Momanifold: Learning to Measure 3D Human Motion via Decoupled Joint Acceleration Manifolds

Ziqiang Dang*[1], Tianxing Fan*[1], Boming Zhao[1], Xujie Shen[1], Lei Wang[2], Guofeng Zhang[1], Zhaopeng Cui†[1]

[1] State Key Lab of CAD&CG, Zhejiang University [2] OPPO

* Equal Contribution † Corresponding Author

Abstract

Incorporating temporal information effectively is important for accurate 3D human motion estimation and generation which have wide applications from human-computer interaction to AR/VR. In this paper, we present MoManifold, a novel human motion prior, which models plausible human motion in continuous high-dimensional motion space. Different from existing mathematical or VAE-based methods, our representation is designed based on the neural distance field, which makes human dynamics explicitly quantified to a score and thus can measure human motion plausibility. Specifically, we propose novel decoupled joint acceleration manifolds to model human dynamics from existing limited motion data. Moreover, we introduce a novel optimization method using the manifold distance as guidance, which facilitates a variety of motion-related tasks. Extensive experiments demonstrate that MoManifold outperforms existing SOTAs as a prior in several downstream tasks such as denoising real-world human mocap data, recovering human motion from partial 3D observations, mitigating jitters for SMPL-based pose estimators, and refining the results of motion in-betweening.

Highlights of this work

- We present a novel human motion prior, i.e., MoManifold, which models plausible human motion in a continuous high-dimensional motion space and can be used to measure human motion plausibility, thus facilitating downstream tasks such as denoising real-world human mocap data, recovering human motion from partial 3D observations, jitter mitigation for human pose estimators and refining the results of motion in-betweening.
- Decoupled joint acceleration manifolds and a weighted design based on human skeleton geometry are adopted to model human dynamics to deal with the dramatic demand for human motion training data.
- We introduce a novel motion optimization method based on MoManifold, which can be applied to various downstream tasks.
- Extensive experiments demonstrate that MoManifold has good generalization ability and outperforms existing SOTAs on multiple motion-related tasks.

Methodology

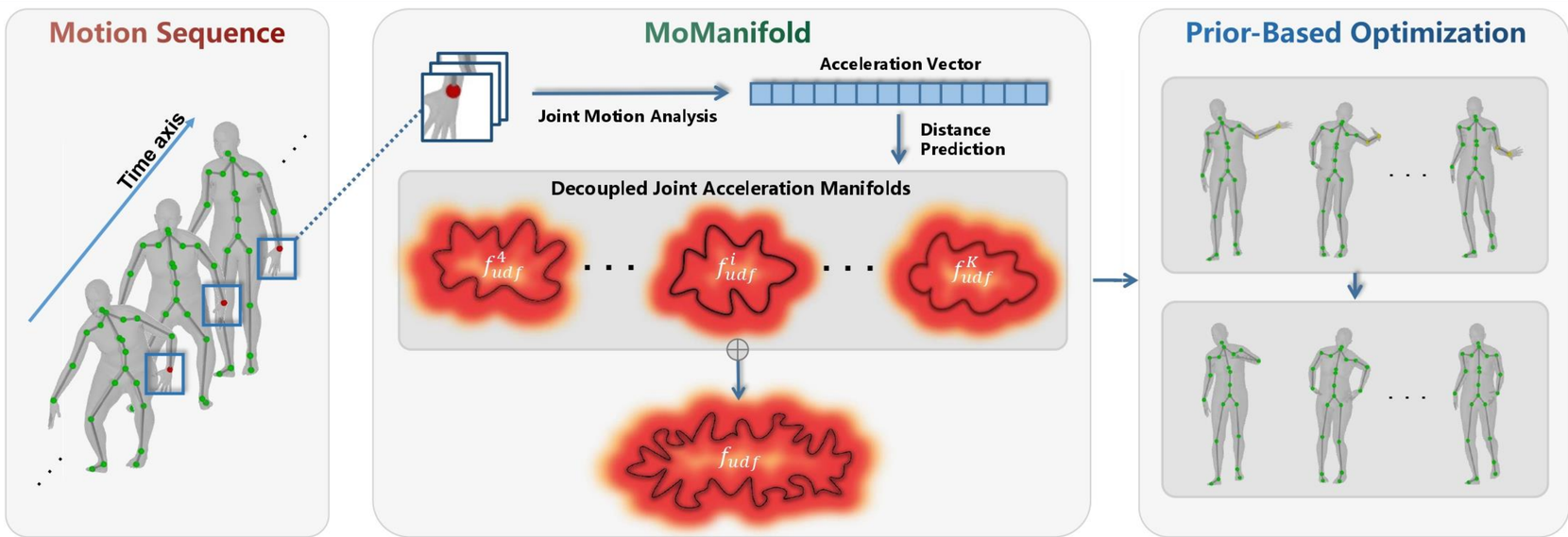

Figure 1: A motion sequence can be divided into different motion segments, represented as displacement vectors of body joints. Instead of directly learning the implicit surface of motion segment ***m***, Momanifold learns an independent implicit surface of plausible acceleration vectors for every joint, and the distance to the manifold measures whether the joint motion complies with human dynamics. With a weighted design based on skeleton, we combine these manifolds to obtain the manifold of motion segments.

After modeling human motion as an unsigned distance field, we can utilize Momanifold as a motion prior for downstream tasks. Here, we introduce a novel optimization method that employs the distance value as a guiding metric for optimization and integrates with a traditional temporal term.

$$\varepsilon_{motion} = f_{udf}(\ddot{\boldsymbol{m}})$$

$$\varepsilon_{fusion} = \varepsilon_{motion} + \sum_{i=1}^{K}\sum_{t=1}^{T}(1-w_i)\left\|\boldsymbol{p}_t^i - \boldsymbol{p}_{t-1}^i\right\|_2$$

Correlation Analysis

Joint	Pearson	Joint	Pearson	Joint	Pearson
leftKnee	0.9869	neck	0.9631	rightElbow	0.9913
rightKnee	0.9863	leftCollar	0.9455	leftWrist	0.9946
leftAnkle	0.9890	rightCollar	0.9638	rightWrist	0.9946
rightAnkle	0.9886	leftShoulder	0.9700	leftHand	0.9908
leftFoot	0.9829	rightShoulder	0.9759	rightHand	0.9910
rightFoot	0.9810	leftElbow	0.9903		

Table 1: Pearson Correlation Coefficient. All Pearson coefficients are very close to 1, indicating strong linear correlations between manifold distances and acceleration error.

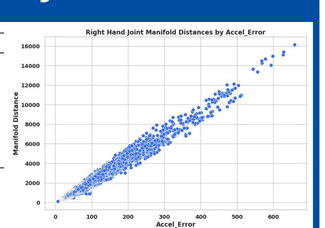

Figure 2: Scatter plot of right hand joint.

Motion Denoising & Fitting to Partial Data

Data	Noisy HPS			Noisy AMASS		
# frames	60	120	240	60	120	240
VPoser-t [34]	3.05	4.43	7.11	5.83	6.55	7.86
HuMoR [37]	6.08	12.67	-	10.28	12.63	-
Pose-NDF [43]	1.17	1.30	1.16	5.03	5.39	5.49
Ours	**0.90**	**0.91**	**0.88**	**1.45**	**1.47**	**1.60**

Table 2: Motion Denoising. We compare PVE in cm.

Data	Occ. Leg		Occ. Arm +Hand		Occ. Shoulder +Upper Arm	
# frames	60	120	60	120	60	120
VPoser-t [34]	8.69	10.77	8.79	10.70	8.74	10.20
HuMoR [37]	9.52	12.70	9.39	13.82	9.02	12.14
Pose-NDF [43]	8.50	9.40	8.66	9.43	8.73	9.47
Ours	**4.83**	**5.07**	**4.83**	**5.01**	**4.93**	**5.04**

Table 3: Fitting to Partial Data. We compare PVE (in cm) on test set of AMASS.

Mitigating Jitters for SMPL-based Pose Estimators

Method	3DPW			
	MPJPE ↓	PA-MPJPE ↓	PVE ↓	Accel ↓
SPIN [24]	99.29	61.71	113.32	34.95
SPIN w/ S [51]	97.81	61.19	111.5	**7.4**
SPIN w/ ours	**97.24**	**60.80**	**111.37**	8.43
EFT [20]	91.6	55.33	110.17	33.38
EFT w/ S [51]	89.57	54.40	**107.66**	**7.89**
EFT w/ ours	**89.35**	**53.83**	107.82	8.94
PARE [23]	79.93	48.74	94.07	26.45
PARE w/ S [51]	78.68	48.47	**92.5**	**6.31**
PARE w/ ours	**78.55**	**47.84**	92.65	7.63
VIBE* [22]	84.28	54.93	99.10	23.59
VIBE* w/ S [51]	83.46	54.83	98.04	**7.42**
VIBE* w/ ours	**83.07**	**54.28**	**97.8**	8.01
TCMR* [7]	88.47	55.70	103.22	7.13
TCMR* w/ S [51]	88.69	56.61	103.40	**6.48**
TCMR* w/ ours	**88.28**	**55.69**	**103.02**	6.72

Table 4: Mitigating Jitters on 3DPW. "w/S“ indicates using SmoothNet. "*"denotes spatio-temporal backbones.

Method	AIST++			
	MPJPE ↓	PA-MPJPE ↓	PVE ↓	Accel ↓
VIBE* [22]	107.41	72.83	127.56	31.65
VIBE* w/ S [51]	105.21	**70.74**	124.78	**6.34**
VIBE* w/ ours	**104.85**	71.60	**124.60**	7.92
TCMR* [7]	106.95	71.58	124.73	6.47
TCMR* w/ S [51]	107.19	**71.43**	124.76	**4.70**
TCMR* w/ ours	**106.51**	71.56	**124.20**	5.29

Table 5: Mitigating Jitters on AIST++.

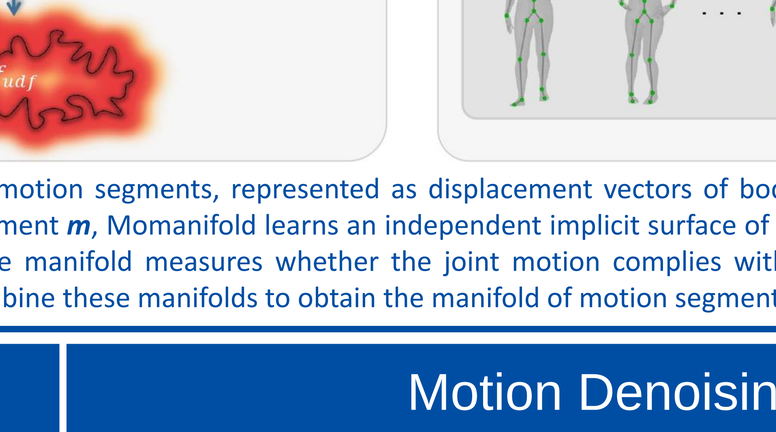

Figure 3: Qualitative Comparison with SmoothNet.

Method	3DPW					
	Left- Leg	Left- Foot	Left- ToeBase	Right- Leg	Right- Foot	Right- ToeBase
VIBE* [22]	99.73	137.11	144.40	101.13	139.86	149.85
VIBE* w/ S [51]	99.76	137.43	144.50	101.20	140.40	150.00
VIBE* w/ ours	**98.66**	**136.21**	**143.65**	**99.86**	**138.70**	**148.93**
TCMR* [7]	99.72	140.21	148.60	101.94	142.29	**152.56**
TCMR* w/ S [51]	100.19	141.18	149.48	103.05	144.25	154.31
TCMR* w/ ours	**99.54**	**140.02**	**148.42**	**101.84**	**142.28**	152.57

Table 6: Mitigating Jitters of Legs and Feet.

Motion In-betweening Refinement

Method	NPSS ↓			Accel ↓		
frames	15	30	45	15	30	45
Two-stage [36]	0.06	0.28	0.68	11.97	11.10	10.47
Two-stage w/ ours	0.06	0.28	0.68	**7.71**	**8.03**	**8.08**

Table 7: Refining Motion In-betweening on LAFAN1. "frames" refers to the number of frames of the generated transitions.

Conclusion

This paper presents a novel human motion prior Momanifold that models plausible human motions in continuous high-dimensional motion space with decoupled joint acceleration manifolds. Extensive experiments demonstrate that Momanifold has good generalization ability and outperforms existing SOTAs on multiple motion-related tasks.Although the relationship between joints is implicitly established through the SMPL tree structure, such relationship is relatively weak. Therefore, as future work, we will explore how to establish explicit relationships between joints under the representation of neural distance field.