Motivation

Hand tracking is a process of accurately reconstructing shape and articulation of human hands. It is a crucial component of natural human-computer interfaces and animation of humanoid avatars. A number of hand tracking algorithms has been recently proposed (citations). However, in most consumer applications hand tracking is just a single components of a bigger pipeline. Before tracking a suitable hand model is obtained. Once the hand pose parameters are found, the tracking result is displayed by posing the model. Both tasks are not trivial.

# Why creating a high precision hand model?

Hand model should be able to accurately represent the observed data. The discrepancy between the optimal model pose given the data and the true hand pose can be significant, especially if the hand model does not reflect all the degrees of freedom of a hand.

# Why realistically animating a hand model?

The hand skinning quality is obviously important for digital avatars applications. In AR and VR applications a 3D hand model can properly interact with 3D objects, establish a realistic contact and disappear behind them. Given that, the degree of immersion into virtual reality depends on whether a user sees own realistic hands (find a study mentioned by Leap Motion).

# Why single representation

Since each stage of the pipeline requires a hand model, it is worth trying to find a model representation suitable for all three stages. If that representation is at least as good as its available alternatives at each stage, it should be preferred to them. Firstly, it eliminates a problem of retargeting, which doesn’t only require additional effort but also brings additional imprecision. Secondly, a system with a single representation is simpler and more elegant.

# Why convolution surfaces representation

We suggest to use convolution surfaces representation of the hand model. Convolution surface is an implicit surface which is described by a control skeleton. The skeleton may consist of points, edges or polygons (Bloomenthal). In each vertex of the skeleton we define a radius. The radius in intermediate points is a linear combination of the radii at the neighboring vertices. Given the topology of the underlying skeleton, the model can be represented with convolution surface up to high precision (find some theoretical estimates in the literature?)

# for model fitting?

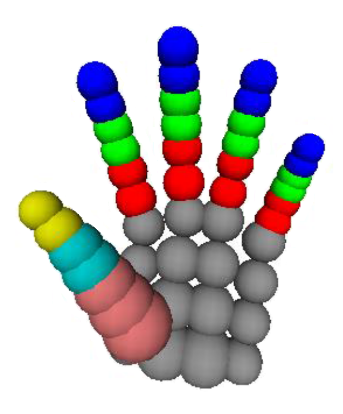
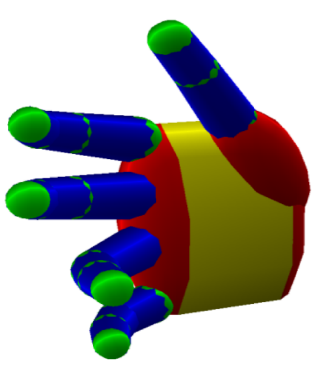


Figure 1. (a) Qian, Sridhar, (b) Oikonomidis, (c) Melax, (d) Sharp, Schroeder

The spheres and mixed cylinders/spheres hand model representations (a) are ubiquitous in hand tracking, because they are well suited for tracking tack per se (see next) and can be quickly to created manually. If a small number of such primitives is used, the precision of the model is low. It is possible to obtain higher precision by increasing the number of primitives. But this defeats the purpose of simplicity if the representation.

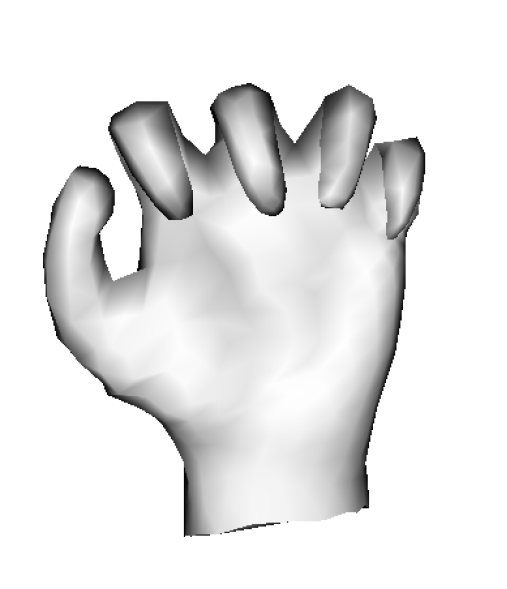
The triangles mesh representation can approximate the hand to a high precision. However, there is obvious way to specify the model parts that are rigid and should be kept the same shape between the poses. This results in overfitting to local skin deformations.

**Convex bodies.** In convex bodies representation (b) each rigid section is approximated with rigid convex body.

# for hand tracking?

For model based tracking the main operation is to find the closest point on the model for a given data point. For spheres/cylinder and convolution surfaces model this operation is done in closed for each rigid segment. For a triangular mesh this operation has complexity linear in number of triangles. Moreover, the triangular mesh has (much) more degrees of freedom than the underlying problem. Without additional regularization, rigid parts of the hand model can deform to fit the data; the individual vertices can shift to fit the sensor noise.

# for hand skinning?



The Linear Blend Skinning approach used to pose the hand model in previous works creates artifacts, the fingers looks like made from rubber. The state of the art approaches in hand skinning are implicit surfaces-based. A convolution surfaces model serves a ready to use input for such an approach.

Contributions

* Suggested a single hand model representation suitable for modeling, tracking and skinning tasks.
* Developed an approach for approximation a model with convolution surface, given a skeleton topology.
* Formulated position-based inverse kinematics algorithm for hand tracking. The position based approach does not require walking through the hierarchical chain of joint, thus saving the computational power. Also, position-based inverse kinematics does not involve linear approximations or rotations, thus making the optimization more stable.
* Suggested an automatic approach for field functions construction in implicit skinning. Replaced newton iteration for vertex projection by a closed form solution.

Related Literature

* Sridhar et. al. perform a grid search over fixed range of on finger thickness, hand length and width.
* Rhee et. al., use a single image of a hand at rest pose to infer joint hand joint locations from skin creases. Given the skeleton obtained at the previous step and the hand contour from the image, they deform a template hand mesh to fit this data.
* Taylor et. al. generate a user-specific hand model from an RGBD video sequence. The model is represented as a triangular mesh with an embedded skeleton. In each frame the hand pose is initialized using an appearance-based tracking algorithm. The hand model parameters are found by solving a single optimization problem formulated for the entire video sequence which also finds hand pose in each frame. This approach is prone to overfitting to local skin deformations, since the triangles mesh has excessive number of degrees of freedom.
* Straka et al. also fit the template mesh with attached skeleton to 3D data. The model is deformed to fit the data while keeping the vertices attached to their corresponding bones. It is on clear if the approach will be able to handle a hand motion sequence. The results are demonstrated on a full body model.
* Khamis et. al. fit a hand model for a specific user by finding its shape coordinates in the basis of mesh matrices and bones locations. As in the approach by Taylor et. al, they optimize simultaneously for pose and shape parameters in all the frames of an RGBD sequence across all the subjects. Requires large number of subjects as a regularization for excessive degrees of freedom. The results generated by the approaches of Taylor and Khamis could be used in our system to create a hand model representation tailored for efficient tracking and animation.
* Albrecht et al. developed an approach for creating an anatomically realistic hand model that includes bones and muscles structure. Their approach requires several prerequisites including plaster cast of a human hand and laser scanner for manually creating a physically realistic hand template. Given user-defined correspondences between 3D feature points and the hand image, a specific hand model is created by deforming a generic hand model.