NEURAL INFORMATION PROCESSING SYSTEMS

WASP WALLENBERG AI, AUTONOMOUS SYSTEMS AND SOFTWARE PROGRAM

Optimal-state Dynamics Estimation for Physics-based Human Motion Capture from Videos




Cuong Le [1], Viktor Johansson [1], Manon Kok [2], Bastian Wandt [1]

[1] Department of Electrical Engineering, Linköping University, Sweden

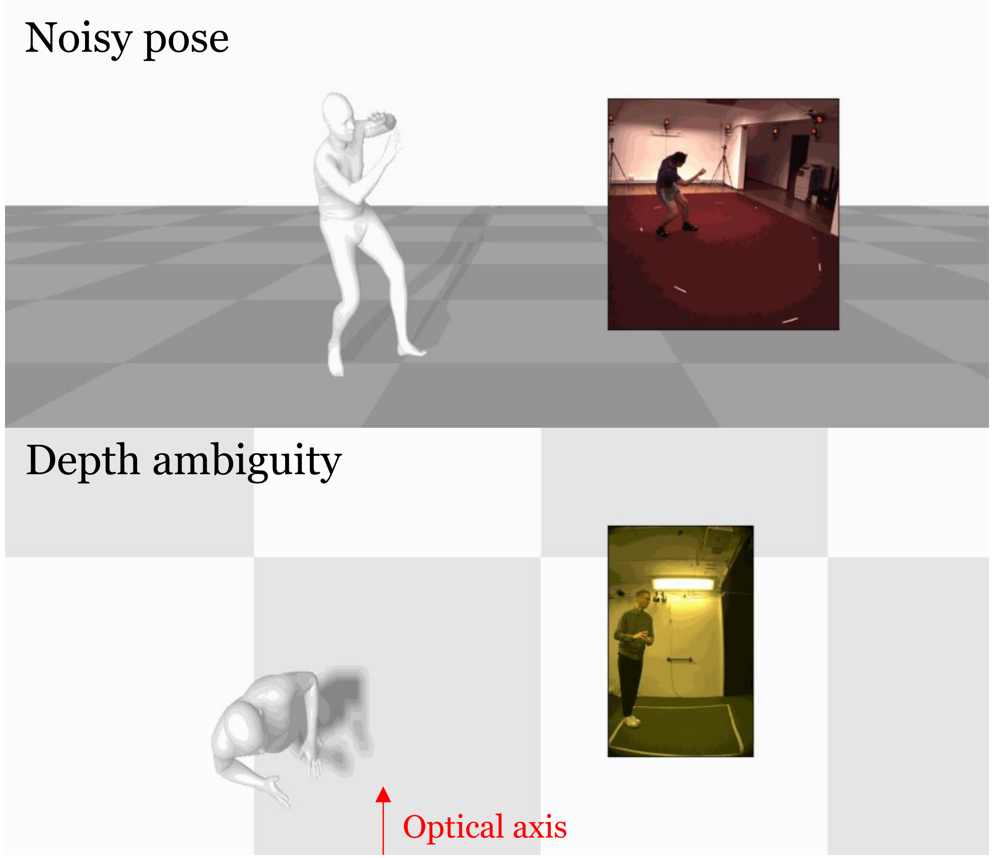
[2] Delft Center for Systems and Control, Delft University of Technology, The Netherlands

LINKÖPING UNIVERSITY

TU Delft

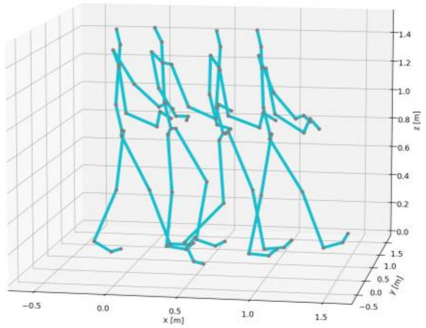
Motivation

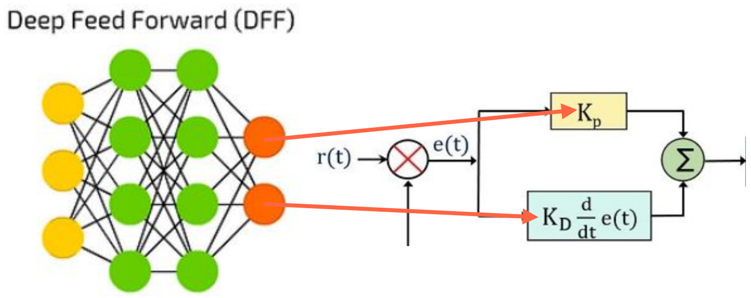
- Current monocular pose tracking in unconstrained environments is only partially solved.
- The challenges:
  - Variability in human appearance
  - Complexity of human motion and interactions of the body with other objects
  - Ambiguity in depth dimension
- Most common problems:
  - Jittering
  - Out-of-plane rotations
  - Ground plane penetration/floating
  - Foot-skating

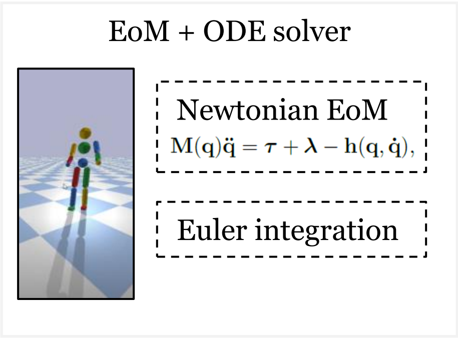


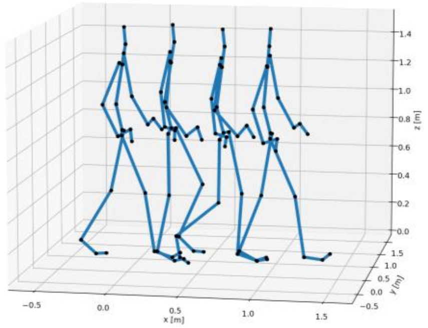
Physics-based Motion Capture

Noisy kinematics → Force estimation PD Controller → Physics constraints "Simulation" → Plausible Motion



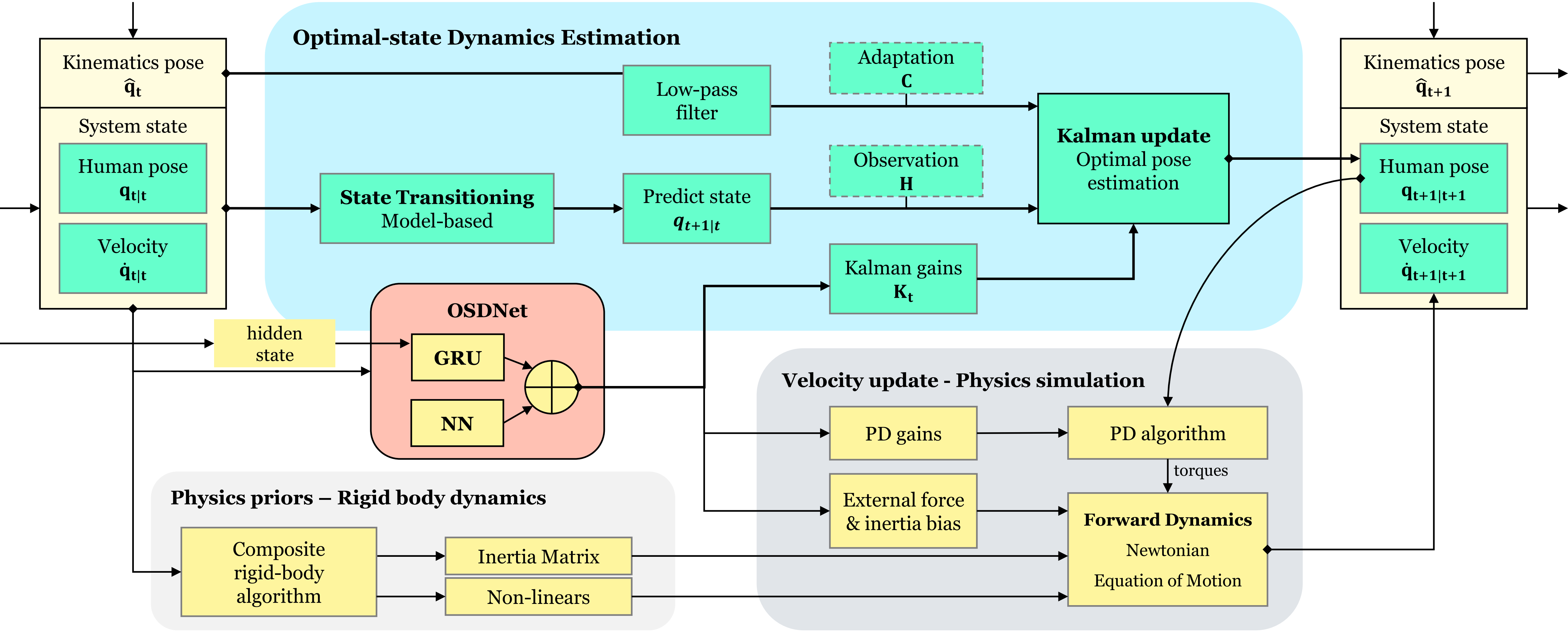






- Typical physics-based approach utilize the equation of motion.
- Reinforcement learning with physics engines.
- Trajectory optimization with physics-inspired losses.
- End-to-end estimation with differentiable physics.

OSDCap – Optimal-state Dynamics Estimation



Results - Comparison with SOTAs

- Human3.6M

Methods	Phys.	Onl.	MPJPE ↓ [mm]	MPJPE-G ↓ [mm]	MPJPE-PA ↓ [mm]	PCK ↑ [%]	CPS ↑ [mm]	GRP ↓ [mm]	Accel ↓ [mm/s^2]
Vnect [27]	✖	✔	89.6	-	62.7	85.1	-	185.1	-
HMMR [17]	✖	✔	79.4	-	55.0	88.4	-	231.1	-
HMR [16]	✖	✔	78.9	-	54.3	88.2	-	204.2	-
TRACE [40]	✖	✔	78.1	152.7	62.5	88.3	169.1	125.9	19.2
VIBE [19]	✖	✔	68.6	207.7	43.6	-	-	-	23.4
Gärtner et al. [9]	✔	✖	84.0	143.0	56.0	-	-	-	-
DiffPhy [8]	✔	✖	81.7	139.1	55.6	-	-	-	-
PhysPT [52]	✔	✖	52.7	-	36.7	-	-	-	-
*DnD [21]	✔	✖	**52.5**	-	**35.5**	-	-	-	-
PhysCap [37]	✔	✔	97.4	-	65.1	82.3	-	182.6	-
NeurPhys [38]	✔	✔	76.5	-	58.2	89.5	-	-	-
Xie et al. [46]	✔	✔	68.1	-	-	-	-	**85.1**	-
IPMAN-R [41]	✔	✔	60.7	-	41.1	-	-	-	-
SimPoE [50]	✔	✔	56.7	-	41.6	-	-	-	**6.7**
OSDCap	✔	✔	54.8±0.1	**132.8**±1.6	39.8±0.1	**95.5**±0.1	**197.7**±0.1	119.1±1.8	8.4±0.2

- Fit3D
- SportPose

Dataset	Methods	MPJPE ↓ [mm]	MPJPE-G ↓ [mm]	MPJPE-PA ↓ [mm]	PCK ↑ [%]	CPS ↑ [mm]	GRP ↓ [mm]	Accel ↓ [mm/s^2]
Fit3D [7]	TRACE [40]	85.4	131.2	65.2	85.5	166.6	178.1	20.2
	OSDCap	**58.7**	**73.8**	**42.6**	**96.7**	**209.4**	**47.2**	**8.2**
SportsPose [14]	TRACE [40]	97.3	361.9	71.1	60.1	168.1	333.0	15.8
	OSDCap	**71.7**	**113.6**	**52.4**	**68.8**	**190.0**	**90.2**	**10.9**

- Physics metrics

Method	GP ↓ [mm]	GD ↓ [mm]	Friction ↓ [mm]	Velocity ↓ [mm/s]	Foot-skating ↓ [%]
TRACE	**2.6**	12.5	31.5	22.4	37.0
OSDCap	5.3	**8.2**	**14.6**	**12.8**	**15.2**

Results – Comparison to classical Kalman

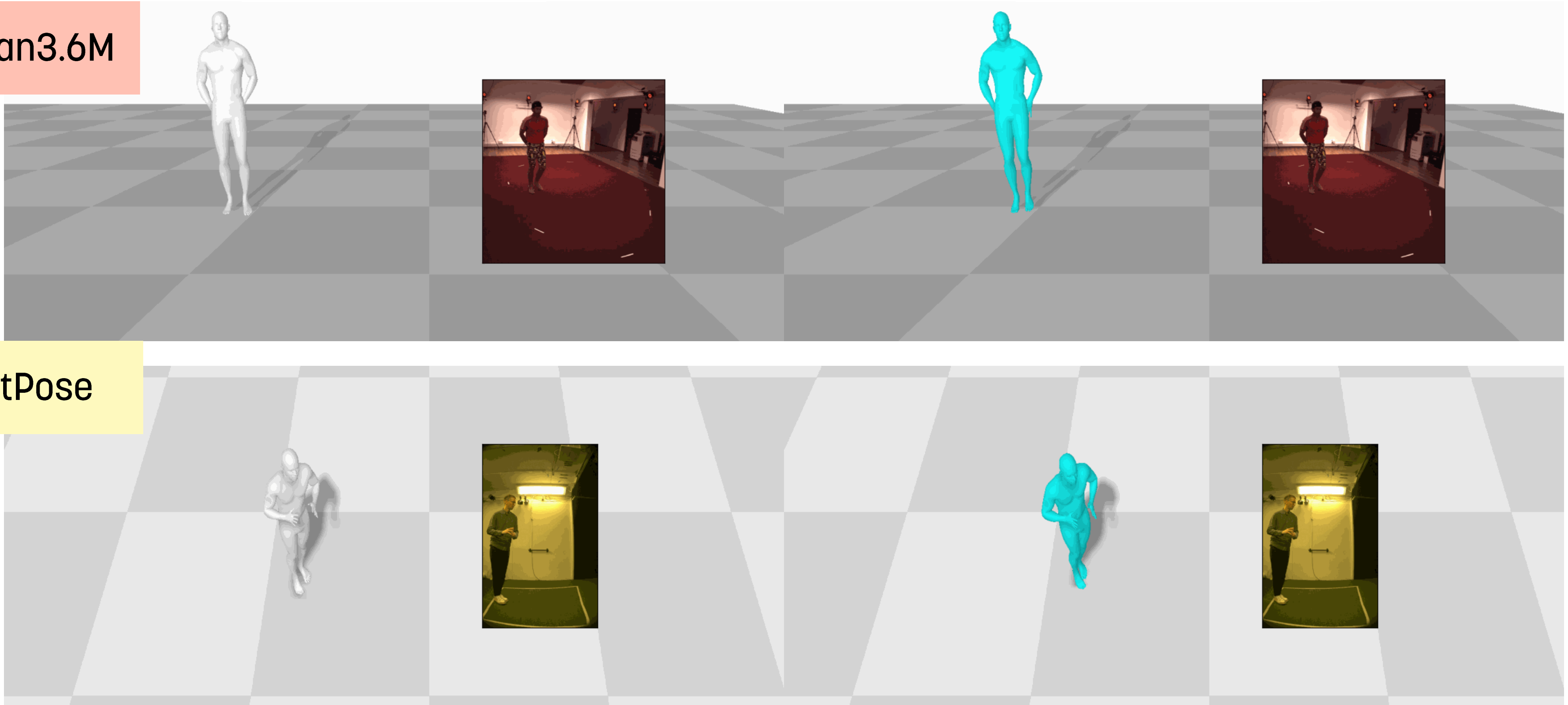
Method	MPJPE ↓ [mm]	MPJPE-G ↓ [mm]	MPJPE-PA ↓ [mm]	PCK ↑ [%]	GRP ↓ [mm]	Accel ↓ [mm/s^2]
TRACE	78.4	153.9	62.7	88.1	128.2	19.7
cKF_kin_only	78.3	153.0	63.0	87.9	127.4	7.8
cKF_100/1	60.9	120.7	43.4	94.3	102.0	7.7
cKF_10/1	61.5	122.6	43.7	94.4	103.0	9.1
cKF_1/1	59.9	117.6	43.2	94.8	100.1	**6.5**
cKF_1/10	63.6	124.0	44.3	93.8	102.7	11.5
cKF_1/100	65.3	132.3	44.0	93.5	110.2	9.7
OSDCap	**54.0**	**111.0**	**40.0**	**95.9**	**94.8**	8.7

- Classical Kalman filters require knowledge about the noise level of the data (hard to measure).
- Learnable neural Kalman relieves us from trial-and-error process, and helps achieving best results.

Results – Qualitatively

Input kinematics | OSDCap motions

- Human3.6M
- SportPose



Thank you for listening!

https://github.com/cuongle1206/OSDCap