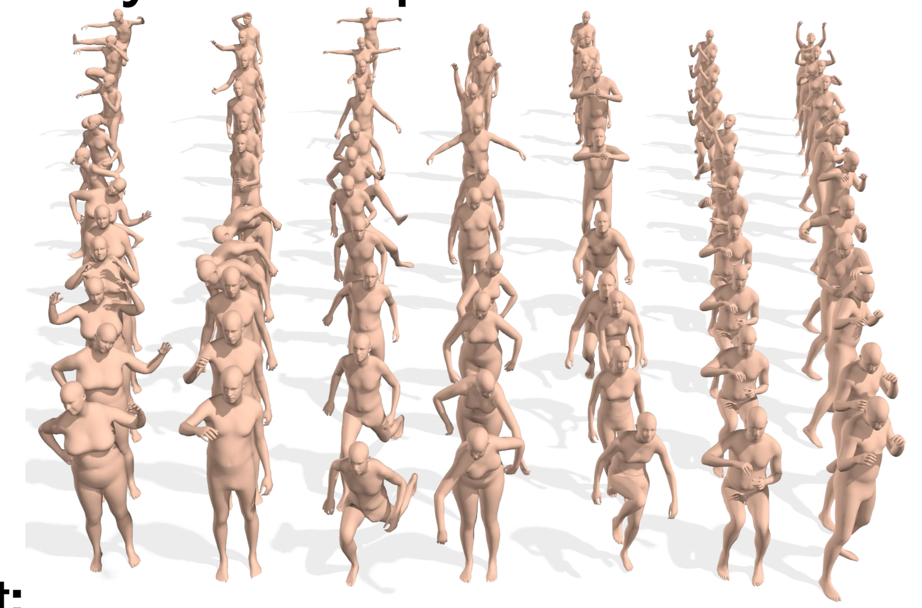
AMASS: Archive of Motion Capture as Surface Shapes

Naureen Mahmood[1], Nima Ghorbani[2], Nikolaus F. Troje[3], Gerard Pons-Moll[4], Michael J. Black[2]
[1]Meshcapade, [2]MPI for Intelligent Systems, [3]York University, [4]MPI for Informatics
amass@tue.mpg.de

ICCV 2019 Seoul, Korea

Unify MoCap Datasets



Context:
- Deep-learning of human motion requires large datasets with natural variation
- Optical marker-based MoCap records human motion with high fidelity

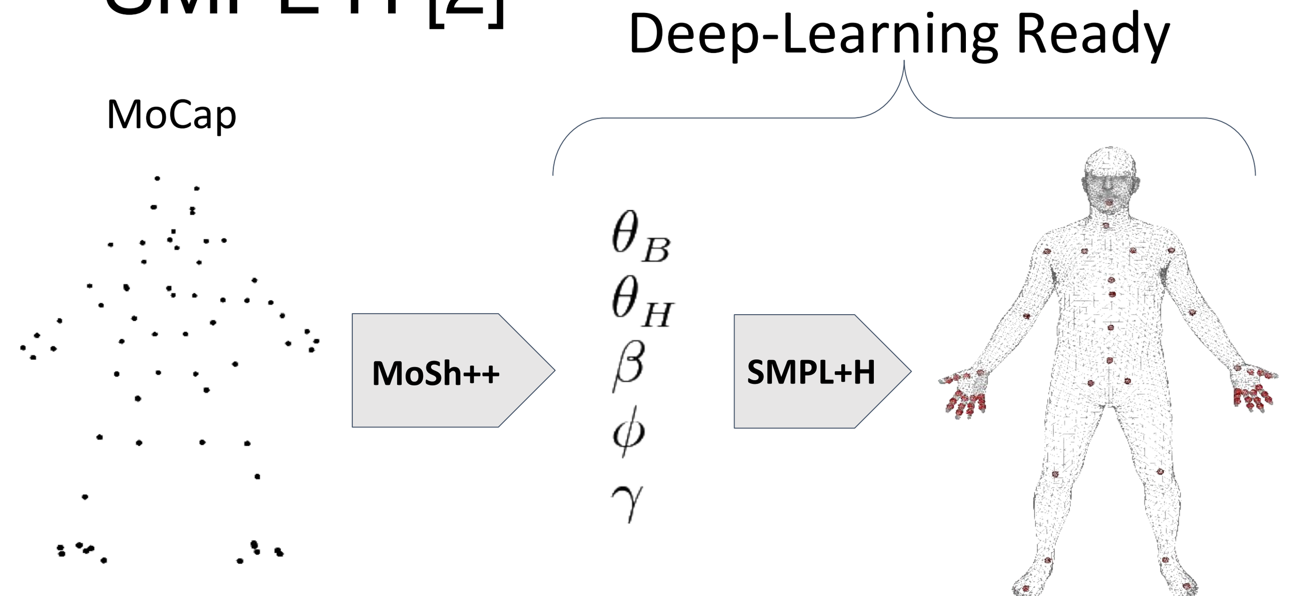
Problem:
- Every MoCap dataset uses different markers and a different representation of body pose

Idea:
- Fit a body model to all MoCap sequences to obtain a single, common parameterization

Approach:
- MoSh++

Contribution:
- Unified dataset with **43 hours of human motion** and growing; already used to train a human pose pior [4]

SMPL-H [2]



Tutorial Code: https://github.com/nghorbani/amass

AMASS Dataset

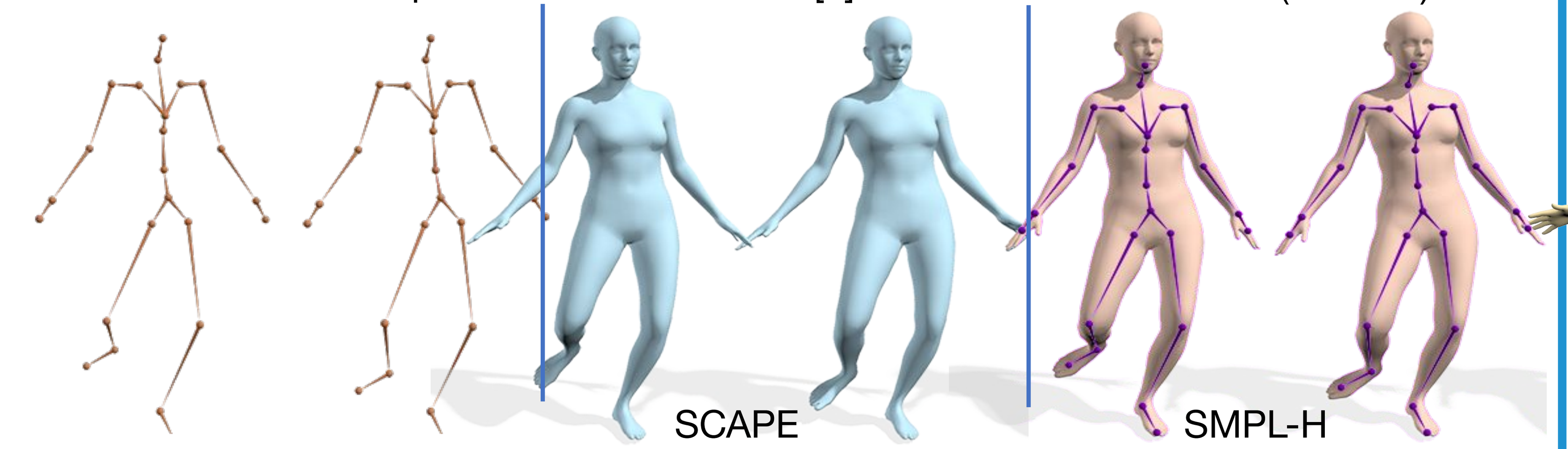
	Markers	Subjects	Motions	Minutes
ACCAD	82	20	252	26.74
BMLmovi New!	67	86	1801	168.99
BMLrub	41	111	3061	522.69
CMU	42	96	1983	543.49
Dfaust Synthetic	67	10	129	10.37
EKUT	51	4	349	30.71
Eyes Japan	37	12	750	363.64
Human Eva	39	3	28	8.48
KIT	100	55	4232	661.84
MPI HDM05	41	4	215	144.54
MPI MoSh	89	19	77	16.53
Pose Prior	53	3	35	20.82
SFU	53	7	44	15.23
SSM	86	3	30	1.87
TCD Hands	85	1	62	8.05
Total Capture	53	5	37	41.10
Transitions	49	1	110	15.10
AMASS (+growing)	-	**440**	**13,195**	**2600.22**

If you had an interesting MoCap dataset we can add it to AMASS.

References

[1] M. Loper et al., MoSh: motion and shape capture from sparse markers, SIGGRAPH Asia, 2014
[2] J. Romero et al., "Embodied Hands: Modeling and Capturing Hands and Bodies Together, SIGGRAPH Asia, 2017
[3] G. Pons-Moll et al., Dyna: A Model of Dynamic Human Shape in Motion, SIGGRAPH Asia 2015
[4] G. Pavlakos et al., Expressive body capture: 3D hands, face, and body from a single image, CVPR 2019
[5] F. Bogo et al., Dynamic FAUST, CVPR 2017

MoSh++



AMASS introduces MoSh++:
- Extends MoSh [1] to use SMPL-H [2]; body model available for research and commercial license, with hand articulation and a body skeleton
- Go beyond traditional MoCap and recover soft tissue dynamics with DMPLs [3]
- Evaluate on a novel dataset: Synchronized MoCap with 4D Scan (SSM)

Model Fitting

MoSh++ Stage I: Optimize shape parameters and marker locations:

$$\begin{aligned}E(\tilde{\mathcal{M}},\beta,\Theta_B,\Theta_H) = {}& \lambda_D E_D(\tilde{\mathcal{M}},\beta,\Theta_B,\Theta_H) \\ &+ \lambda_\beta E_\beta(\beta) + \lambda_{\theta_B} E_{\theta_B}(\theta_B) + \lambda_{\theta_H} E_{\theta_H}(\theta_H) \\ &+ \lambda_R E_R(\tilde{\mathcal{M}},\beta) + \lambda_I E_I(\tilde{\mathcal{M}},\beta).\end{aligned}$$

MoSh++ Stage II: Optimize pose parameters (body, hands, soft-tissue deformation):

$$\begin{aligned}E(\theta_B,\theta_H,\phi) = {}& \lambda_D E_D(\theta_B,\theta_H,\phi) \\ &+ \lambda_{\theta_B} E_{\theta_B}(\theta_B) + \lambda_{\theta_H} E_{\theta_H}(\theta_H) \\ &+ \lambda_u E_u(\theta_B,\theta_H) \\ &+ \lambda_\phi E_\phi(\phi) + \lambda_\nu E_\nu(\phi).\end{aligned}$$

Regularizers:

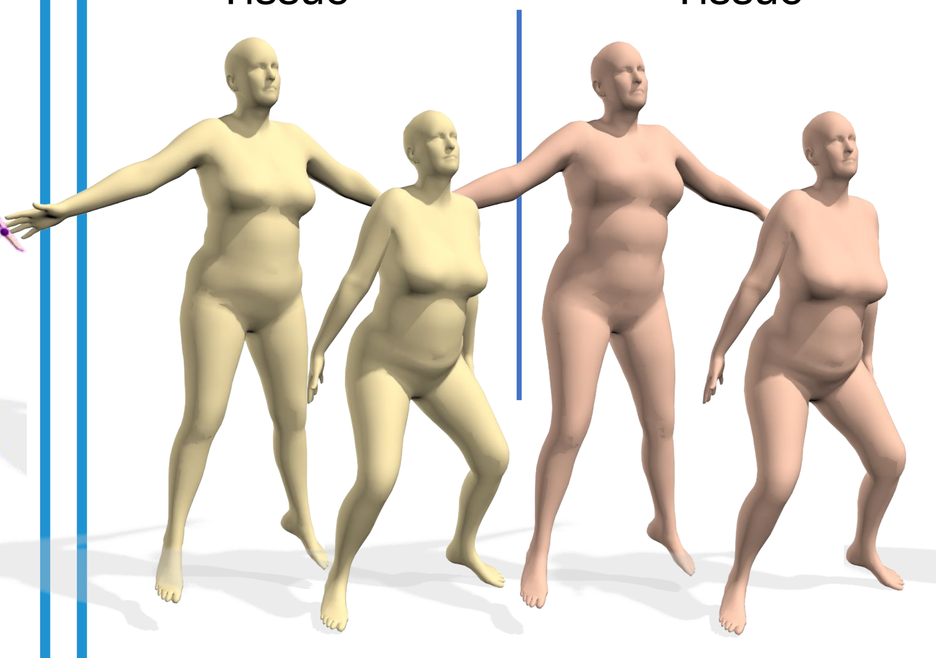
Data: $E_D(.) = \Sigma_{i,t} ||\hat{m}(\tilde{m}_i,\beta,\theta_t) - m_{i,t}||^2$

Shape: $E_\beta(\beta) = \beta^T \Sigma_\beta^{-1} \beta$

Hands: $E_{\theta_H}(\theta_H) = \theta_H^T \Sigma_{\theta_H}^{-1} \theta_H$

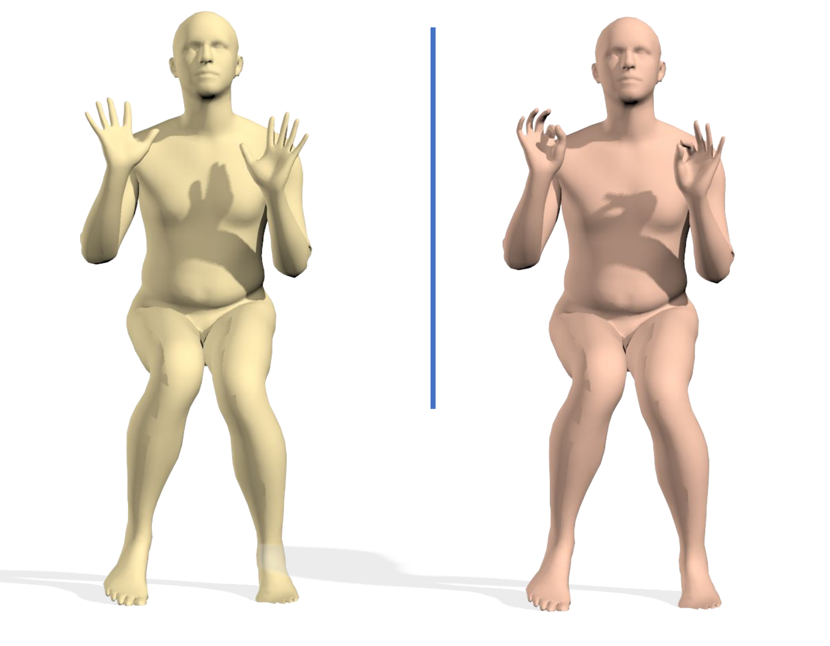
Soft-tissue: $E_\phi(\phi) = \phi_t^T \Sigma_\phi^{-1} \phi_t$

Soft Tissue



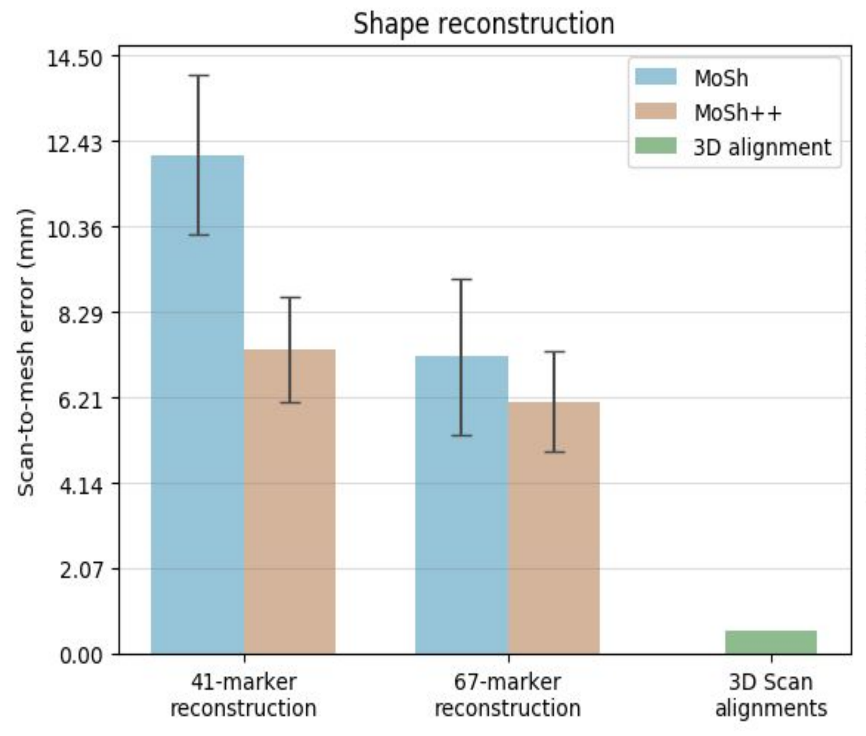
MoSh++ captures motion of the chest and stomach better than MoSh

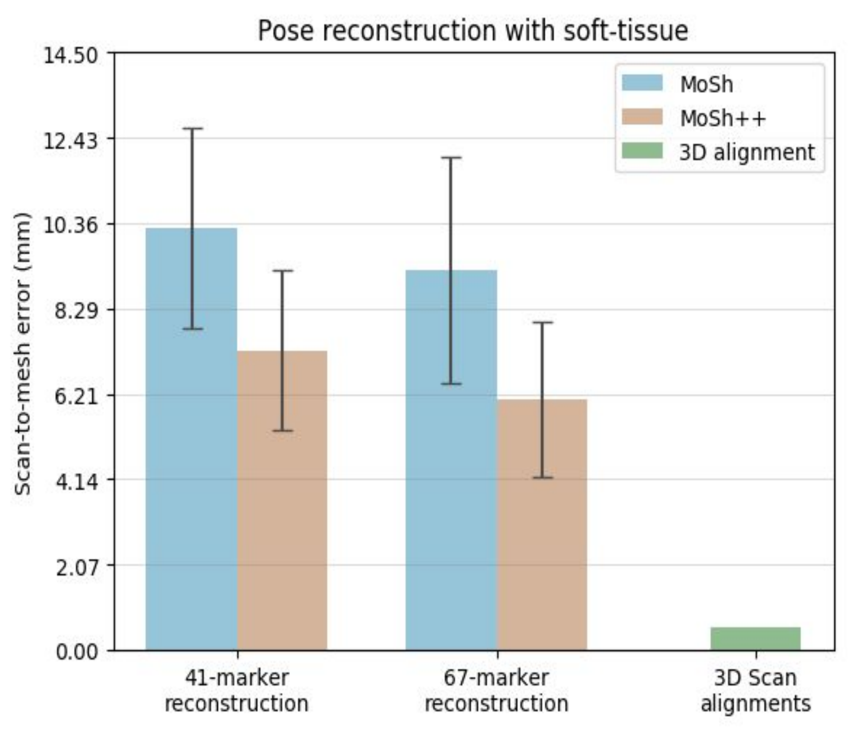
Articulated Hands



If hand markers are present MoSh++ fits hand poses

Evaluation





SSM:
- Novel synchronized MoCap + 4D scans, 30 motions across 3 subjects
- Splits for hyper-parameter optimization & evaluation

Synthetic DFAUST:
- Virtual MoCap from DFAUST meshes [5], 129 motions, 10 subjects
- Used only for evaluation

https://amass.is.tue.mpg.de/