Supplemental Material - Walking on Thin Air: Environment-Free Physics-based Markerless Motion Capture

Micha Livne*, Leonid Sigal†, Marcus A. Brubaker‡ and David J. Fleet* .

*Department of Computer Science
University of Toronto
Toronto, Canada
{mlivne, fleet}@cs.toronto.edu

†Department of Computer Science
University of British Columbia
Vancouver, Canada
lsigal@cs.ubc.ca

‡Lassonde School of Engineering
York University
Toronto, Canada
mab@eecs.yorku.ca

	Our Model	SCAPE	Hasler	Allen
Mesh recon.	Linear	Nonlinear (Least Squares)	Nonlinear (Poisson)	Linear
Mesh recon. speed	1.82 [mSec] (full) 1.17 [mSec] (pose)	1 [Sec] mesh only (given matrices)	25 [Sec]	13 [mSec]
Semantics	Skeleton Body shape Anthropometrics	Skeleton Body shape	None	Skeleton
Mean reconstruction error	5.3 [mm]	N/A	54 [mm]	4.9 [mm]

Table III: A comparison between our model, SCAPE [1], Hastler [2], and Allen [3]. **Semantics** refers to the direct interpretation of model parameters. Semantic parameters, such as explicit anthropometrics (skeleton) representation, is useful in the context of tracking. In our case, it allows direct control over which set of parameters to optimize.

***Abstract*—Supplemental material for "Walking on Thin Air: Environment-Free Physics-based Markerless Motion Capture" paper.**

VI. Results

The accompanying video can be found in the project website: https://research.seraphlabs.ca/u-of-t/performance-capture/.

VII. Learning the Body Model

In order to learn the body mesh model in Sec. III-A we use the Hastler dataset [2], which consists of 111 subjects with 520 poses, all with registered meshes. We learn the model by minimizing Eq. 14 w.r.t. the weights $W = \{w_{ib}\}$, the different mesh templates $\{\tilde{\mathbf{p}}^s\}_s$ per subject s, and the template pose and anthropometrics $\{\mathbf{q}^s, \boldsymbol{\ell}^s\}_s$ per subject s. We define the number of weights per vertex based on joints proximity along the kinematic tree, with BFS of distance 3. Since we optimize all parameters w.r.t. to the same reconstruction error function (Eq. 14), we get an accurate reconstruction despite the simplicity of the model, when compared with other state-of-the-art models (Table III).

Notice that some of the models in Table III were trained on different datasets. However, our main goal is to demonstrate that our model is comparable to state-of-the-art models, rather than a comprehensive comparison. In our dataset only 43 out of 111 subjects have more than a single pose, which is required for our training. However, those subjects account for 86% of the total number of poses (450 out of 520 poses).

A. Model Parameters Optimization

While it is possible to optimize for the weights, mesh templates and pose simultaneously, it is a slow, non-convex and nonlinear optimization. Instead we alternated the optimization between the parameters, which yields a much faster optimization process, and is also convex w.r.t. the mesh templates and weights. Our learning process includes the following steps:

1) Initialize $\mathbf{q}^s$ for all subjects by fitting landmarks (based on the registered meshes).
2) Repeat until convergence:
   a) Optimize weights $W = \{w_{ib}\}$ given current poses and mesh templates. We optimize a global reconstruction error function as W is shared among all subjects

$$\mathcal{F} = \sum_{s \in \mathcal{S}} \mathcal{F}^s \left(\Theta^s, \mathcal{D}\right) \tag{17}$$

where $\mathcal{S}$ are all subjects with more than a single pose in our dataset. By examining Eq. 3, it is clear that Eq. 17 is convex w.r.t W. Thus, we can define $\mathbf{A}_i^{s,j}$ by rewriting Eq. 3 as

$$\begin{aligned} &\left(\ \dots \quad \mathbf{M}_b\left(\mathbf{q}_j\right) \cdot \tilde{\mathbf{M}}_b\left(\tilde{\mathbf{q}}^s\right)^{-1} \tilde{\mathbf{p}}_i^s \quad \dots \ \right) \mathbf{w}_i = \mathbf{p}_i^j \\ &\Rightarrow \mathbf{A}_i^{s,j} \cdot \mathbf{w}_i = \mathbf{p}_i^j \end{aligned} \tag{18}$$

where j is a pose index (over all poses of subject s), and $\mathbf{w}_i$ is the weights of vertex i as a vector. By concatenating $\mathbf{A}_i^j, \mathbf{p}_i^j$ of all subjects s and all poses j it is easy to calculate the least-squares solution.

b) Optimize mesh template per subject, given current weights and poses. We optimize the mesh template independently per subject. By examining Eq. 3, it is clear that Eq. 14 is convex w.r.t to $\tilde{\mathbf{p}}^s$, and we can define $\mathbf{T}_i^{s,j}$ by rewriting Eq. 3 as

$$\mathbf{T}_i^{s,j} \cdot \tilde{\mathbf{p}}_i^s = p_i^j \tag{19}$$

per vertex i and per pose j. By concatenating matrix $\mathbf{T}_i^{s,j}$ for all poses j per subject, a simple least-squares solution can be used here as well.

c) Optimize pose $\mathbf{q}^j$ for all poses of all subjects. All poses can be estimated independently, by using nonlinear and non-convex optimization of Eq. 14 w.r.t. the pose parameter $\mathbf{q}$. We used BFGS to optimize for pose parameters $\mathbf{q}_s^k$ for all poses of all subject. Note: Since the optimization is local, good initialization is required.

Practically, two full iterations iterations of $(a), (b), (c)$ were enough to get close to convergence. The result of the model parameters optimization phase are W, shared weights to be used in LMB, the mesh template per subject $\tilde{\mathbf{p}}^s$, the bones length $\boldsymbol{\ell}^s$ per subject, and the pose vector $\mathbf{q}_s^j$ per pose j and subject s.

B. Basis Learning

Once we learn the model parameters as explained above, we can train a linear regressor with basis $\mathbf{B}_{\boldsymbol{\ell}}$ from bones length to a mesh template $\tilde{\mathbf{p}}$, s.t.

$$\tilde{\mathbf{p}}^s \approx \mathbf{B}_{\boldsymbol{\ell}} \cdot \boldsymbol{\ell}^s \tag{20}$$

where $\mathbf{B}_{\boldsymbol{\ell}}$ is learned with a least squares formulation.

By applying PCA to the null space of the linear regression basis (difference between regressed mesh template and $\tilde{\mathbf{p}}$), we can learn the body shape basis $\mathbf{B}_{\boldsymbol{\beta}}$. We used the first 10 PC as a linear basis. Once we have the two basis, $\mathbf{B}_{\boldsymbol{\ell}}, \mathbf{B}_{\boldsymbol{\beta}}$, we can generate new mesh templates given any desired bones length $\boldsymbol{\ell}$ and body shape score $\boldsymbol{\beta}$, as shown in Eq. 2, and generate a mesh for any given pose by using Eq. 3.

References

[1] D. Anguelov, P. Srinivasan, and D. Koller, "Scape: shape completion and animation of people," *ACM Transactions on Graphics*, 2005.

[2] N. Hasler, C. Stoll, M. Sunkel, B. Rosenhahn, and H.-P. Seidel, "A statistical model of human pose and body shape," *Computer Graphics Forum*, vol. 28, no. 2, pp. 337–346, 2009.

[3] B. Allen, B. Curless, and A. Hertzmann, "Learning a correlated model of identity and pose-dependent body-shape variation for real-time synthesis," *CM SIGGRAPH/Eurographics Symposium on Computer Animation*, 2006.