SOMA: Solving Optical Marker-Based MoCap Automatically

Nima Ghorbani, Michael J. Black

MPI for Intelligent Systems Tübingen

soma@tue.mpg.de

MAX PLANCK INSTITUTE FOR INTELLIGENT SYSTEMS

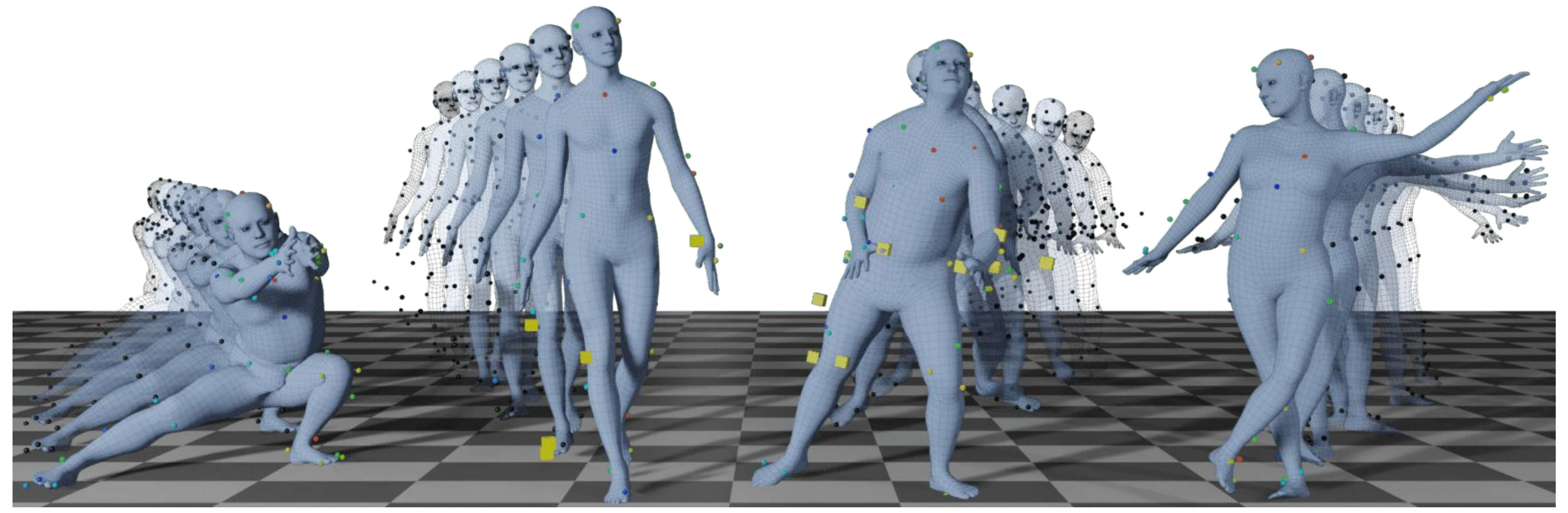
2021 ICCV OCTOBER 11-17 VIRTUAL

Overview

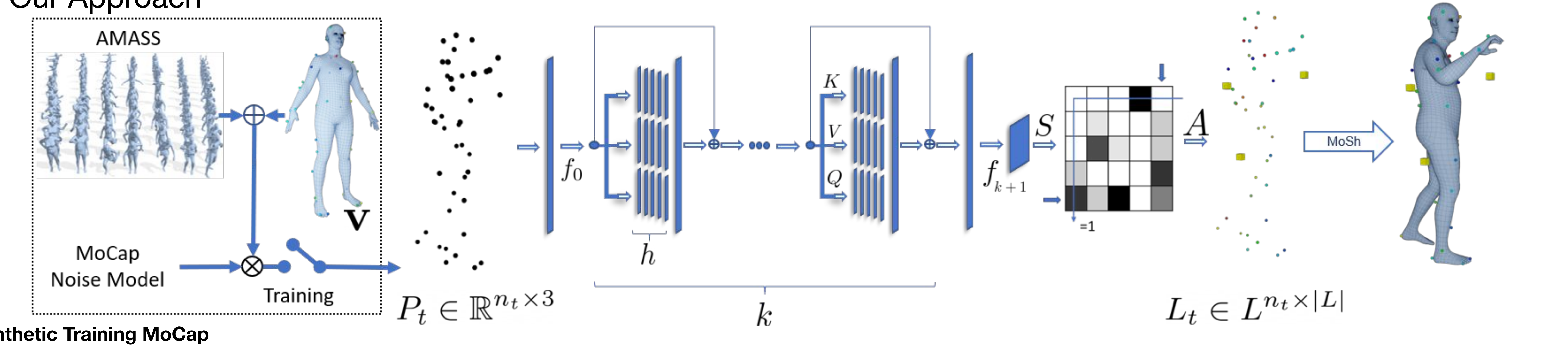
Goal: Automatically *label* raw mocap point clouds (MPC) and solve SMPL-X bodies.

Problem: A MPC has missing data, noise, and “ghost” points. The assignment of points to labels is unknown. Existing methods are constrained to small range of motions, a single body shape, a certain capture scenario, a special marker layout, or require a subject-specific calibration sequence.

Solution: A robust, automatic mocap solving pipeline that works with archival data, different mocap technologies, poor data quality, and varying subjects and motions.



Our Approach



Synthetic Training MoCap

Place virtual markers on synthetic SMPL-X [2] bodies of AMASS [3],

$$\text{SMPL-X}\ \ (\theta_b, \beta, \gamma) : \mathbb{R}^{|\theta_b| \times |\beta| \times |\gamma|} \to \mathbb{R}^{3N}.$$

Augment synthetic markers with controllable noise.

Self-attention on Sparse 3D Point Cloud

Layers of **Transformer** [4] elements process varying number of mocap points and produce a score matrix S.

Constrained Inexact Matching

Normalization to encourage bijective label-point correspondence.
Unassigned points and labels are set to null by the normalization layer.

Training Loss

Log-likelihood of augmented assignment matrix, and its ground truth.
Down-weight the influence of the over-represented class by the reciprocal of its occurrence frequency.

$$\mathbb{L}_A = \frac{-1}{\sum_{i,j} G'_{i,j}} \sum_{i,j} W_{i,j} \cdot G'_{i,j} \cdot log(A'_{i,j})$$

References

[1] M. Loper et al., MoSh: motion and shape capture from sparse markers, SIGGRAPH Asia, 2014

[2] G. Pavlakos et al., Expressive body capture: 3D hands, face, and body from a single image, CVPR 2019

[3] N. Mahmood et al.,AMASS, Archive of Motion Capture as Surface Shapes, ICCV 2019

[4] Vaswani et al. Attention is all you need. NIPs 2017

Experiments

Ablation Study

Version	Accuracy	F1
Base	95.50 ± 5.33	94.66 ± 6.03
- AMASS Noise Model	94.73± 5.52	93.73 ± 6.25
- CAESAR bodies	95.21 ± 6.83	94.31 ± 7.57
- Log-Softmax Instead of Sinkhorn	91.51 ± 10.69	90.10 ± 11.47
- Random Marker Placement	89.41 ± 8.06	87.78 ± 8.85
- Transformer	11.36 ± 6.54	7.54 ± 6.22

Comparison With Previous Work

Method	Number of Exact Per-Frame Occlusions						
	0	1	2	3	4	5	5+G
Holzreiter et al.	88.16	79.00	72.42	67.16	61.13	52.10	—
Maycock et al.	83.19	79.35	76.44	74.91	71.17	65.83	—
Ghorbani et al.	97.11	96.56	96.13	95.87	95.75	94.90	—
SOMA-Real	**99.08**	**98.97**	**98.85**	**98.68**	**98.48**	**98.22**	**98.29**
SOMA-Synthetic	99.16	98.92	98.54	98.17	97.61	97.07	95.13
SOMA*	98.38	98.28	98.17	98.03	97.86	97.66	97.56

MoCap Solving Performance

Name	# Frames	# Motions	Acc.	F1	$V2V_{mm}^{mean}$	$V2V_{mm}^{median}$
Clap	7572	6	100.00 ± 0.00	100.00 ± 0.00	0.00 ± 0.08	0.00
Dance	15023	8	99.78 ± 0.87	99.68 ± 1.29	0.15 ± 1.76	0.00
Jump	9621	6	99.99 ± 0.13	99.99 ± 0.25	0.03 ± 0.72	0.00
Kick	10787	6	99.59 ± 1.18	99.48 ± 1.50	0.75 ± 6.92	0.00
Lift	16932	6	100.00 ± 0.00	100.00 ± 0.00	0.00 ± 0.06	0.00
Random	19617	7	100.00 ± 0.05	100.00 ± 0.09	0.00 ± 0.21	0.00
Run	9356	6	100.00 ± 0.00	100.00 ± 0.00	0.00 ± 0.06	0.00
Sit	9829	6	100.00 ± 0.00	100.00 ± 0.00	0.00 ± 0.09	0.00
Squat	11287	6	100.00 ± 0.00	100.00 ± 0.00	0.01 ± 0.13	0.00
Throw	9292	6	99.99 ± 0.15	99.99 ± 0.22	0.00 ± 0.09	0.00
Walk	12264	6	100.00 ± 0.00	100.00 ± 0.00	0.00 ± 0.11	0.00
	131580	69	99.94 ± 0.47	99.92 ± 0.64	0.08 ± 2.09	0.00