Literature Review - 3

- Narasimha Prasanth Chintarlapalli Reddy
- Narasimha_Reddy@student.uml.edu
- 01669930

The **primary** research paper I chose for this Literature Review is **"Online Generative Model Personalization for Hand Tracking"**[1] by **Tkach, Anastasia and Tagliasacchi, Andrea and Remelli, Edoardo and Pauly, Mark and Fitzgibbon, Andrew**.

```
@article{Tkach:2017:OGM:3130800.3130830,
 author = {Tkach, Anastasia and Tagliasacchi, Andrea and Remelli, Edoardo and Pauly, Mark and Fitzgibbon, Andrew},
 title = {Online Generative Model Personalization for Hand Tracking},
 journal = {ACM Trans. Graph.},
 issue_date = {November 2017},
 volume = {36},
 number = {6},
 month = nov,
 year = {2017},
 issn = {0730-0301},
 pages = {243:1--243:11},
 articleno = {243},
 numpages = {11},
 url = {http://doi.acm.org/10.1145/3130800.3130830},
 doi = {10.1145/3130800.3130830},
 acmid = {3130830},
 publisher = {ACM},
 address = {New York, NY, USA},
 keywords = {articulated registration, generative tracking, motion capture, real-time calibration, real-time hand tracking},
}
```

The **secondary** research paper I chose is **"Construction and animation of anatomically based human hand models**
"[2] by **Albrecht, Irene and Haber, Jorg and Seidel, Hans-Peter**.

```
@inproceedings{Albrecht:2003:CAA:846276.846290,
 author = {Albrecht, Irene and Haber, J\"{o}rg and Seidel, Hans-Peter},
 title = {Construction and Animation of Anatomically Based Human Hand Models},
 booktitle = {Proceedings of the 2003 ACM SIGGRAPH/Eurographics Symposium on Computer Animation},
 series = {SCA '03},
```

year = {2003},
isbn = {1-58113-659-5},
location = {San Diego, California},
pages = {98--109},
numpages = {12},
url = {http://dl.acm.org/citation.cfm?id=846276.846290},
acmid = {846290},
publisher = {Eurographics Association},
address = {Aire-la-Ville, Switzerland, Switzerland},
}

Human hand has ability to perform fine motor manipulations and powerful work alike. It is a masterpiece of mechanical complexity. Designing an animatable human hand requires a lot of detail. It is not easy to replicate all the natural motions performed by hand.

Both our secondary and primary papers discuss on various ways of animating a hand possible.

Our secondary paper presents a human hand model with underlying anatomical structure. Muscle contraction values play key role in controlling animation of hand model. The secondary paper employs a physically based hybrid muscle model to convert these contraction values into movement of skin and bones. Pseudo muscles directly control the rotation of bones based on anatomical data and mechanical laws, while geometric muscles deform the skin tissue using a mass-spring system. Thus, resulting animations automatically exhibit anatomically and physically correct finger movements and skin deformations. The secondary paper present a deformation technique to create individual hand models from photographs. A radial basis warping function is set up from the correspondence of feature points and applied to the complete structure of the reference hand model, making the deformed hand model instantly animatable.

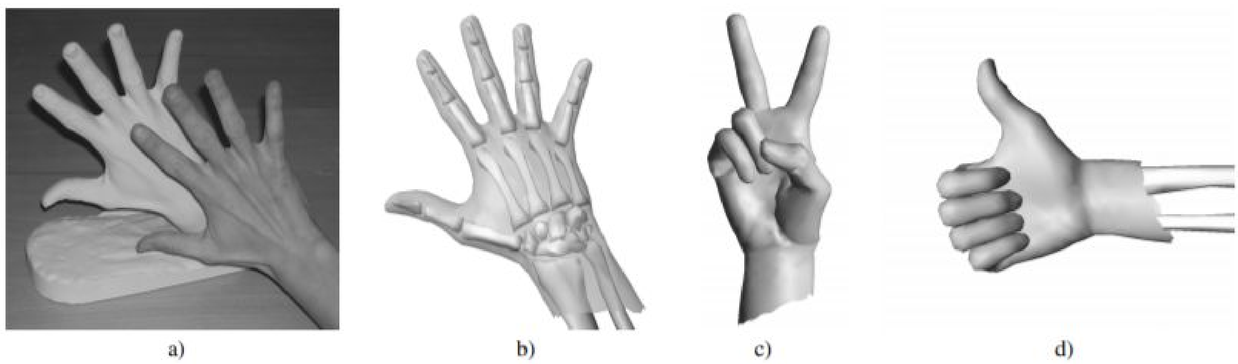
Let's go into the fine details of our secondary paper. Hands play vital role in our everyday life. Hands play key role in expressing simple to complex feelings like shaking hands, expressing gratitude or sympathy. We need them for eating, playing, writing, working, communicating, in a nutshell: for everything.

Some uses of modeling/animating hand movements are: models can be used for teaching and practicing sign language. Also to teach someone to operate machines or to give online assembly instructions.

The three main contributions of this paper are:

1. A human hand model with anatomical structure, structure, suitable for real-time animation using physics-based simulation of muscles and elastic skin properties.
2. a hybrid muscle model that comprises pseudo muscles and geometric muscles. Pseudo muscles directly control the rotation of bones based on anatomical data and mechanical

laws, while geometric muscles deform the skin tissue using a mass-spring system.
3. A deformation technique based on feature points to warp the complete structure of a reference hand model to an individual hand model taken from a photograph.



Construction and animation of the reference hand model: a) plaster cast of a human hand ready for 3D scanning; b) assembly of skin mesh and individual bone meshes; c) and d) skin tissue deformation during animation.

The first contribution is **The Reference Hand Model**. Our secondary paper has a reference hand model with the following building blocks:

- the skin surface, which is represented by a triangle mesh consisting of 3000 triangles;
- the skeleton of the hand, composed of 29 triangle meshes corresponding to the individual bones of the human hand and forearm
- a set of virtual muscles, which are embedded in between the skin surface and the skeleton;
- a mass-spring system, interlinking the skin, skeleton, and muscles;
- a joint hierarchy, which matches the structure of the skeleton, with an individually oriented coordinate system at each joint center defining valid axes of joint rotation.

Using hierarchy of coordinate systems, they modeled the degrees of freedom(DOF) for each joint individually. The only joints ignored in this model are the joints between individual wrist bones.

The second contribution is **A Hybrid Muscle Model.** Muscle mechanics of the human hand have evolved to a degree of complexity that is unique among mammals. This evolutionary process took place in order to allow us to perform fine motor manipulations and powerful manual work alike. Modeling and simulating all the subtle anatomical details of the muscles of the human hand is an impractical approach. In this section, we present a hybrid muscle model, which is flexible enough to cover the rich variety of muscle mechanics in the human hand and yet is easy to use.

In this model, rotation of single bone is calculated and modeled with all the other bones. If we want to rotate a chain of bones, the moment of Inertia depends on the position of all bones in that chain. See the following figure for reference. Moment of Inertia can be determined only by accounting for all the bones in the figure below.

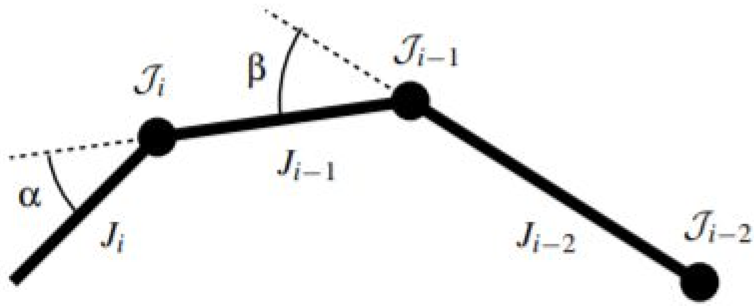


Figure 3: *Moments of inertia for a chain of bones.*

The third contribution is **New Hand Models from Photographs.** This model does not require a 3D target hand model to be obtained in time-consuming scanning process. Instead, this model use a simple photograph of the individual hand to be modeled. The photograph merely needs to show a simple ruler. . Since there are no other prerequisites for the photograph, low-cost consumer cameras can be used for the acquisition.

It is obvious that one of the major limitations is that if animations falis, the model is compromised hugely. Since this is built around a reference model, a lot can be done to make this a dynamic model to be used for all sorts of users.

Now let's discuss our primary paper i.e Online Generative Model Personalization for Hand Tracking. We need to realize that this paper was published in Nov 2017 edition and our secondary paper was published in the year 2003. There are lot of developments in computer graphics during this period. This paper presents a new algorithm for real-time hand tracking on commodity depth sensing devices. This method does not require a user specific calibration, but learns the geometry as the user performs live in front of the camera, thus enabling seamless virtual interaction at the consumer level. The key novelty in this approach is an online optimization algorithm that jointly estimates pose and shape in each frame, and determines the uncertainty in such estimates. This knowledge allows the algorithm to integrate per-frame estimates over time, and build a personalized geometric model of the captured user. This approach can easily be integrated in state-of-the art continuous generative motion tracking software.

Let’s go into the details of the algorithm, but let's discuss the importance of this paper. In our everyday life we interact with the surrounding environment using our hands. The focus of recent research has been to seamlessly get such interaction to virtual objects such as VR/MR headsets. This is also essential in games and films for pre-visualization, where motion can be transferred in real time to a virtual avatar. It is desirable for this type of technology to be accurate, robust and have seamless and real time deployment.

Fig. 1. Our adaptive hand tracking algorithm optimizes for a tracking model on the fly, leading to progressive improvements in tracking accuracy over time. **Above**: hand surface color-coded to visualize the spatially-varying confidence of the estimated geometry. Insets: color-coded *cumulative* certainty. Notice how in the last frame all parameters are certain. **Below**: histograms visualize the certainty of each degree of freedom, that is, the diagonal entries of the inverse of the covariance estimate from: (a) data in the current frame Σ^*, or (b) the information $\hat{\Sigma}$ accumulated through time by our system.

The model we discussed in our **secondary** paper is **offline** model calibration. But in this **primary** paper of our's the model is **online** model calibration. The advantages of online model calibration are that it is quick, real-time and more accurate.

The contributions of this primary paper are a principled way to integrate per-frame information into an online real-time pose/shape tracking algorithm: one that successfully estimates the hand's pose, while simultaneously refining its shape. That is, as more of the user's hand and articulation is observed during tracking, the more the tracking template is progressively adapted to match the performer, which in turns results in more accurate motion tracking. From single frame, not all degrees of freedom can be estimated. For example, just by observing a straight finger, we cannot accurately estimate length of a phalanx. This model automatically estimates the confidence in each frame and builds the model. If the user maintains an unreasonable shape, the system goes back to its default pose. While estimating this, it simultaneously models the hand. That is the beauty of this model. This is the quality which makes it very useful in games and films. The key concept used in this model is e Levenberg-Marquardt Kalman Filter. Overall, this solution yields a fully automatic, real-time hand tracking system that is well-suited for consumer applications.

Let's describe the joint calibration and tracking algorithm, which combines the Levenberg-style optimization of previous hand trackers with the uncertainty maintenance framework of Kalman filtering. Previous hand tracking work has made use of temporal smoothing priors to propagate pose information from previous frames, without the use of filtering. However this approach cannot be used for shape because it is so weakly constrained in any given frame, and because its temporal prior is so strong, as shape parameters are persistent over time: It is observed that the same user performing in front of camera for thousands of frames. But sufficient info required is not present in every frame, the example of which is discussed above.

$$\hat{x}_n = \arg\max_{x_n} \underbrace{\log\left(p(x_n^*|x_n)\,p(x_n|\hat{x}_{n-1})\right)}_{L(x_n)}$$

$$p(x_n^*|x_n) = \exp\left(-\tfrac{1}{2}(x_n^* - x_n)^T \Sigma_n^{*-1}(x_n^* - x_n)\right)$$

$$p(x_n|\hat{x}_{n-1}) = \exp\left(-\tfrac{1}{2}(x_n - \hat{x}_{n-1})^T \hat{\Sigma}_{n-1}^{-1}(x_n - \hat{x}_{n-1})\right)$$

$$\hat{\Sigma}_n^{-1} = \frac{\partial^2 L}{\partial x_n^2}\Big|_{\hat{x}_n} \approx \begin{bmatrix}\sqrt{\Sigma_n^{*-1}} \\ \sqrt{\hat{\Sigma}_{n-1}^{-1}}\end{bmatrix}^T \begin{bmatrix}\sqrt{\Sigma_n^{*-1}} \\ \sqrt{\hat{\Sigma}_{n-1}^{-1}}\end{bmatrix} = \Sigma_n^{*-1} + \hat{\Sigma}_{n-1}^{-1}$$

Table 1. *Split* cumulative regression – Kalman Filter (KF)

$$\hat{x}_n = \arg\max_{x_n} \underbrace{\log\left(p(d_n|x_n)\,p(x_n|\hat{x}_{n-1})\right)}_{L(x_n)}$$

$$p(d_n|x_n) = \exp\left(-\tfrac{1}{2}(d_n - F(x_n))^T (d_n - F(x_n))\right)$$

$$p(x_n|\hat{x}_{n-1}) = \exp\left(-\tfrac{1}{2}(x_n - \hat{x}_{n-1})^T \hat{\Sigma}_{n-1}^{-1}(x_n - \hat{x}_{n-1})\right)$$

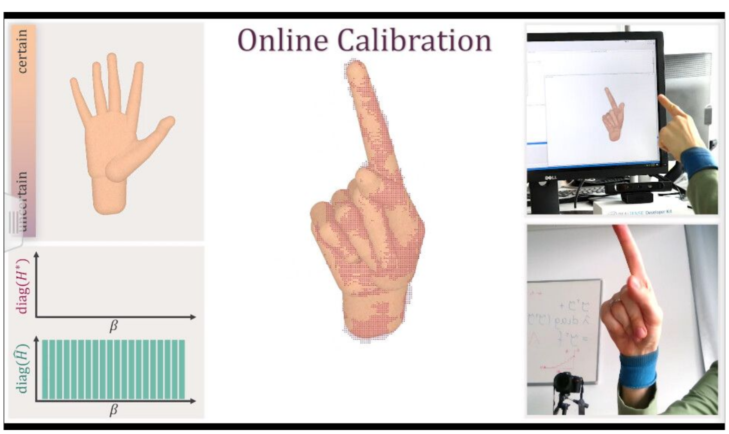
$$\hat{\Sigma}_n^{-1} = \frac{\partial^2 L}{\partial x_n^2}\Big|_{\hat{x}_n} \approx \begin{bmatrix}-\frac{\partial F(\hat{x}_n)}{\partial x_n} \\ \sqrt{\hat{\Sigma}_{n-1}^{-1}}\end{bmatrix}^T \begin{bmatrix}-\frac{\partial F(\hat{x}_n)}{\partial x_n} \\ \sqrt{\hat{\Sigma}_{n-1}^{-1}}\end{bmatrix} = \Sigma_n^{-1} + \hat{\Sigma}_{n-1}^{-1}$$

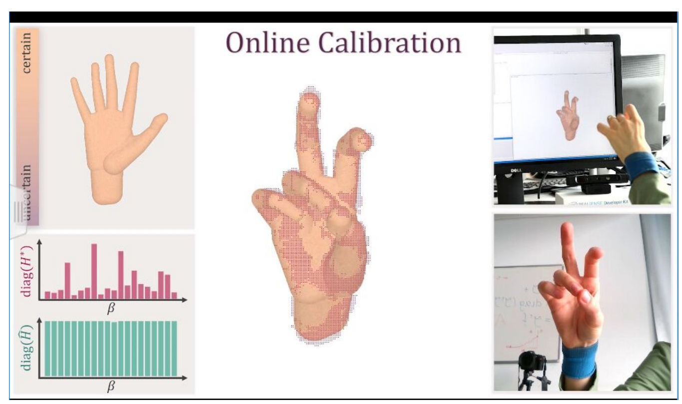
Table 2. *Joint* cumulative regression – Iterated Extended KF (IEKF)

The above tables represent the filter expressions used to calculate the per-frame confidence.

Generic algorithms register geometric model of hand movements to sensor data. The accuracy of this depends on how well the user fits the tracking paths. Our secondary paper calibrate by jointly fitting a model to a small set of input frames. Current requirement is for the user to select any frame from an uncalibrated model and then wait for the calibration algorithm to finish. TO regress a given parameter, there are good and bad poses. Again, this depends on the parameter. We cannot expect to find the length of phalanx by looking at a straight finger. So, this model detects the requirements for parameters in real time. So this means that it is not required for each user to calibrate individually as mentioned previously.

Per frame regression really helps to get this model together.





We can see in the above picture how the image rendered in the screen adjusted to user’s hand in real time. This model will be very useful in games and films. VR/MR has very good applications for this model.

The major limitation of both our primary and secondary paper are that: if tracking fails, the model's quality is compromised. For our primary model, the optimization relies on heavy parallelization and high-end GPU hardware. This can be reduced in future work.