Pose estimation with limited data

SOA analysis and research roadmap

Debaditya Roy

GOAL

State-of-the-art in pose estimation with limited data.

- IMU for our case.
- Extension to our research.
- Possible applications

Agenda

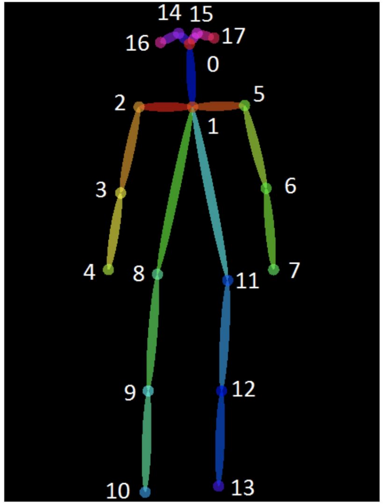
- What is pose estimation?
- Evolution over some research works.
- Background.
- Discussion on selected papers.
- From application to research.
- Possible areas to explore.
- Collaborative learning factor.
- Other applications.

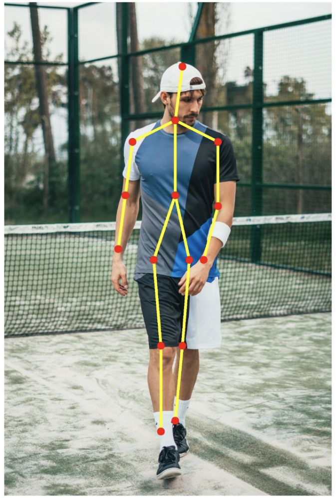
Agenda

- What is pose estimation?
- Evolution over some research works.
- Background.
- Discussion on selected papers.
- From application to research.
- Possible areas to explore.
- Collaborative learning factor.
- Other applications.

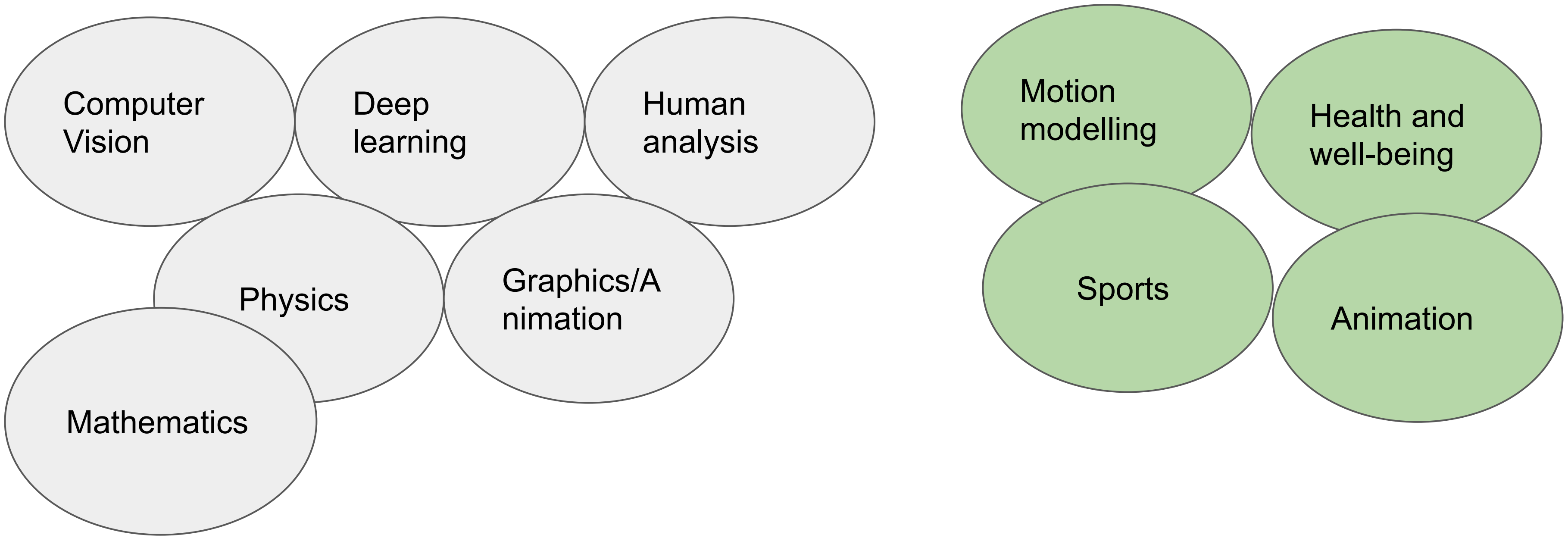
What is pose estimation?

- Computer vision problem.
- Detect human figures in image/video/other data.
- Detect key joints of the figure.
- Not a person identification process.





Research and application domains



Agenda

- What is pose estimation?
- Evolution over some research works.
- Background.
- Discussion on selected papers.
- From application to research.
- Possible areas to explore.
- Collaborative learning factor.
- Other applications.

Evolution over some research works

- Pose estimation from video and imus. [1]
- Pose estimation from only imu. (Sparse inertial poser[2])
- Pose estimation with only imu and deep learning (Deep interial poser[3]).

Agenda

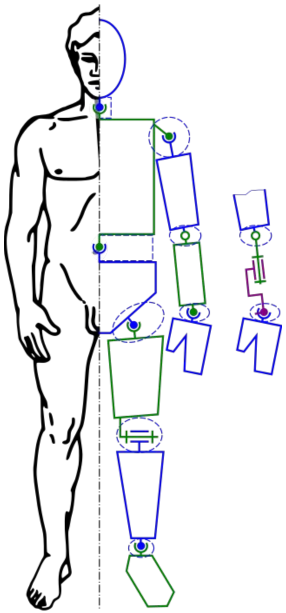
- What is pose estimation?
- Evolution over some research works.
- Background.
- Discussion on selected papers.
- From application to research.
- Possible areas to explore.
- Collaborative learning factor.
- Other applications.

Background - Model based pose estimation[4]

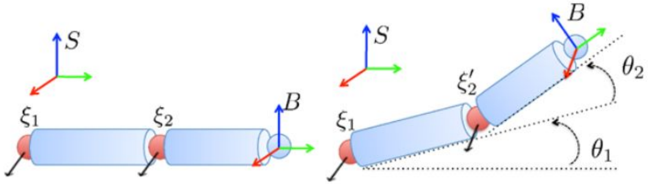
- Kinematic parametrization.
- Model creation.
- Optimization.

Background - Kinematic parametrization[4]

- Good parametrization requirements:
  - Pose configuration represented with minimum number of parameters.
  - Natural human motion.
  - Easy computation of derivatives of segment positions and orientation w.r.t to parameters.
  - Simple rules of concatenating motions.

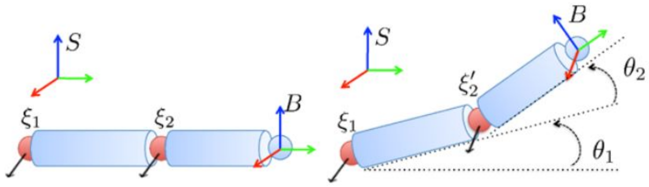


- Kinematic chain.



Background - Kinematic chain[4]

- Encodes motion of a body segment as motion of previous segment in the chain.
- Angular motion about the body joint.
- Motion of lower arm is parameterized by motion of upper arm and rotation about the elbow.



Background - Model based pose estimation[4]

- Kinematic parametrization.
  - Rotation matrices.
  - Axis-Angle
    - Exponential maps of rigid body motion.
  - Kinematic chains.
  - Human pose parametrization.
- Model creation.
- Optimization.

Background - Rotation Matrices

- Encodes orientation of body frame w.r.t spatial frame.

  Ps = RsbPb

  Rsb = Rotation matrix

- Rotation and translation

  Ps = RsbPb + ts

- **Rigid body motion: g = (R,t).**

Background - Axis Angle representation[4]

- Describe rotations as angle θ and axis $\boldsymbol{\omega}$.
- Exponential formulae: R = exp(θώ)
    - ώ = skew symmetric matrix of $\boldsymbol{\omega}$

$$\begin{bmatrix} 0 & -\omega_3 & \omega_2 \\ \omega_3 & 0 & -\omega_1 \\ -\omega_2 & \omega_1 & 0 \end{bmatrix}$$

- We can express exp(θώ) as

$$\exp(\theta\widehat{\omega}) = I + \widehat{\omega}\sin(\theta) + \widehat{\omega}^2\big(1 - \cos(\theta)\big)$$

Background - Extending to rigid bodies[4]

- Rotation + translation = Twist
- Twist denoted by $\theta\xi = \theta(v_1, v_2, v_3, \omega_1, \omega_2, \omega_3)$
- Rigid body motion expressed as

$$\mathbf{G}(\theta, \omega) = \begin{bmatrix} \mathbf{R}_{3\times3} & \mathbf{t}_{3\times1} \\ \mathbf{0}_{1\times3} & 1 \end{bmatrix} = \exp(\theta\widehat{\xi}),$$

Background - Extending to rigid bodies[4]

- Rotation + translation = Twist
- Twist denoted by $\theta\xi = \theta(v_1, v_2, v_3, \omega_1, \omega_2, \omega_3)$
- Rigid body motion expressed as

$$\mathbf{G}(\theta, \omega) = \begin{bmatrix} \mathbf{R}_{3\times3} & \mathbf{t}_{3\times1} \\ \mathbf{0}_{1\times3} & 1 \end{bmatrix} = \exp(\theta\widehat{\xi}),$$

Twist action

Background - Extending to rigid bodies[4]

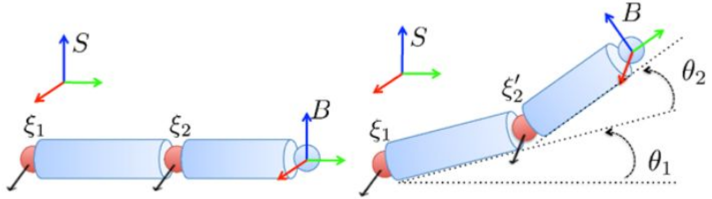
- Rotation + translation = Twist
- Twist denoted by $\theta\xi = \theta(v_1, v_2, v_3, \omega_1, \omega_2, \omega_3)$
- Rigid body motion expressed as

$$\mathbf{G}(\theta, \omega) = \begin{bmatrix} \mathbf{R}_{3\times3} & \mathbf{t}_{3\times1} \\ \mathbf{0}_{1\times3} & 1 \end{bmatrix} = \exp(\theta\widehat{\xi}),$$

Twist action

$$\theta\widehat{\xi} = \theta \begin{bmatrix} 0 & -\omega_3 & \omega_2 & v_1 \\ \omega_3 & 0 & -\omega_1 & v_2 \\ -\omega_2 & \omega_1 & 0 & v_3 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Background - Representing kinematic Chains[4]

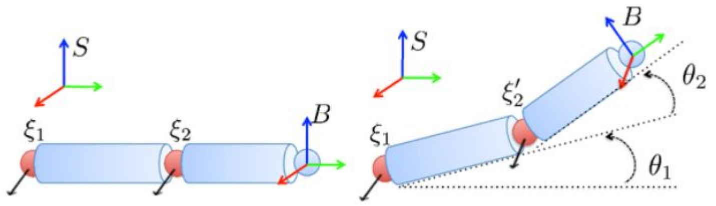


- Obtain control point in the hand in spatial coordinates ps from body cooridnate pb.

$$\bar{\mathbf{p}}_s = \mathbf{G}_{sb}\bar{\mathbf{p}}_b = \mathbf{G}_1\mathbf{G}_2\mathbf{G}_{sb}(\mathbf{0})\bar{\mathbf{p}}_b$$

G1, G2 = rigid body matrices of upper and lower arm, Gsb = rest pose transformation.

Background - Generalized forward kinematics map[4]



$$\bar{\mathbf{p}}_s = \mathbf{G}_{sb}(\theta_1, \theta_2) = e^{\widehat{\xi}_1\theta_1} e^{\widehat{\xi}_2\theta_2} \mathbf{G}_{sb}(\mathbf{0})\bar{\mathbf{p}}_b.$$

Generalized forward kinematic map:

$$\mathbf{G}_{sb}(\Theta) = e^{\widehat{\xi}_1\theta_1} e^{\widehat{\xi}_2\theta_2} \cdots e^{\widehat{\xi}_n\theta_n} \mathbf{G}_{sb}(\mathbf{0})$$

Background - Pose parametrizations[4]

Joint	DoF	Unknown parameter	Example
Root	6	$\xi = \theta[v\ \omega]^T$	All body
Ball	3	$\theta\omega$	Hips
Saddle	2	θ_1, θ_2	Wrist
Revolute	1	θ	Knee

$$\mathbf{x}_t := (\xi, \Theta), \quad \Theta := (\theta_1 \theta_2 \ldots \theta_n).$$

- Root joint and revolute joint
- Root joint represents the twist parameter and revolute joints rest of the angles.

Background - Model Creation

- Geometric primitives.
- Detailed body scans.
- Detailed shape from images.

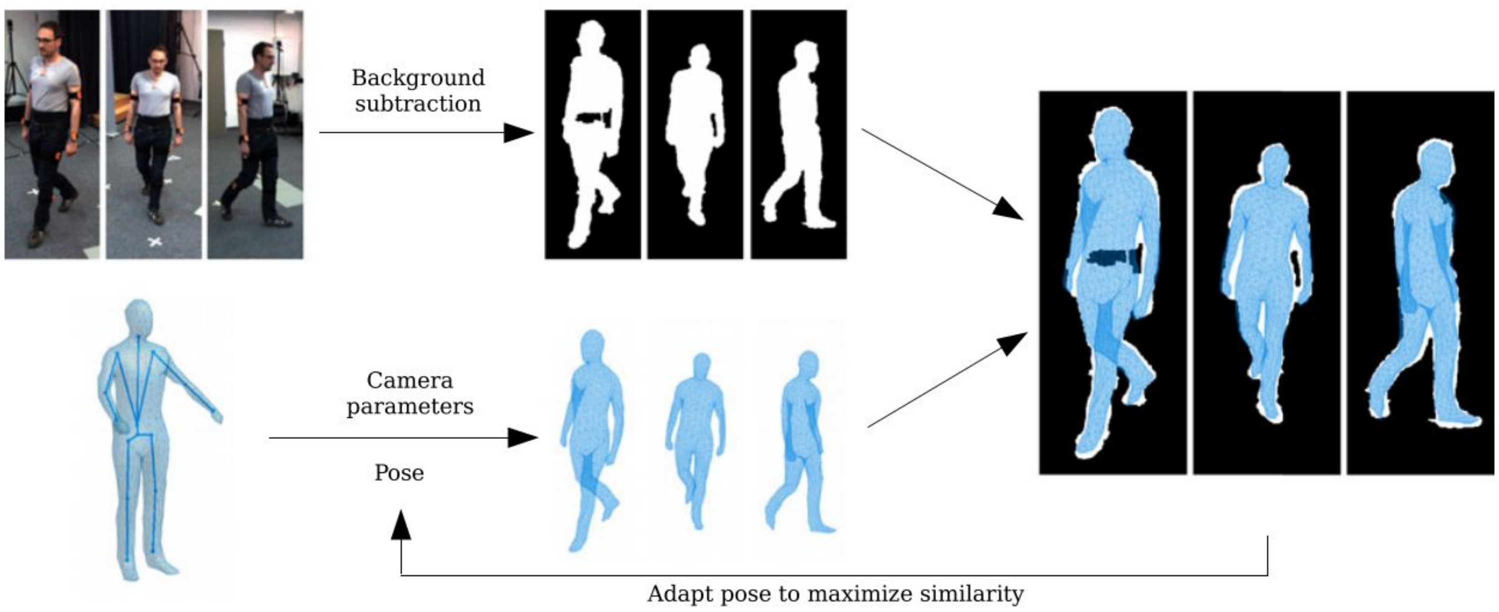
We can use the parameterizations learned before to create body models.

Background - Optimization

- **Model the likelihood of the observations for a given configuration of pose parameters**

- **Pose that best explains observation: Minimizes error function that fits the model data to the given data.**
  - **Model-image association and then error minimization.**
  - **Model-imu data association and then error minimization.**

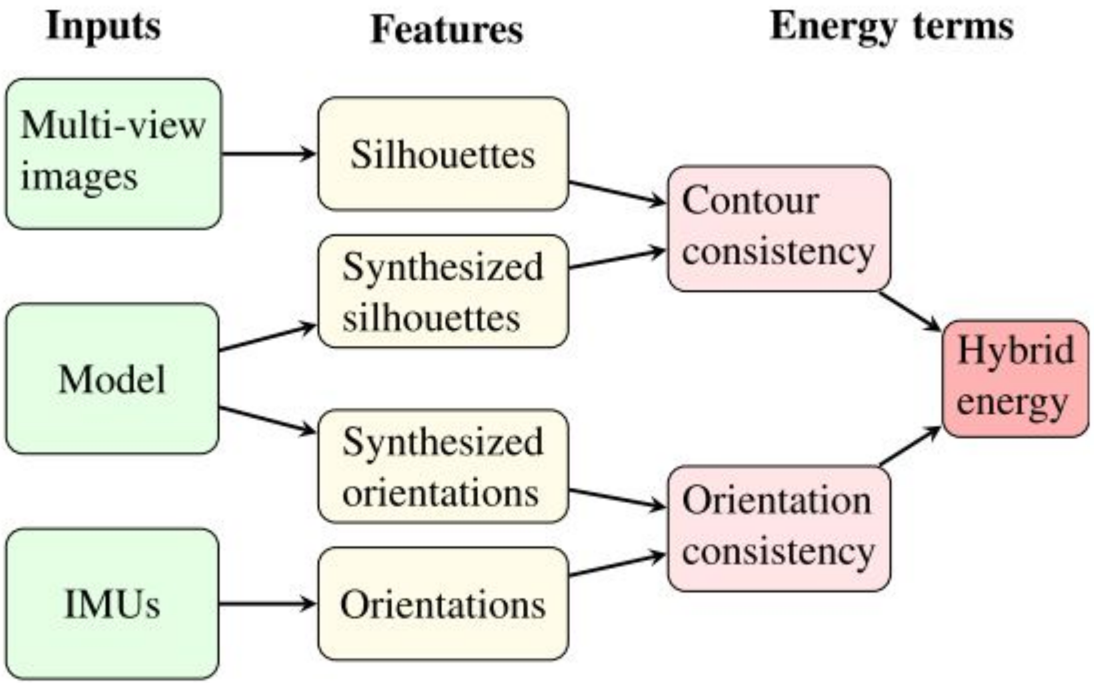
Background - Example[4]



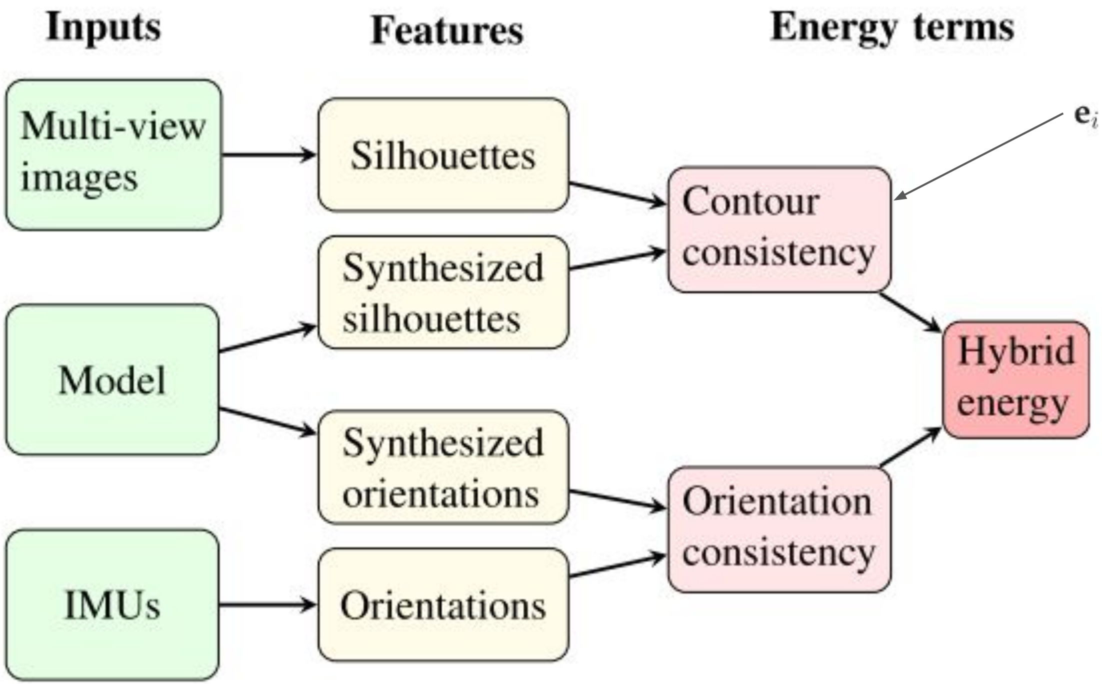
Agenda

- What is pose estimation?
- Evolution over some research works.
- Background.
- Discussion on selected papers.
- From application to research.
- Possible areas to explore.
- Collaborative learning factor.
- Other applications.

Pose estimation from videos and imus[1]



Pose estimation from videos and imus[1]



$$\mathbf{e}_i = \mathcal{F}(\mathbf{x}; \mathbf{p}_i) \times \mathbf{n}_i - \mathbf{m}_i$$

$$\arg\min_{\mathbf{x}} E_{\text{video}}(\mathbf{x}) = \sum_{i=1}^{N} \|\mathbf{e_i}\|^2$$

$$= \sum_{i=1}^{N} \|\mathcal{F}(\mathbf{x}; \mathbf{p}_i) \times \mathbf{n}_i - \mathbf{m}_i\|^2$$

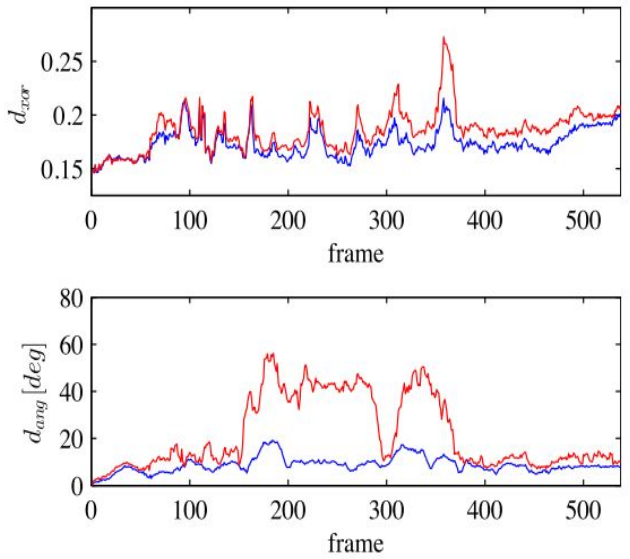
$$\arg\min_{\mathbf{x}} E_{\text{hybrid}}(\mathbf{x}) = E_{\text{video}}(\mathbf{x}) + \lambda E_{\text{sens}}(\mathbf{x})$$

Pose estimation from videos and imus - Measures[1]

- TNT-15 dataset.
  - 5 activity sequence: Walking, running ….
  - Multi-view camera images.
  - 10 sensors: 5 sensors for tracking, 5 sensors for validation.
- Error measure:
  - Angular error w.r.t bone orientations. -> Good measure for orientation consistency.
  - Silhouette overlap between projected and original image.
    - $$d_{xor}(S^{video}, S^{model}) = \frac{1}{K}\sum_{j=1}^{K} \frac{S_j^{video} \oplus S_j^{model}}{S_j^{video} \vee S_j^{model}}$$

d=0 means identical silhouette and d=1 means no overlap.

Pose estimation from videos and imus - Results[1]



- Walking sequence.
- Red = Video tracker.
- Blue = Hybrid tracker.

Mean Angular Error $\mu_{ang}[deg]$ of the Validation Sensors Attached to Thighs, Chest and Upper Arms for the Video-Based and Hybrid Tracker for All Sequences of the Database

	lThigh	rThigh	chest	lUArm	rUArm
video tracker	19.12	12.36	11.97	61.03	46.28
hybrid tracker	8.64	6.75	6.88	27.30	28.96

Sparse Interial Poser[2]

- Recovers full human 3D pose using only 6 IMUs.
- Sensors measure orientation and acceleration.
- Uses SMPL body model.

Sparse Interial Poser[2]

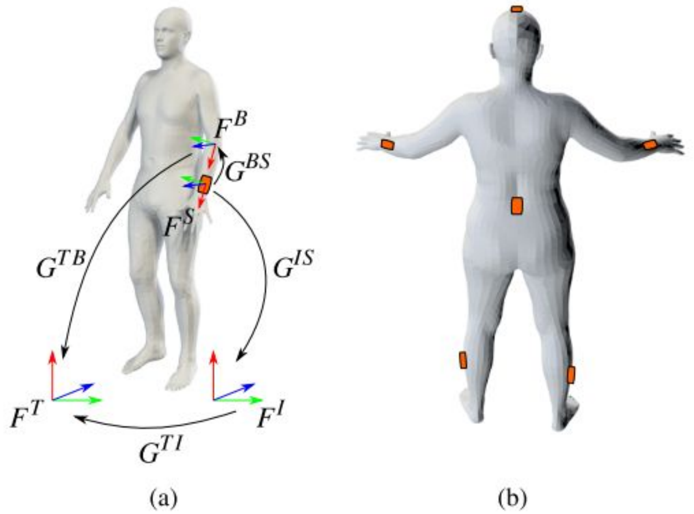


Figure 3: *(a) Coordinate frames: Global tracking coordinate frame F^G, Inertial coordinate frame F^I, Bone coordinate frame F^B and Sensor coordinate frame F^S. (b) Sensor placement at head, lower legs, wrists and back.*

- Sensors attached to the marked locations.
- Translation along each coordinate system gives us the translation for the rigid body motion.
- Formulate the kinematic chain.

Sparse Interial Poser[2]

- Creates SMPL body model over multiple frames.
- Measures orientation and acceleration of the actor over the frames.
- Learn the pose parameters for the SMPL body model.

$$\mathbf{x}_{1:T}^{*} = \underset{\mathbf{x}_{1:T}}{\arg\min}\, E_{\text{motion}}(\mathbf{x}_{1:T}, \mathbf{R}_{1:T}, \mathbf{a}_{1:T})$$

- Aim to recover sequence of poses s.t actual sensor acceleration matches corresponding vertex acceleration.
- Incorporates anthropometric term for realism.

Sparse Interial Poser - Evaluation[2]

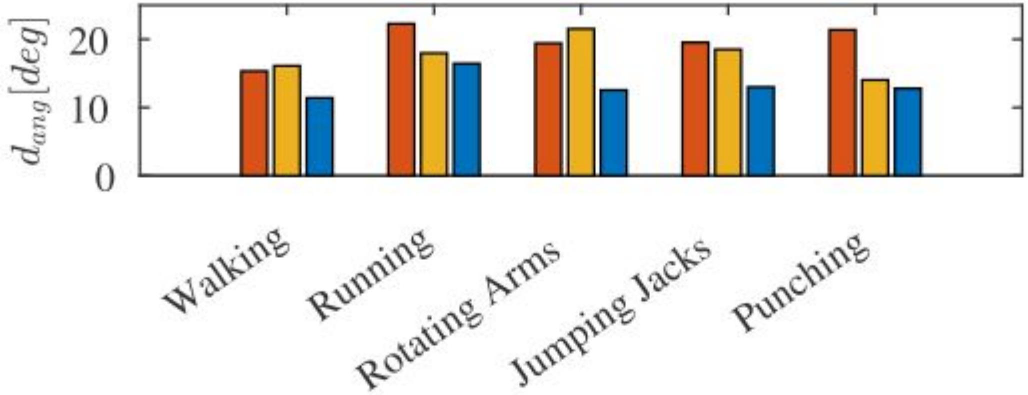
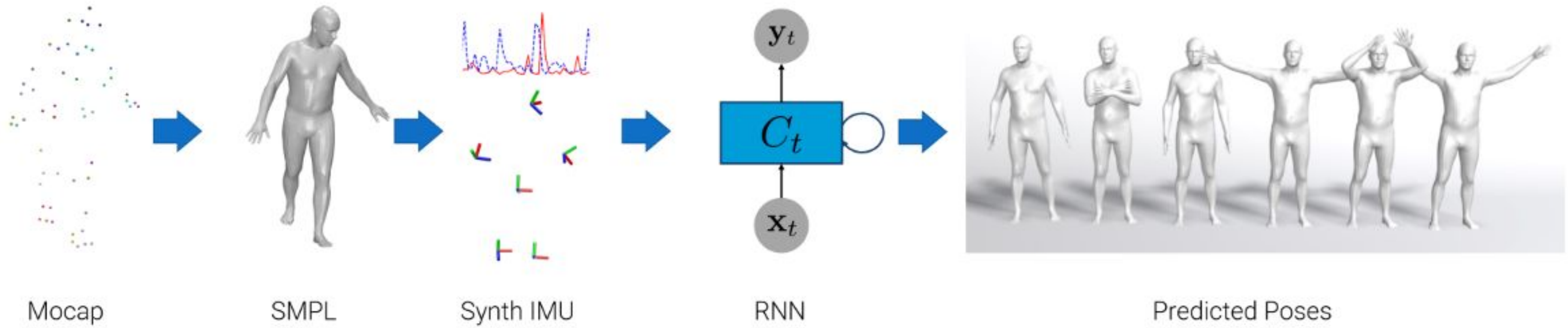


Figure 9: *Mean orientation error on the TNT15 data set: comparison of SOP(red), SIP-M(yellow) against our proposed SIP (blue).*

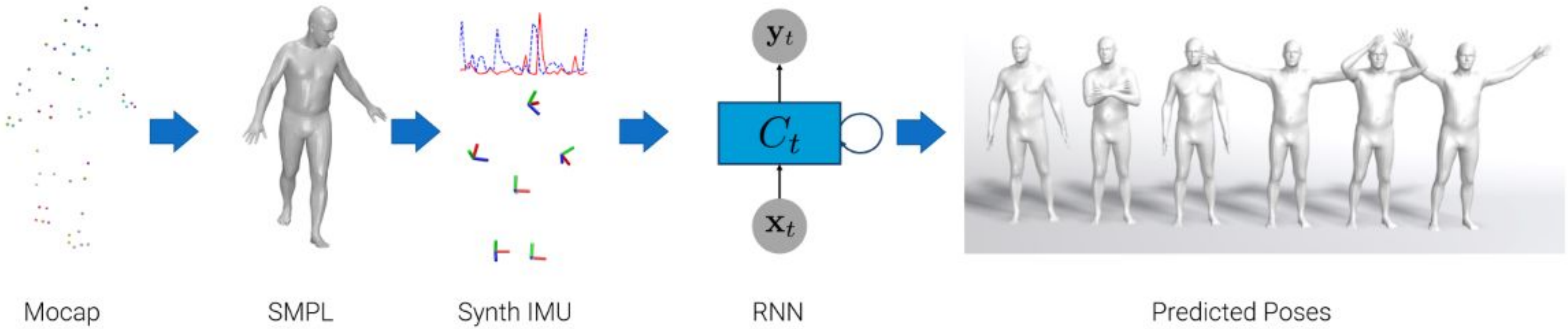
13.32 over all sequences - Mean angular error

Deep Inertial Poser[3]



- Computationally more efficient than the previous work.
- Real-time predictor.
- Synthesizes IMU data from motion capture dataset.

Deep Inertial Poser[3]



- Model long-range temporal dependencies using RNN to map orientation and acceleration to SMPL parameters.
- Reconstructs acceleration during training.
- Uses bidirectional RNN for using both past and future information.

Deep Inertial Poser - Evaluation[3]

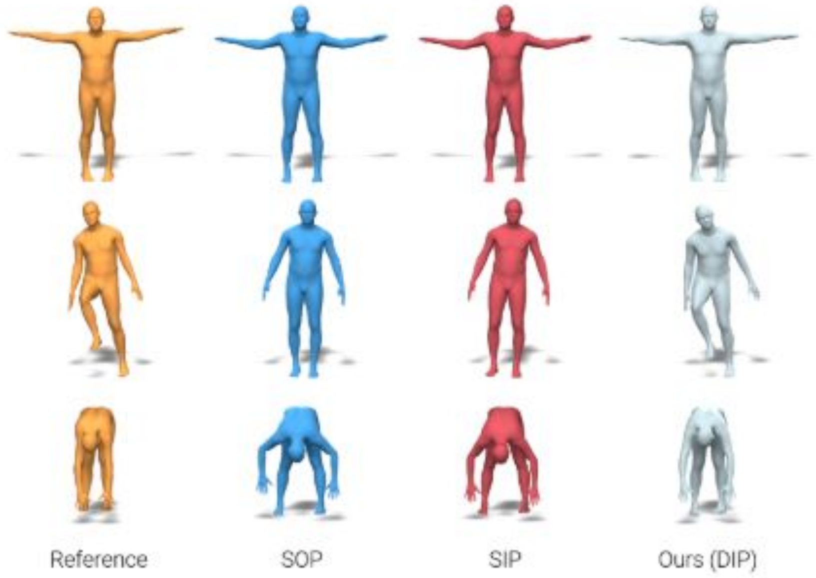


Fig. 8. Sample frames from TotalCapture data set (S1, ROM1).

Agenda

- What is pose estimation?
- Evolution over some research works.
- Background.
- Discussion on selected papers.
- From application to research.
- Possible areas to explore.
- Collaborative learning factor.
- Other applications.

From application to research

- Personalized pose models for athletes.
  - Providing pose-estimate.
  - Providing motion-estimate.
  - Technique improvement.
  - Technique analysis.

Our research setup - super constrained input space.

- Accelerometer recordings (Mobile, sensor).
- Pressure sensor.
- Heart-rate monitor.

Our research setup - Constrain output space.

- Knowledge about motion prior.
- Specific nature of motion (e.g. running)
  - May use the fact that the motions may arise from similar distribution.
- Parts affecting the motion might be limited.
  - E.g. Leg for running might be more interesting to analyze.

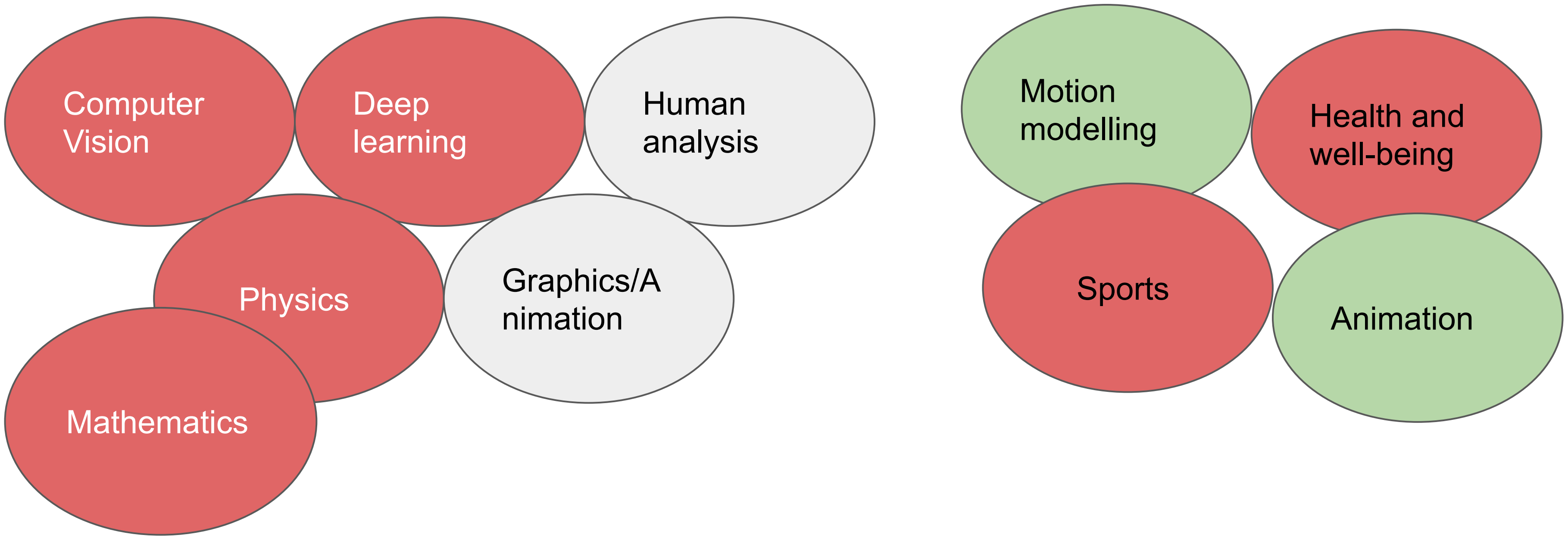
Agenda

- What is pose estimation?
- Evolution over some research works.
- Background.
- Discussion on selected papers.
- From application to research.
- Possible areas to explore.
- Collaborative learning factor.
- Other applications.

Possible areas to explore

- Spatio-temporal nature of the problem..
- Usage of deep neural networks in the problem.
    - Usage of autoencoder to generate motion patterns from same distribution.
- Pose estimation problems with deep neural networks.
- Effect of incorporation of prior knowledge.
- Exploring other latent signals arising from sensor data.
- Physics based models for human motion.

Target imapct circles.



Agenda

- What is pose estimation?
- Evolution over some research works.
- Background.
- Discussion on selected papers.
- From application to research.
- Possible areas to explore.
- Collaborative learning.
- Other applications.

Collaborative learning.

Explore possibility of having collaborative learning in the setup.

- Explore if interaction among multiple actors can improve individual models.

Agenda

- What is pose estimation?
- Evolution over some research works.
- Background.
- Discussion on selected papers.
- From application to research.
- Possible areas to explore.
- Collaborative learning factor.
- Other applications.

Other applications.

- Medical domain.
- Gait analysis.
- Injury prevention.
- Real-time pose estimates in sports like football.

Future works

- Explore other data for pose estimation.
- Explore real-time angle of the same.
- Publishing a survey paper on the same.
- Exploration of other possible application domains.
- Present on individual smaller topics in study-groups.

References

1. Von Marcard, Timo, Gerard Pons-Moll, and Bodo Rosenhahn. "Human pose estimation from video and imus." *IEEE transactions on pattern analysis and machine intelligence* 38.8 (2016): 1533-1547.
2. Von Marcard, Timo, et al. "Sparse inertial poser: Automatic 3d human pose estimation from sparse imus." *Computer Graphics Forum*. Vol. 36. No. 2. 2017.
3. Huang, Yinghao, et al. "Deep inertial poser: learning to reconstruct human pose from sparse inertial measurements in real time." *ACM Transactions on Graphics (TOG)* 37.6 (2018): 1-15.
4. Pons-Moll, Gerard, and Bodo Rosenhahn. "Model-based pose estimation." *Visual analysis of humans*. Springer, London, 2011. 139-170.