is extensible and flexible. Conjectures, heuristics and externally derived information can all be formulated as additional constraints in the solver.

To test a variety of competing hypotheses, we run multiple biased simulations, each incorporating the corresponding constraints. The outcomes of the simulations are evaluated and the best fit is returned.

Voxel Subsampling
We utilize volumetric subsampling to generate improved depth samples from the camera. It allows us to both discard sparse camera noise (by rejecting insufficiently dense voxels) and produce a joint cloud with uniform spatial sampling across the surface of the tracked object.

Surface Constraints

Surface Constraints: Fitting Error

Surface constraints that attract and limit motion to be parallel to surface

Temporal Coherence
Due to camera quality, occlusions, and pose ambiguity from a single depth image, using temporal coherence is necessary for any hand tracking solution.

3D Error Measure
We evaluate the fitting error with a lower bound approximation of the distance between the model and what we can determine about the location of the user’s hand. In particular, for each bone in the tracking model we sum:

• The distance that the bone is known to be in front of (not occluded by the point cloud if applicable)
• The distance from the bone surface to the furthest depth sample point mapped to that bone

Results & Usages
Our approach is able to track the fully articulated pose of a user’s hand in a wide variety of configurations and motions. We are even able to handle the difficult case of a largely occluded hand model, where the depth information is ambiguous. For example, we are able to correctly determine which finger is bent from a side profile configuration.

By using our tracked pose information, it is possible to use a powered ragdoll approach to create an interactive physics playground. In these examples, we use the Bullet physics engine to enable interaction with blocks and swing sets without any specialized code. The level design becomes the interaction design.

References