南京航空航天大學
Nanjing University of Aeronautics and Astronautics

HMD-Poser: On-Device Real-time Human Motion Tracking from Scalable Sparse Observations

JOSE Peng Dai,Yang Zhang,Tao Liu,Zhen Fan,Tianyuan Du,

Zhuo Su,Xiaozheng Zheng,Zeming Li,

PICO,ByteDance

CVPR2024

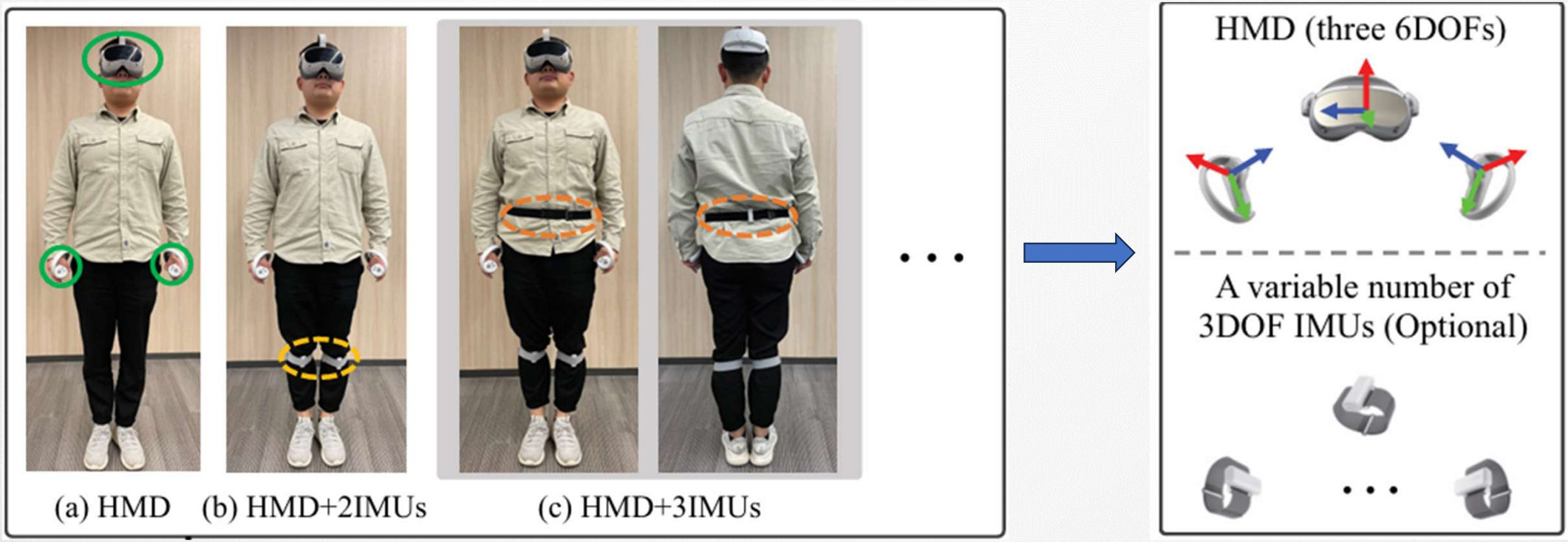
Introduction

Sparse Observations

Motion Tracking

Virtual Avatar

Introduction



Scalable Sparse Observations

Introduction

For only HMD(including hand controllers):

When estimating the user’ s lower-body motions, is inherently an **under-constrained problem** with such sparse tracking signals.

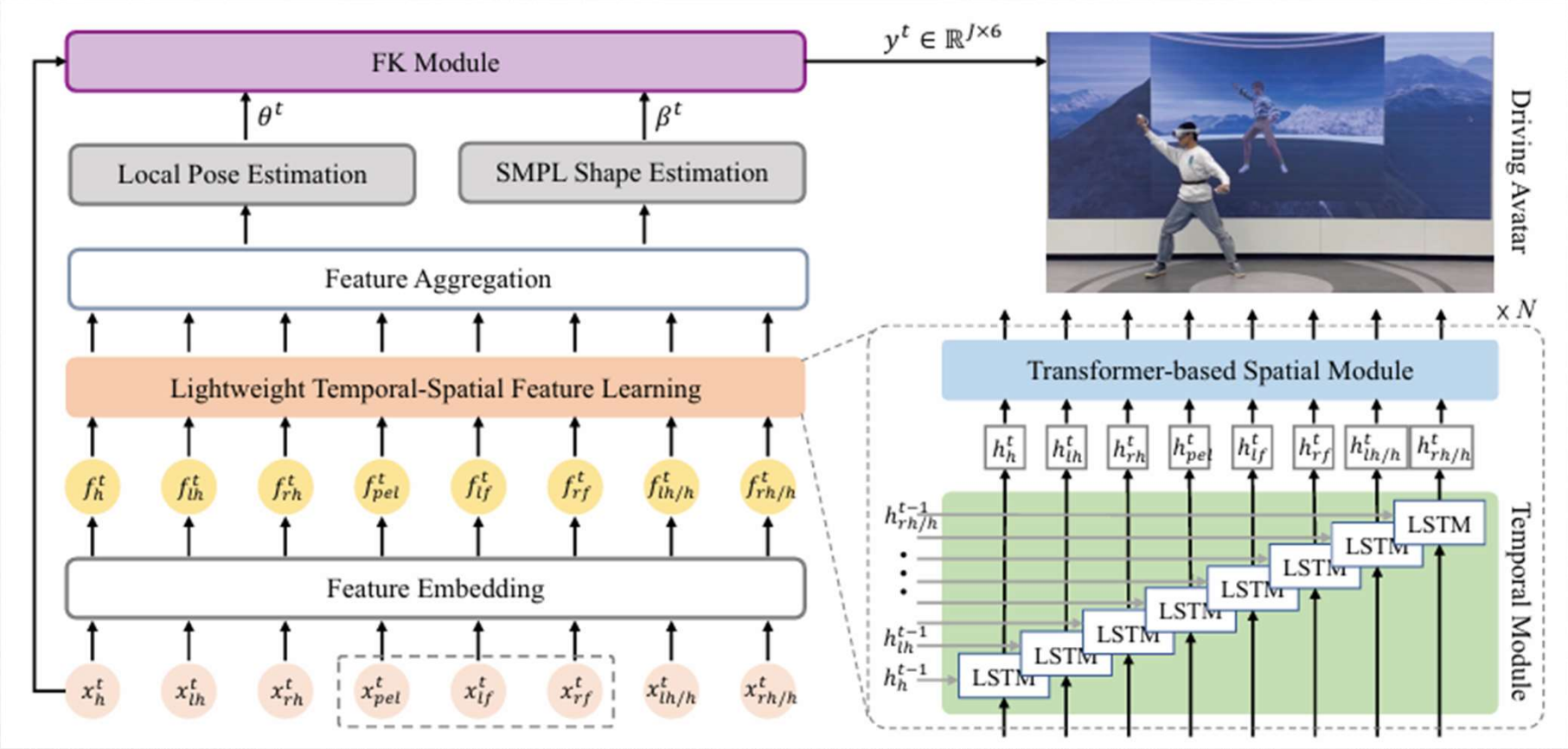
For IMU sensors:

Prone to **positional drift** due to the **inevitable accumulation errors** of IMU sensors

Introduction

Driving an Avatar on HMD in Real-time

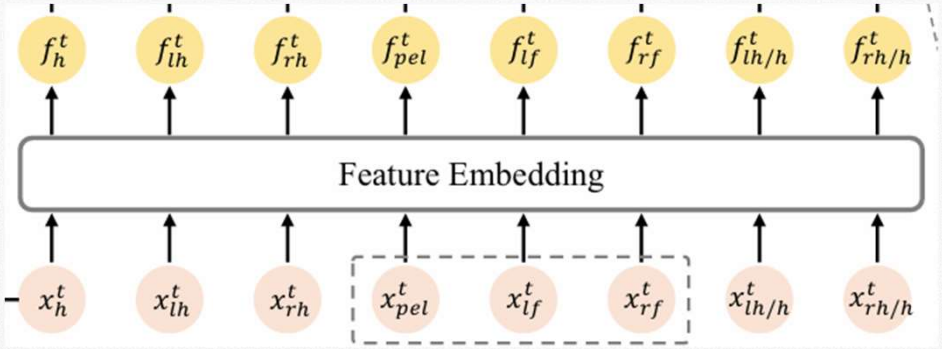
Method



Overview of HMD-Poser

Method

Feature Embedding Module



Input :

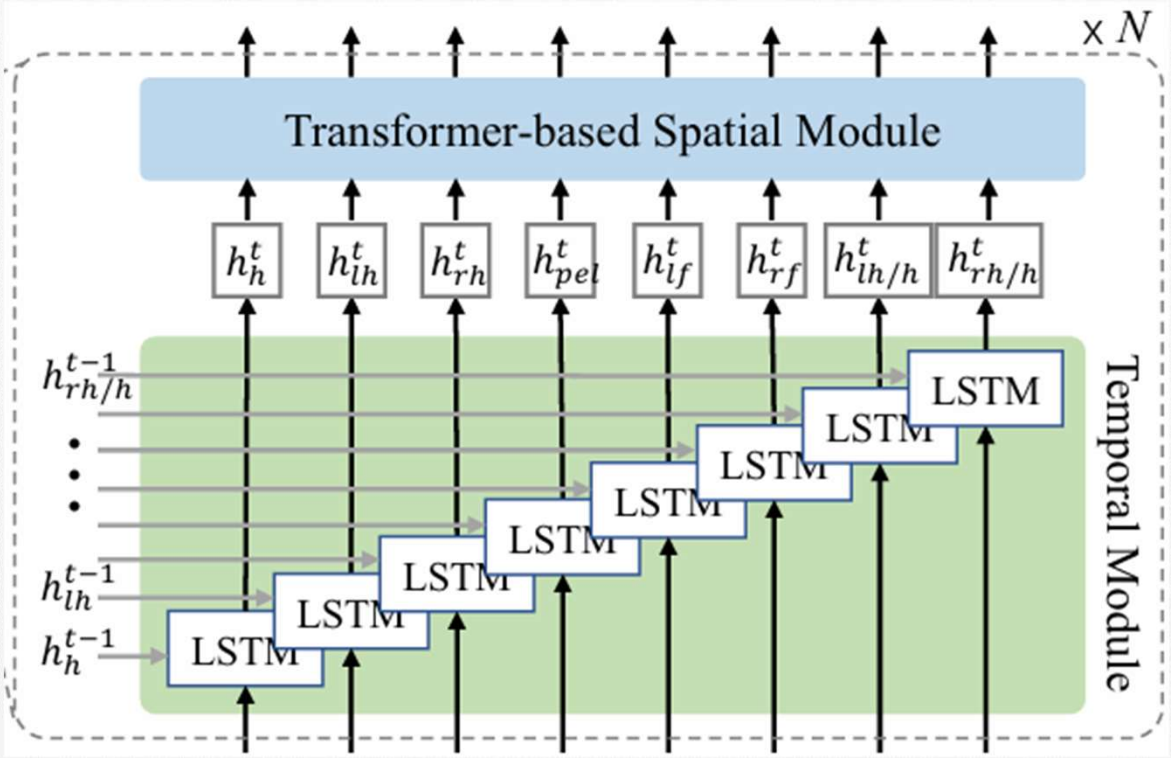
Use a concatenated vector of position, linear velocity, rotation, and angular velocity to obtain the representation for the **head** x_h^t, the **left hand** x_{lh}^t, and the **right hand** x_{rh}^t .

adopt a concatenated vector of rotation, angular velocity, and acceleration to obtain the representation for the **pelvis** x_{pel}^t, the **left leg** x_{lf}^t, and the **right leg** x_{rf}^t.

$$x^t = [x_h^t, x_{lh}^t, x_{rh}^t, x_{pel}^t, x_{lf}^t, x_{rf}^t, x_{lh/h}^t, x_{rh/h}^t]$$

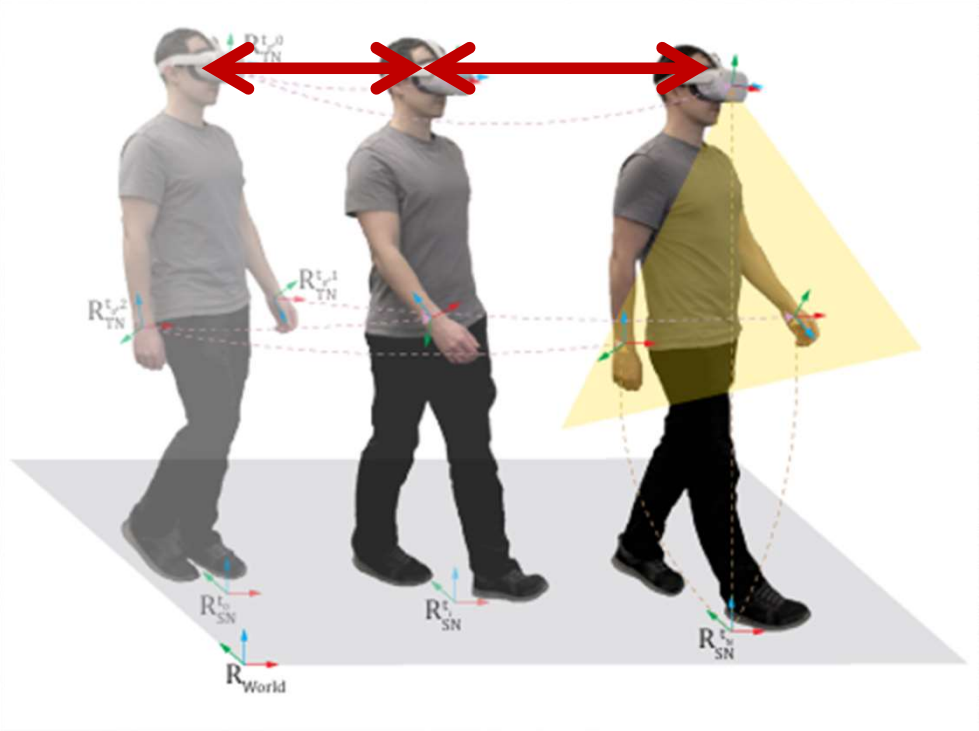
Method

lightweight temporal-spatial feature learning (TSFL) network

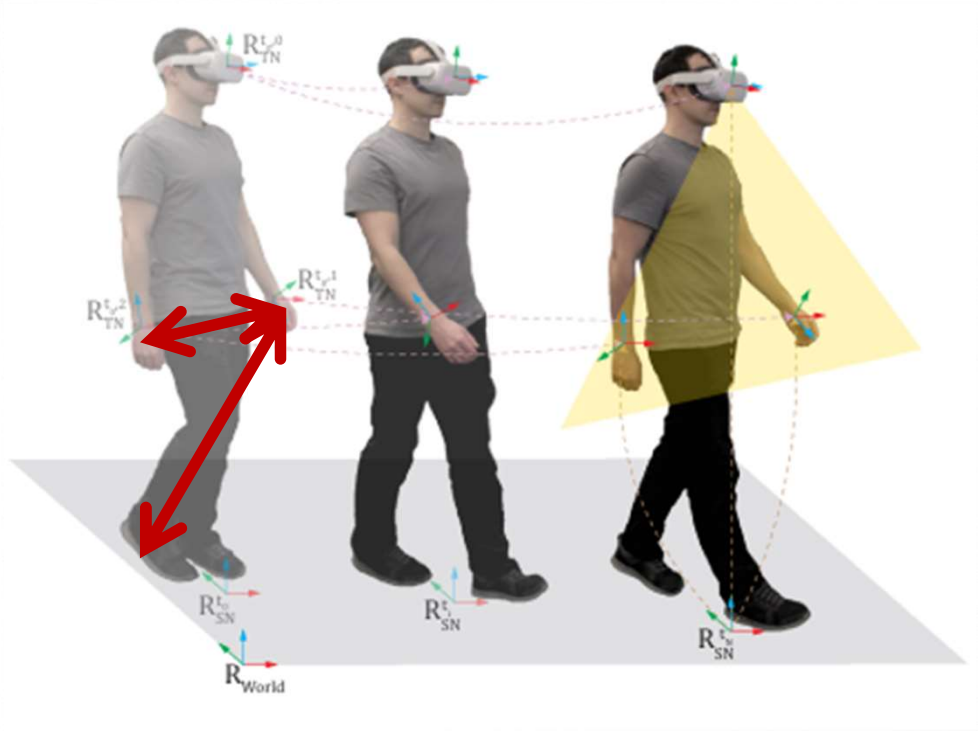


Method

temporal correlation

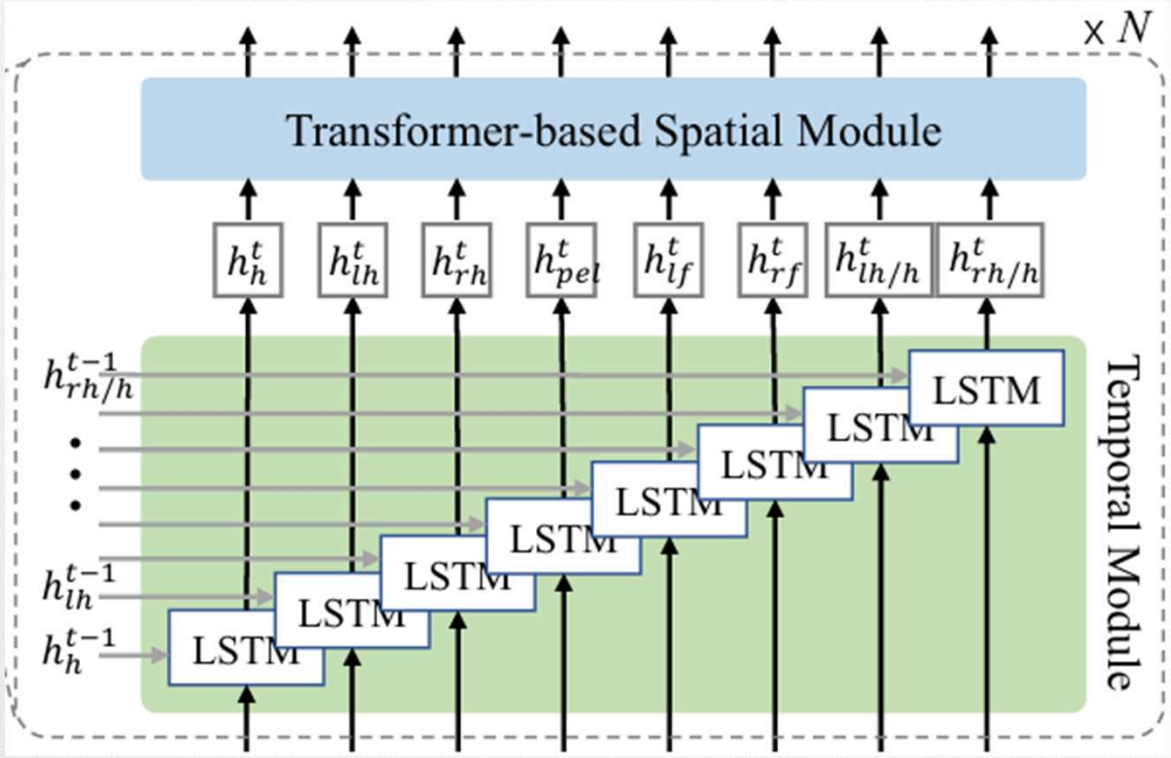


spatial correlation



Method

lightweight temporal-spatial feature learning (TSFL) network



LSTM: learn temporal correlation

Transformer: learn spatial correlation

Time Complexity

Only Transformer : $o(M^2d + Md^2)$

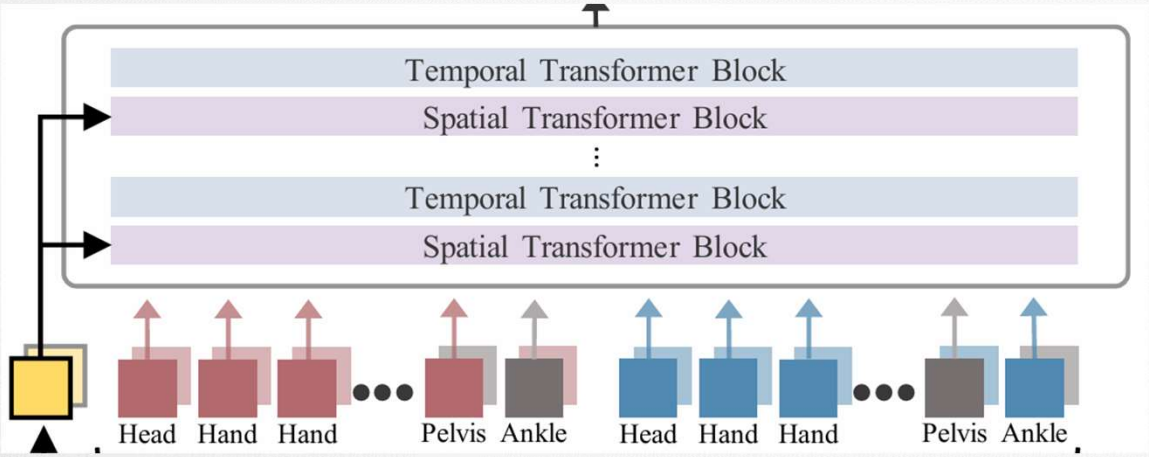
TSFL: $o(d^2)$

M :sequence of length

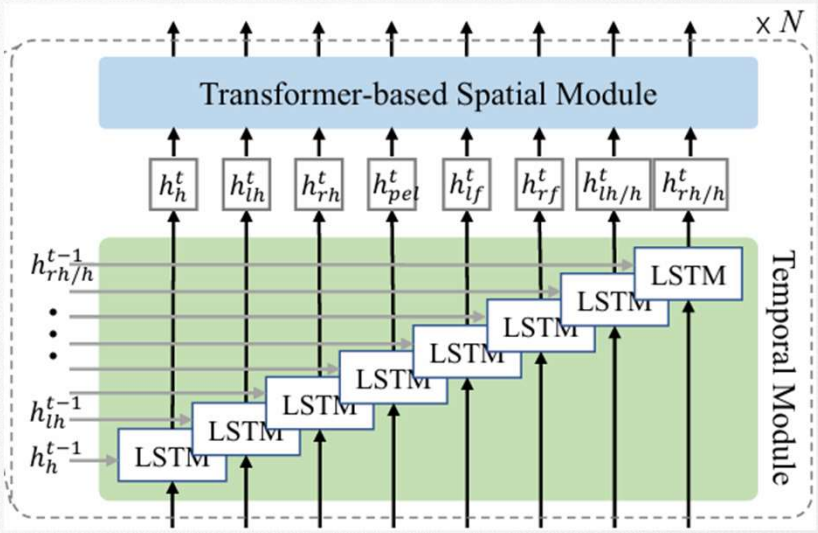
d :dimension of the hidden state

Method

Only Transformer : $o(M^2d + Md^2)$

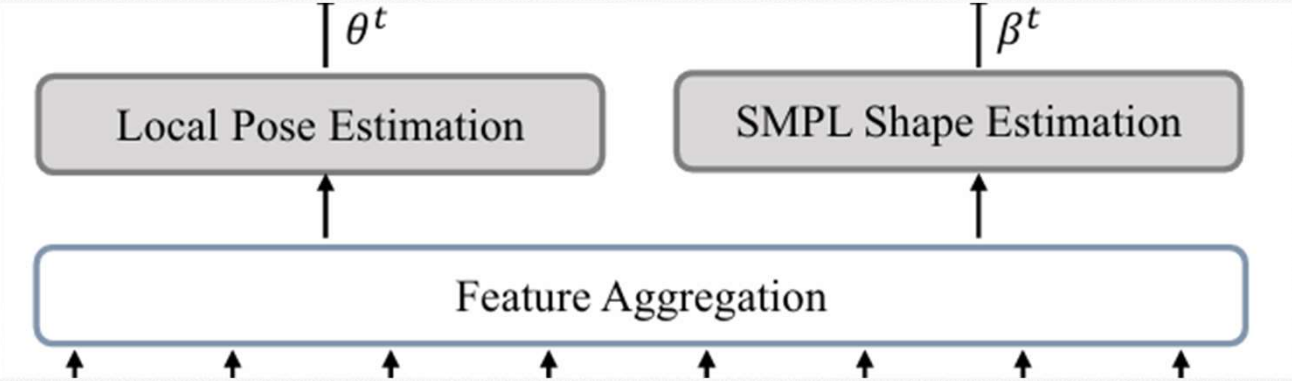


TSFL: $o(d^2)$



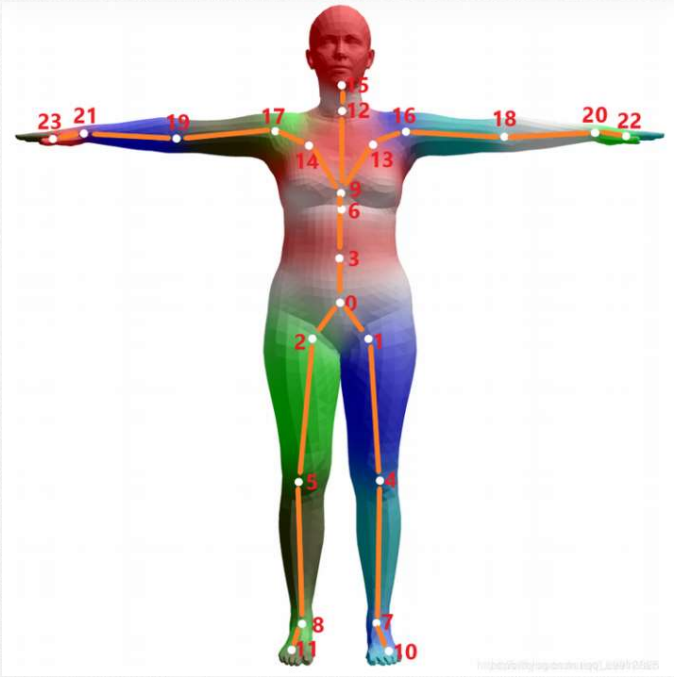
Method

regress the local pose parameters θ and the shape parameters β

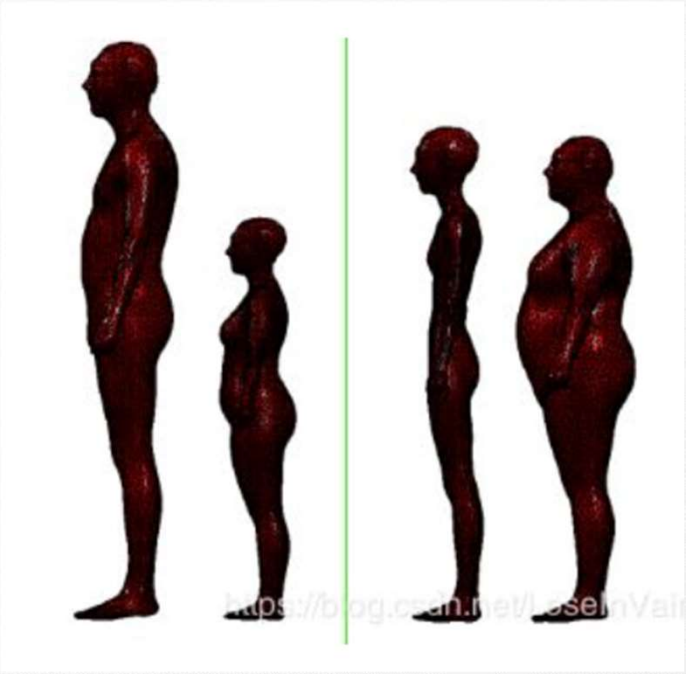


Method

pose parameters θ



shape parameters β



Skinned Multi-Person Linear (SMPL) Model
参数化人体蒙皮模型

Method

Loss Design

Overall Loss : $L = \alpha_{ori}L_{ori} + \alpha_{lrot}L_{lrot} + \alpha_{grot}L_{grot} + \alpha_{joint}L_{joint} + \alpha_{smooth}L_{smooth}$

Where $L_{ori}, L_{lrot}, L_{grot}, L_{joint} and L_{smooth}$ are root orientation loss, local pose loss, global pose loss, joint position loss and joint position loss

set α_{ori}, α_{lrot}, α_{grot}, α_{joint} and α_{smooth} to 1.0 , 5.0 , 1.0 , 1.0 and 0.5, respectively

Smooth Loss : $L_{smooth} = \frac{1}{(T-2)\times(3J)} \sum_{t=1}^{T-1} \sum_{i=0}^{3J} \left| a_i^t - \hat{a}_i^t \right|_1$

Where a_i^t and $\hat{a}_i^t$ are the computed and the ground-truth acceleration at time t, respectively, and T is the sequential length in the training and J is the number of joints.

Experiments

Quantitative.

Method	MPJRE↓	MPJPE↓	MPJVE↓	Jitter↓	H-PE↓	U-PE↓	L-PE↓	R-PE↓
AvatarPoser [16]	2.94	5.84	26.60	13.97	4.58	3.24	9.59	5.05
AGRoL [9]	2.70	5.73	19.08	7.65	4.29	3.16	9.44	5.15
AvatarJLM [59]	2.81	5.03	20.91	6.94	2.01	3.00	7.96	4.58
Transpose [50]	3.05	4.57	22.41	7.98	3.83	3.05	6.76	4.62
PIP [51]	2.45	4.54	19.02	8.13	4.54	3.15	6.53	4.54
HMD-Poser: HMD	2.28	3.19	17.47	6.07	1.65	1.67	5.40	3.02
HMD-Poser: HMD+2IMUs	1.83	2.27	13.28	5.96	1.39	1.51	3.35	2.74
HMD-Poser: HMD+3IMUs	**1.73**	**1.89**	**11.03**	**5.35**	**1.27**	**1.46**	**2.46**	**2.37**

Experiments

Quantitative.

Method	MPJRE↓	MPJPE↓	MPJVE↓	Jitter↓	H-PE↓	U-PE↓	L-PE↓	R-PE↓
AvatarPoser [16]	4.68	6.62	33.16	10.79	3.93	2.97	11.89	5.30
AGRoL [9]	4.38	6.74	24.14	6.33	3.53	3.02	12.11	5.86
AvatarJLM [59]	4.45	5.96	27.50	6.91	2.30	2.97	10.28	5.22
Transpose [50]	4.31	5.29	28.18	5.16	7.38	3.86	7.36	4.80
PIP [51]	3.61	4.16	22.22	6.89	4.28	2.97	5.89	4.30
HMD-Poser: HMD	4.27	5.44	30.15	5.62	2.56	2.44	9.77	4.83
HMD-Poser: HMD+2IMUs	3.66	3.68	20.29	6.22	**1.65**	**2.14**	5.92	4.51
HMD-Poser: HMD+3IMUs	**3.49**	**3.13**	**16.17**	**4.93**	1.81	2.17	**4.51**	**3.88**

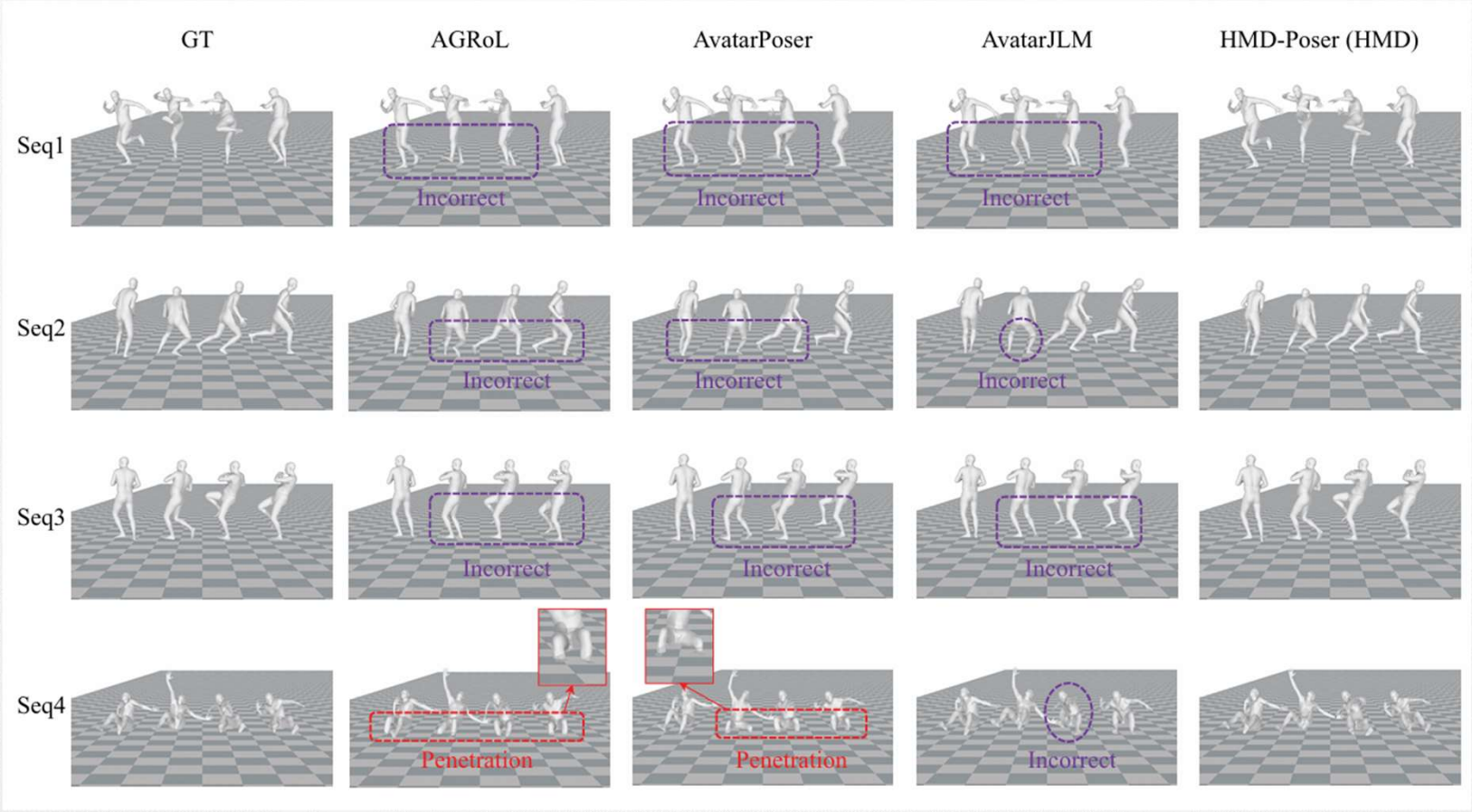
Experiments

Quantitative.

Method	FPS (GPU)↑	FPS (HMD)↑
AvatarPoser [16]	114.1	-
AGRoL [9]	60.8	-
AvatarJLM [59]	1.9	-
Transpose [50]	123.0	-
PIP [51]	62.5	-
HMD-Poser (Ours)	**205.7**	**90.0**

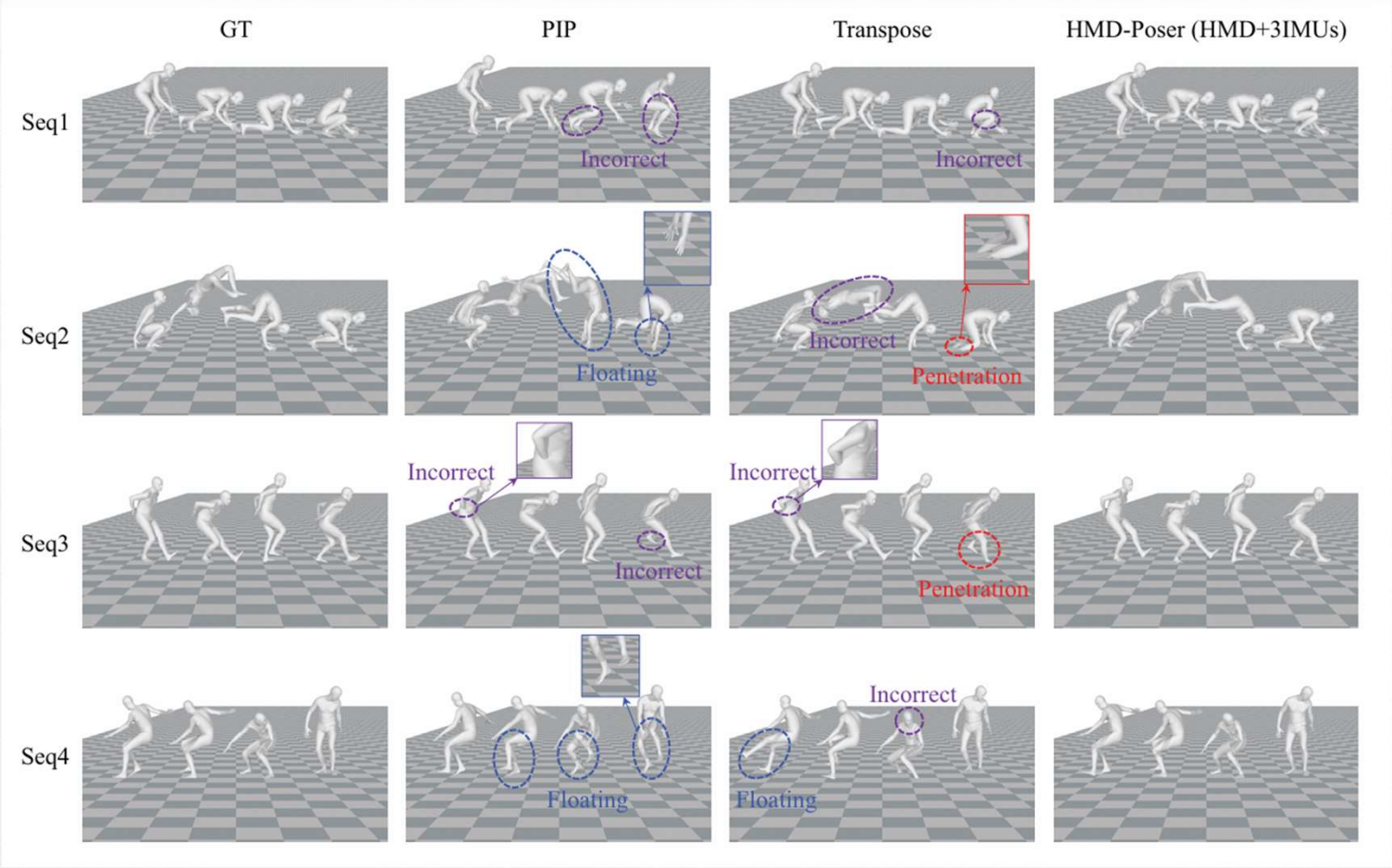
Experiments

Qualitative



Experiments

Qualitative



Experiments

Ablation Study

Method	MPJRE	MPJPE	H-PE	Jitter
w/o $\{x^t_{lh/h}, x^t_{rh/h}\}$	2.45	3.43	2.36	6.25
with $\{x^t_{lh/h}, x^t_{rh/h}\}$	**2.28**	**3.19**	**1.65**	**6.07**

Table 3. Evaluating the effect of adding hand representations relative to the head coordinate frame to input representation.

Method	MPJRE	MPJPE	H-PE	Jitter
w/o ShapeHead	2.32	5.08	4.25	6.11
with ShapeHead	**2.28**	**3.19**	**1.65**	**6.07**

Table 4. Evaluating the effect of the shape regression head. The default shape is used when there is no shape regression head.

Thank you for watching!