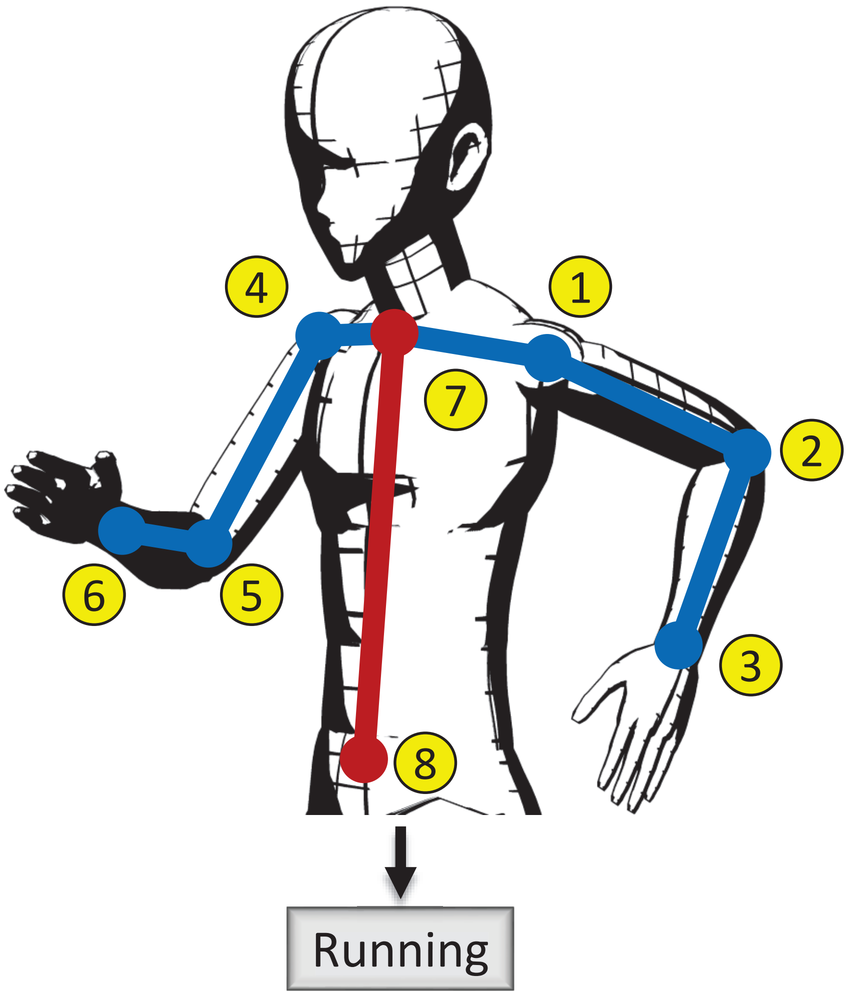
Real-time Arm Skeleton Tracking and Gesture Inference Tolerant to Missing Wearable Sensors

Yang Liu[1], Zhenjiang Li[1], Zhidan Liu[2], Kaishun Wu[2]

City University of Hong Kong[1], Shenzhen University[2]

香港城市大學
City University of Hong Kong

Understanding Human Arm Motions



- **How is the arm moving?**

  Skeleton tracking

- **What is the meaning of this arm motion?**

  Motion inference

Elderly Care

Elderly diseases

- Parkinson
- Alzheimer

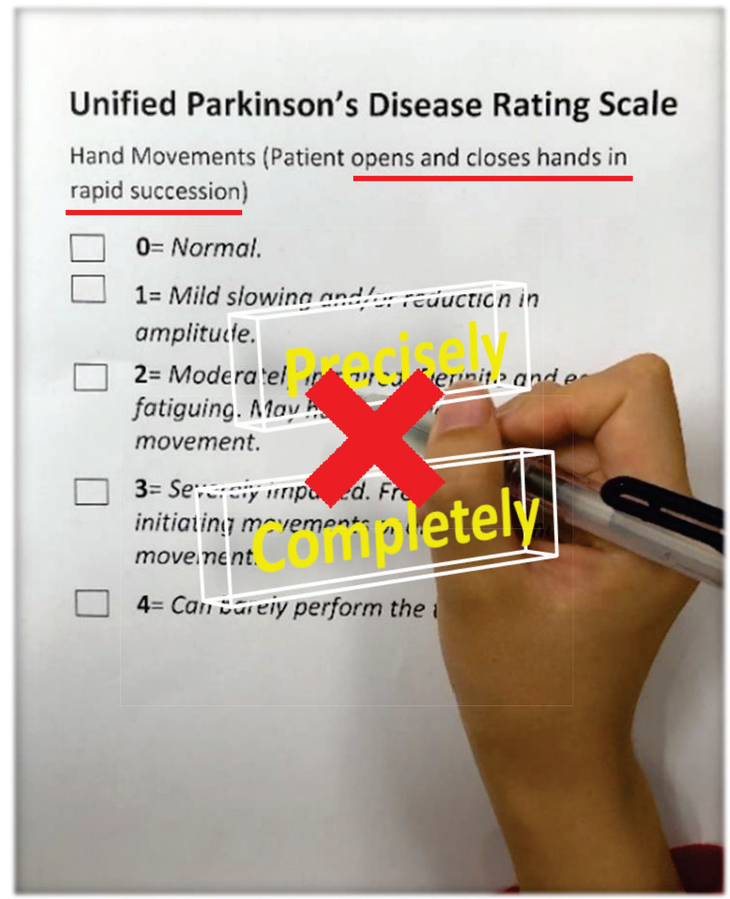
Problems with arm

- Slow motion
- Repeated motion
- Instability
- …

Last treatment

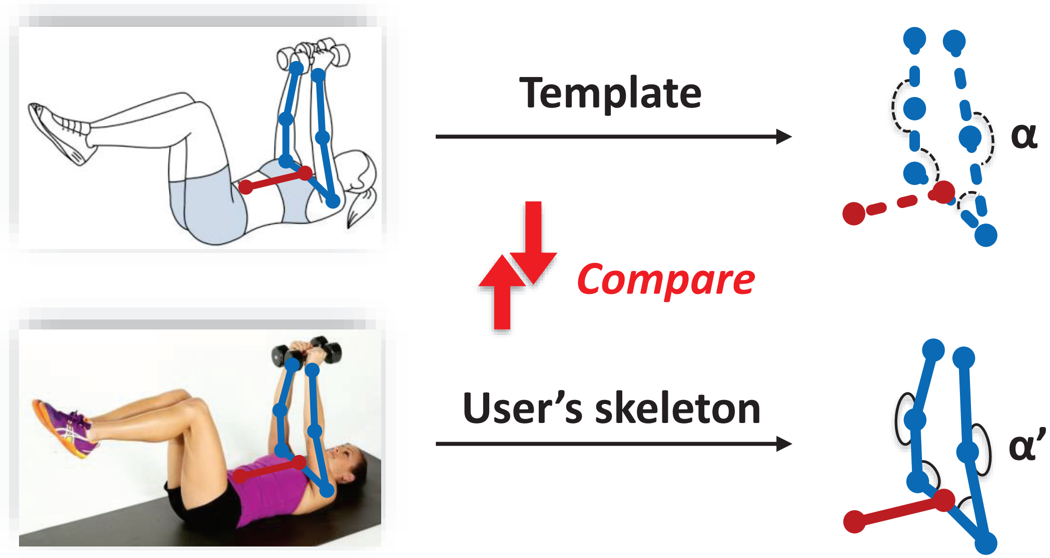
Several weeks

Next treatment



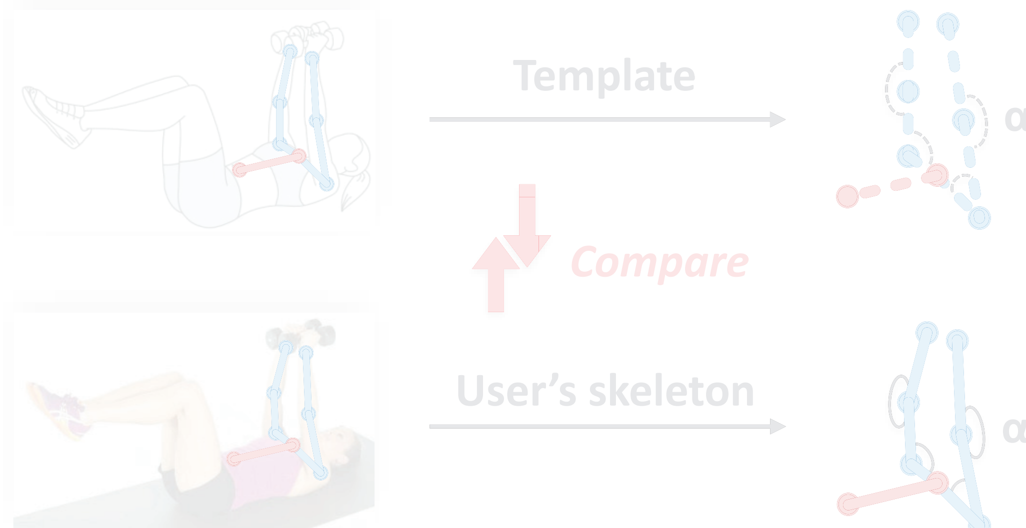
Other Applications

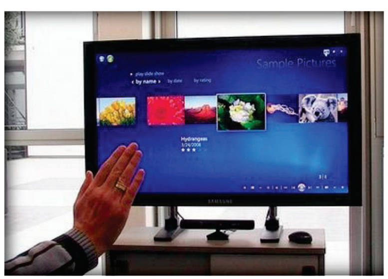
80USD/hour



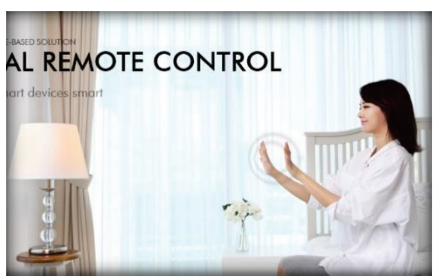
Other Applications

80USD/hour

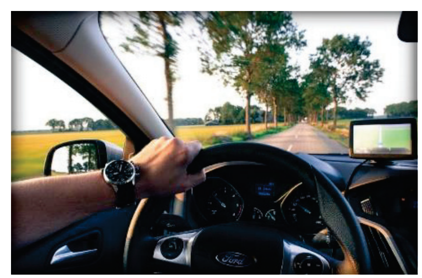




HCI



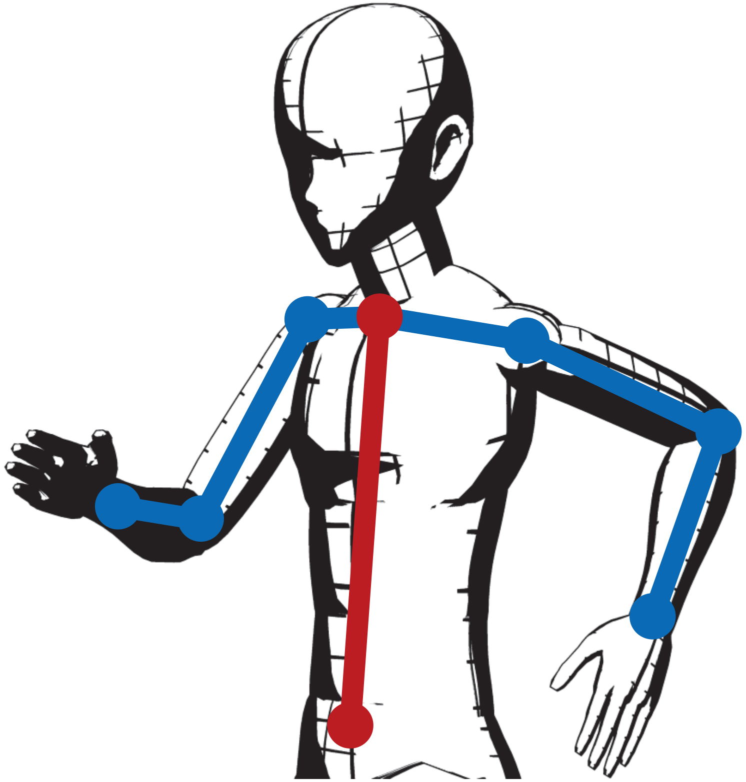
Smart home



Smart car

Gaming

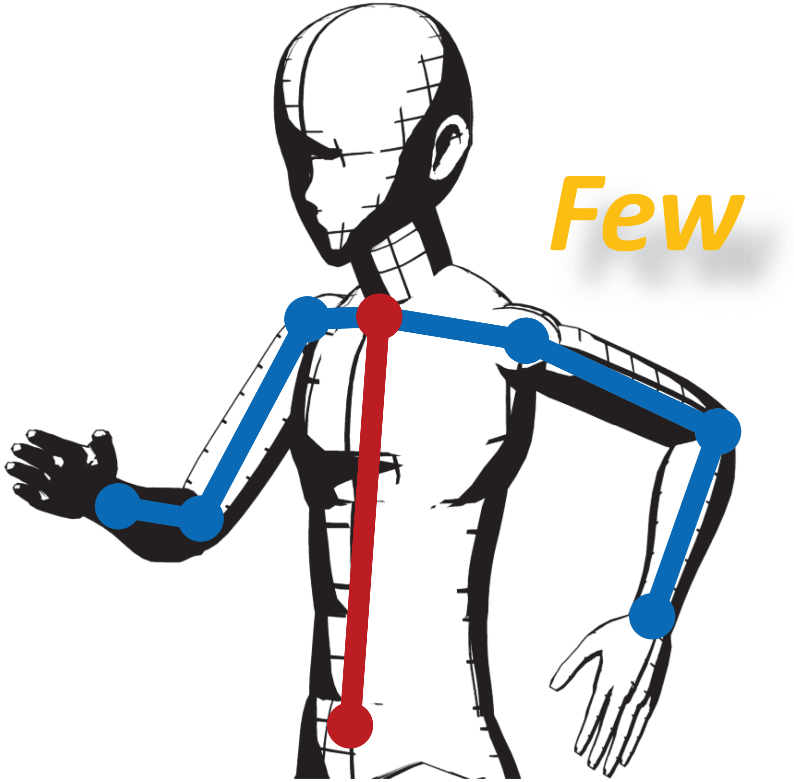
Existing Solutions



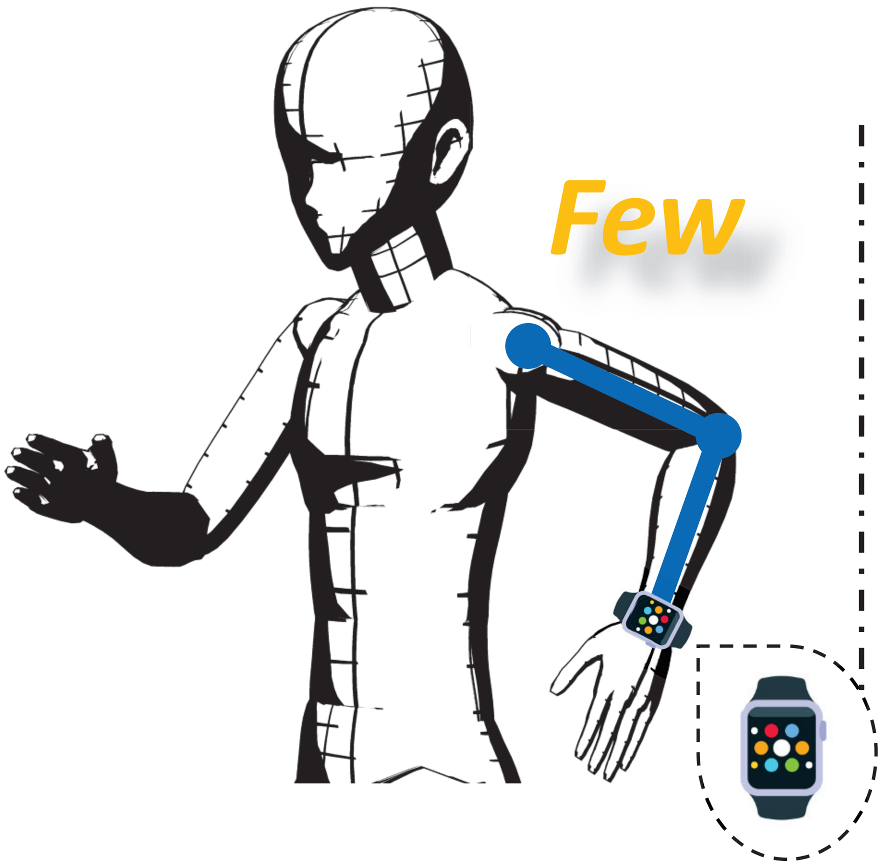
- Service coverage
- System cost
- Privacy

- Convenience
- User-friendly

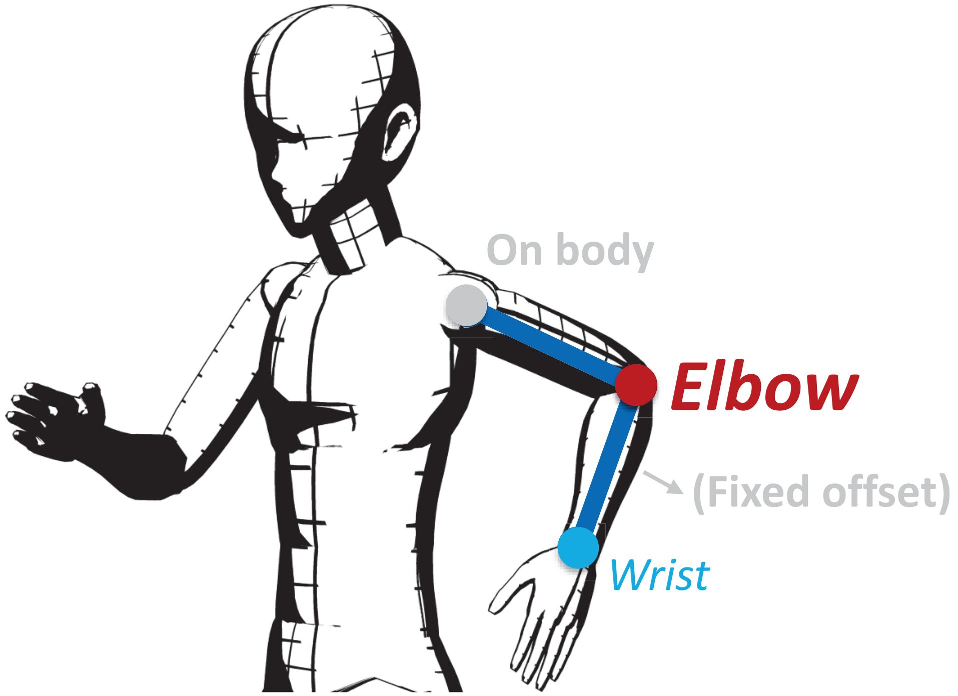
Existing Solutions



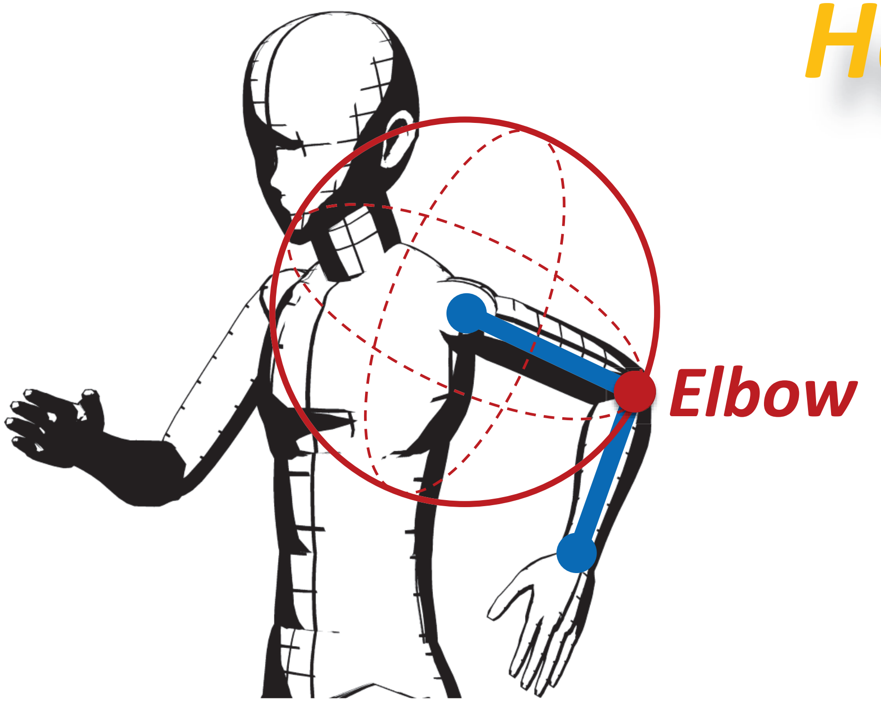
Existing Solutions



Key Problem

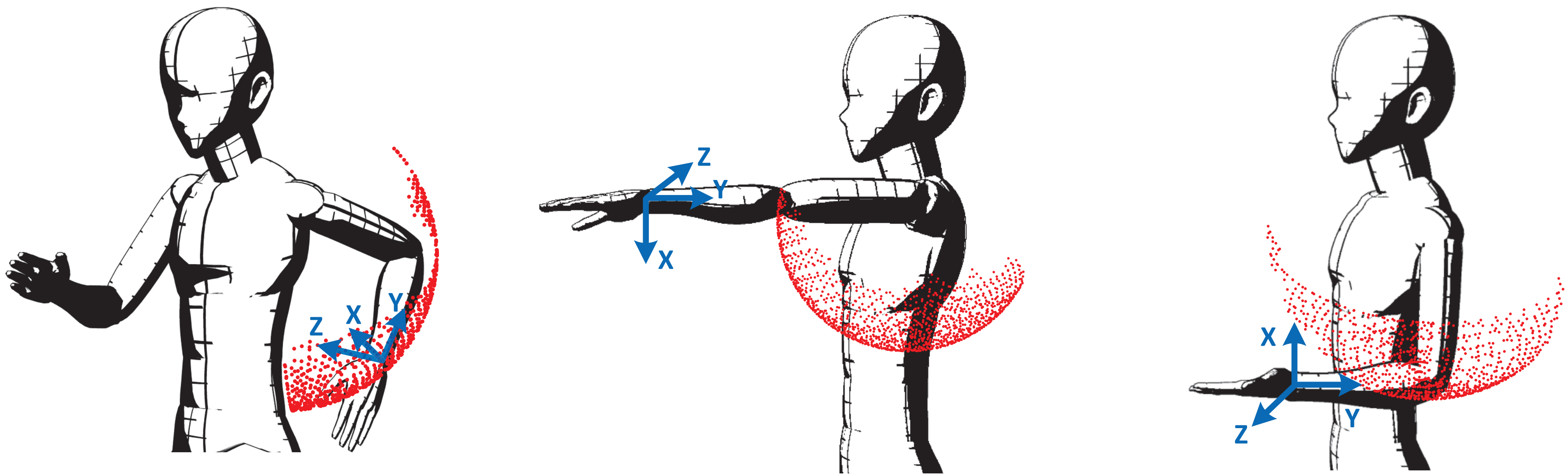


Key Problem



How?

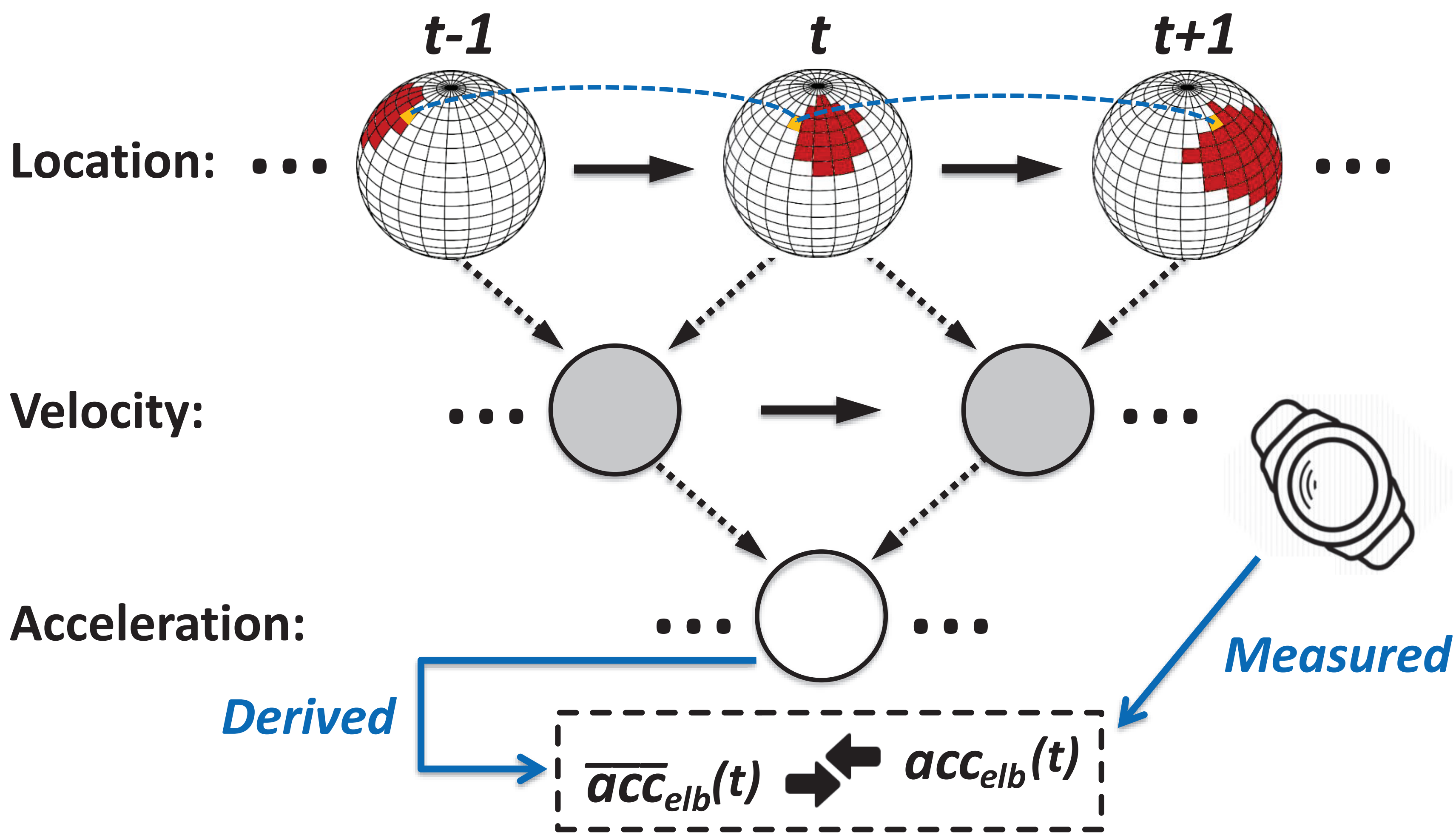
Tracking Principle



For a given wrist orientation, possible elbow locations are within a limited range [1].

[1] "I am a smartwatch and I can track my user's arm", in Proc. of ACM MobiSys, 2016.

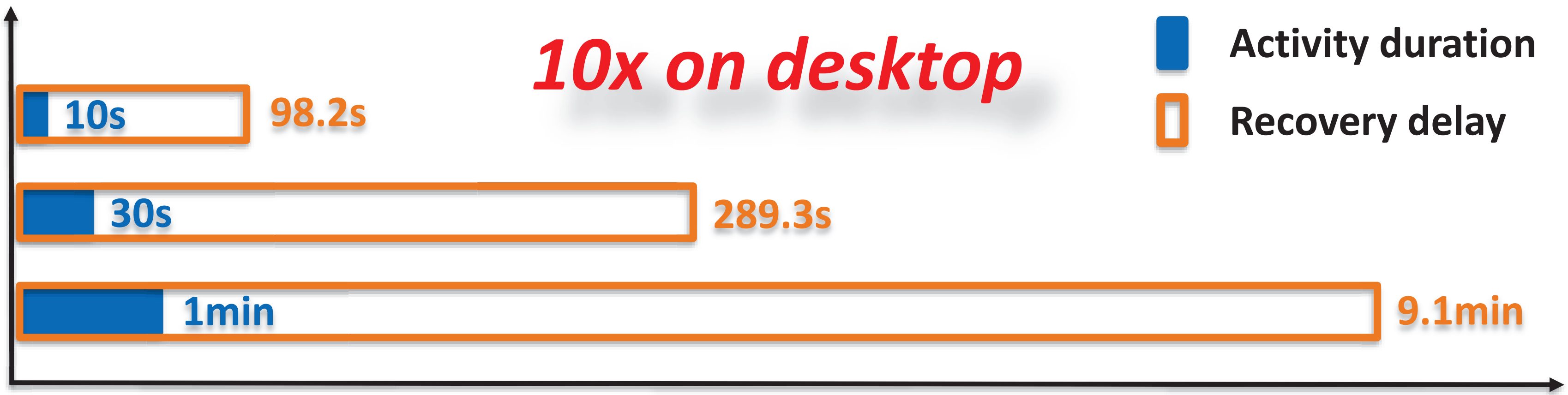
Tracking Principle



Ranges across time [1]

[1] "I am a smartwatch and I can track my user's arm", in Proc. of ACM MobiSys, 2016.

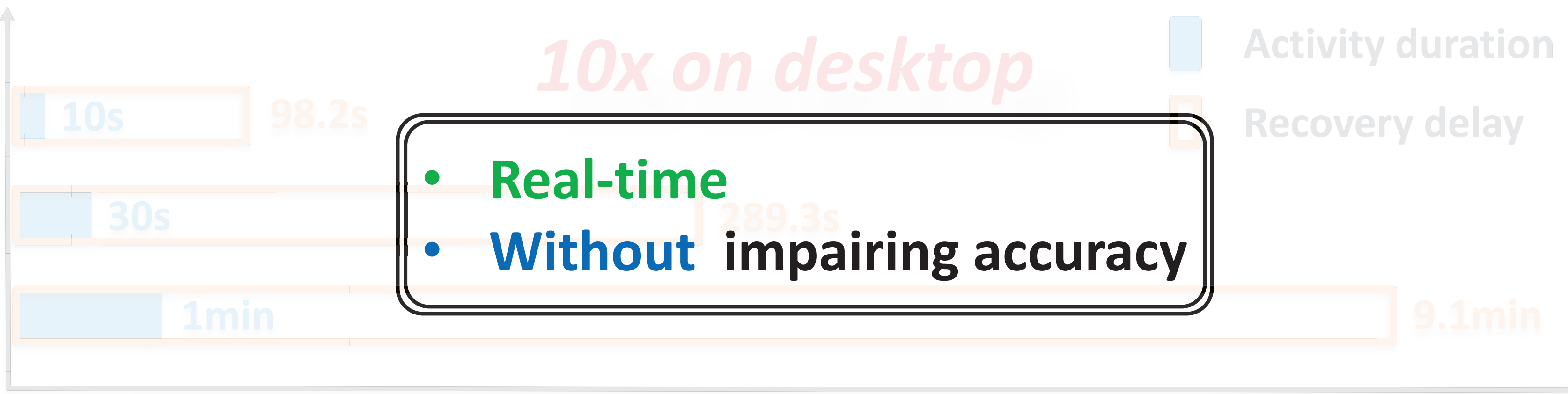
Latency



Time delay of existing work [1]

[1] “I am a smartwatch and I can track my user’s arm”, in Proc. of ACM MobiSys, 2016.

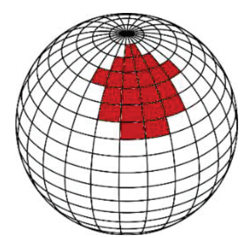
Latency



- **Real-time**
- **Without impairing accuracy**

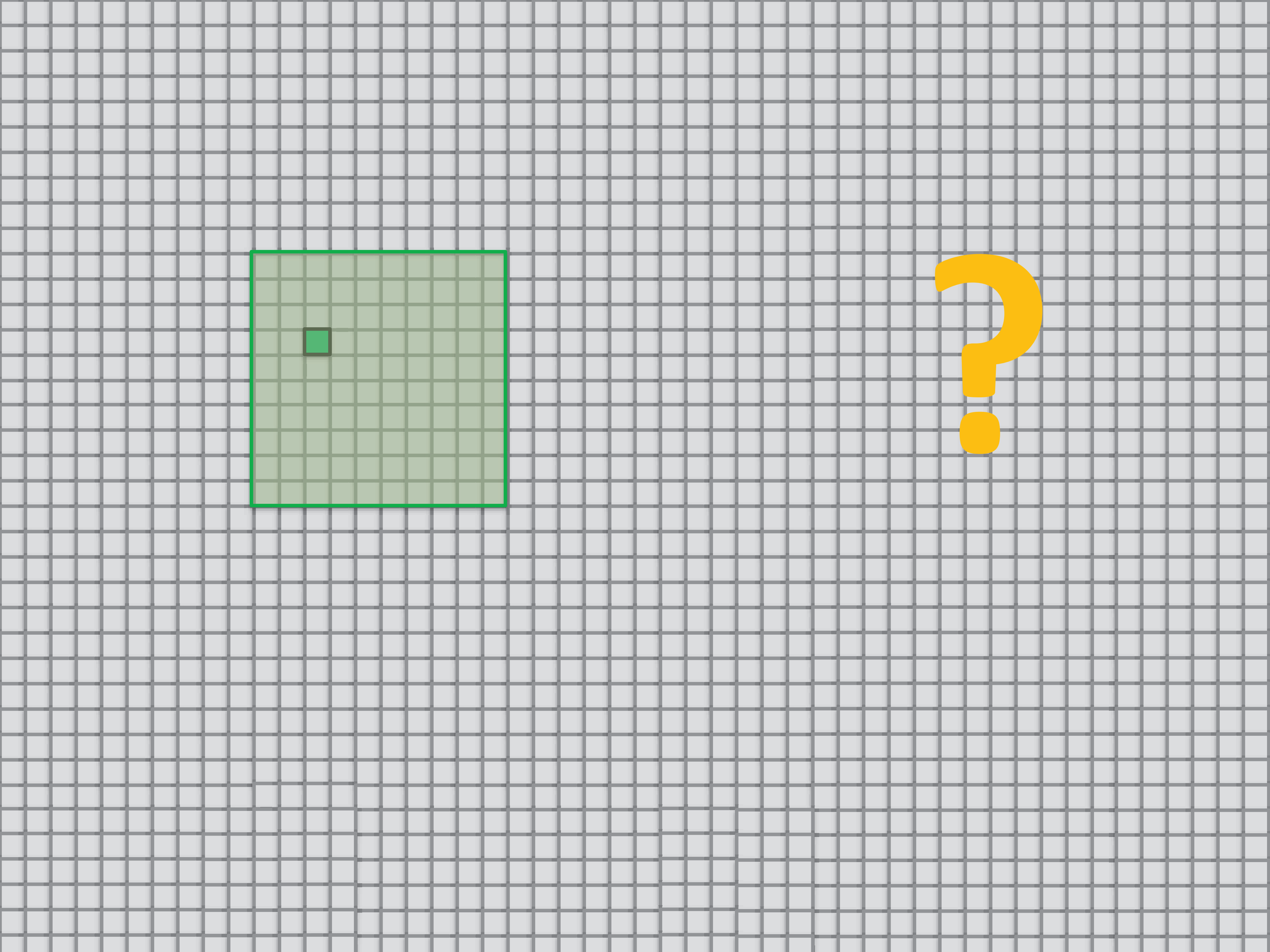
Time delay of existing work [1]

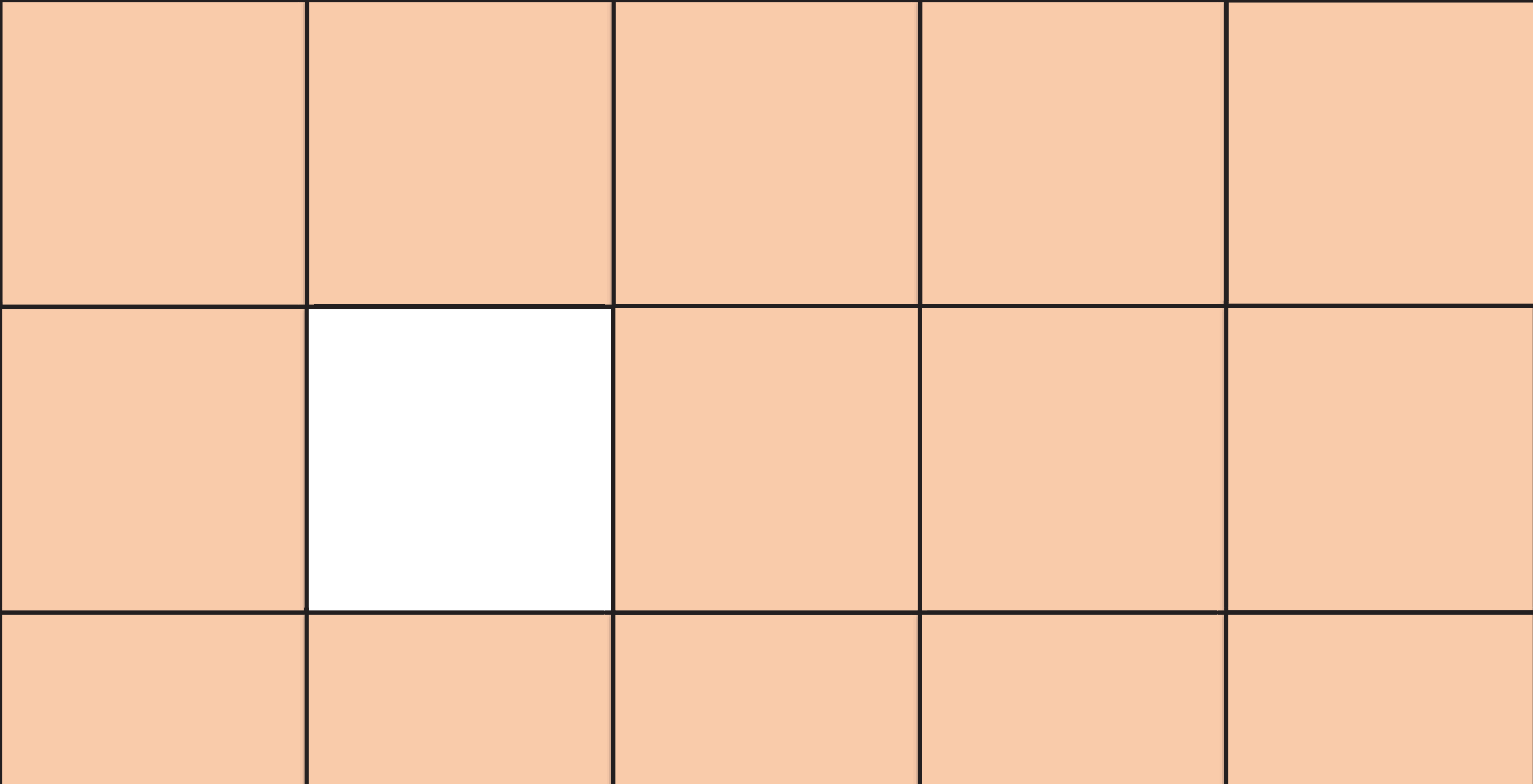
- **Our solution [ArmTroi]:**
  - HMM state reconstruction
  - Hierarchical search



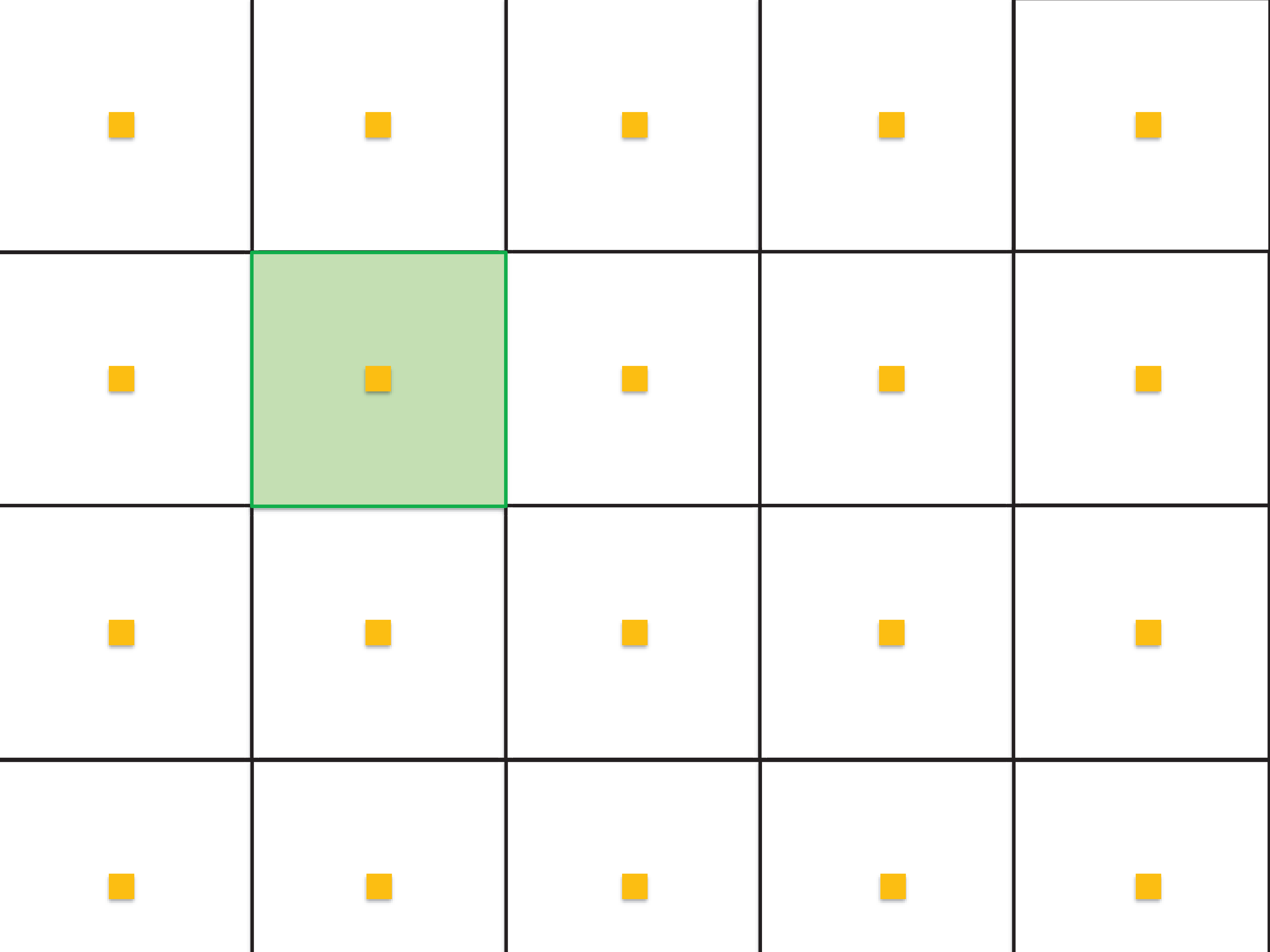
One search space

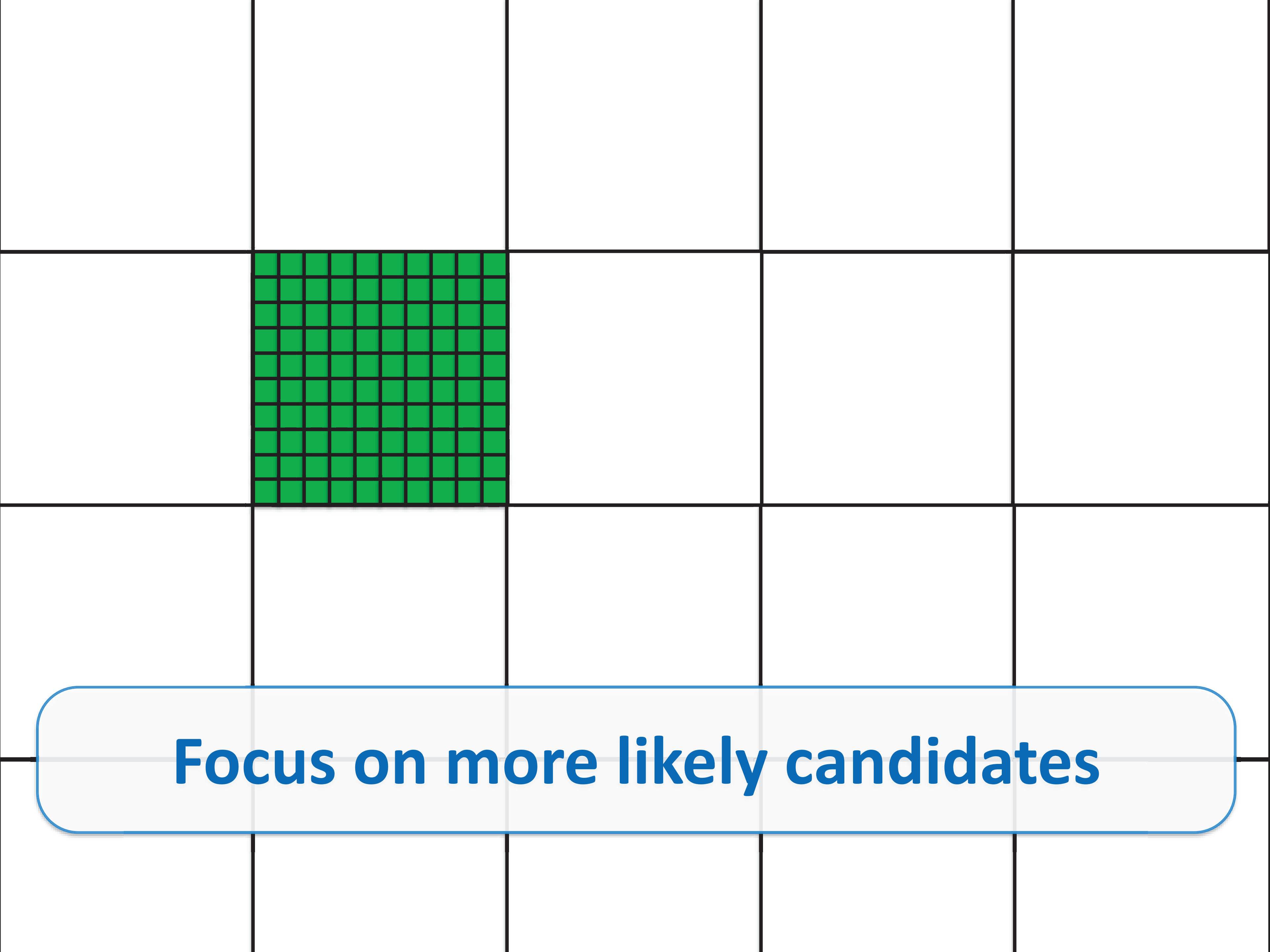
[1] “I am a smartwatch and I can track my user’s arm”, in Proc. of ACM MobiSys, 2016.



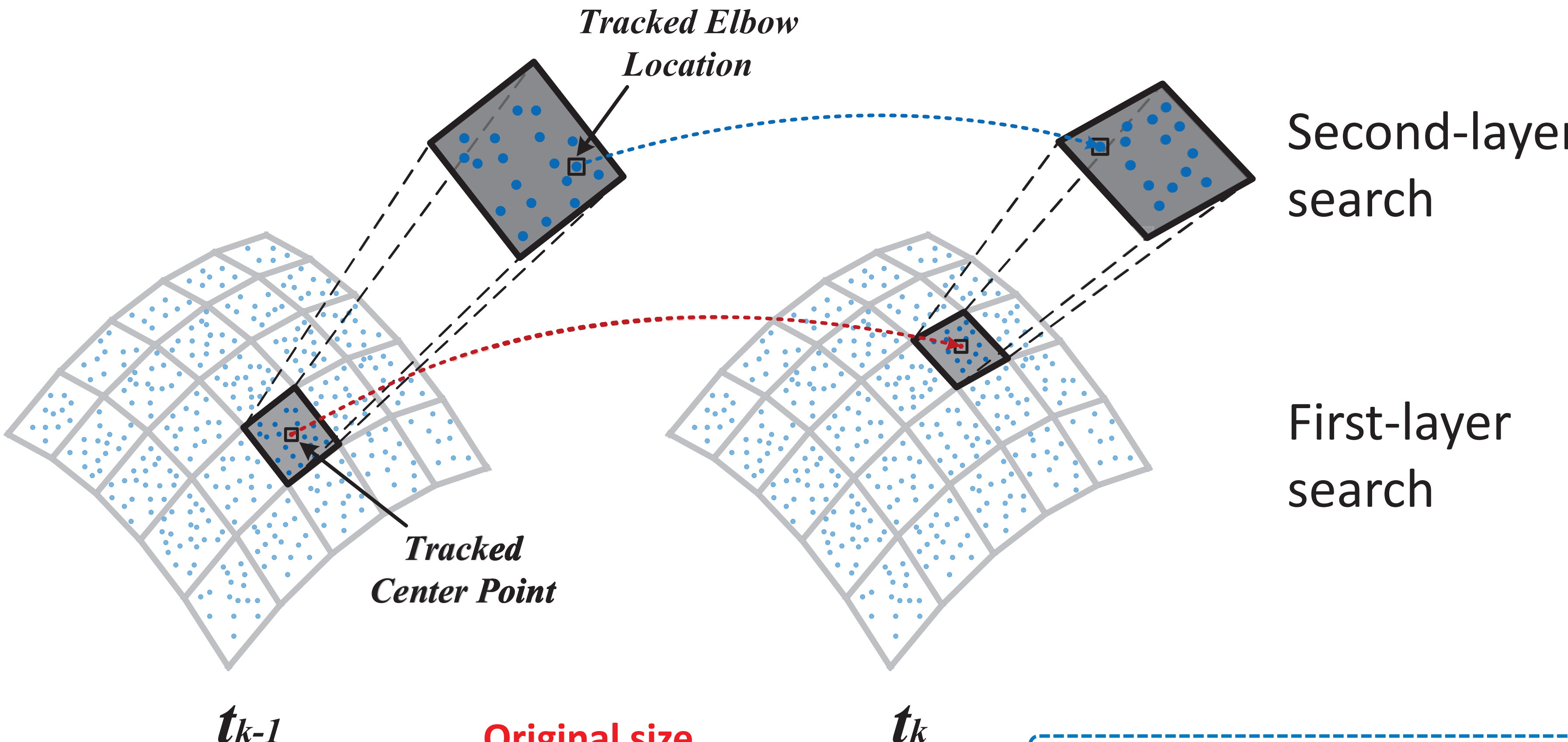


Our idea: exclude the unlikely locations using as little effort as possible




Focus on more likely candidates

Hierarchical Search



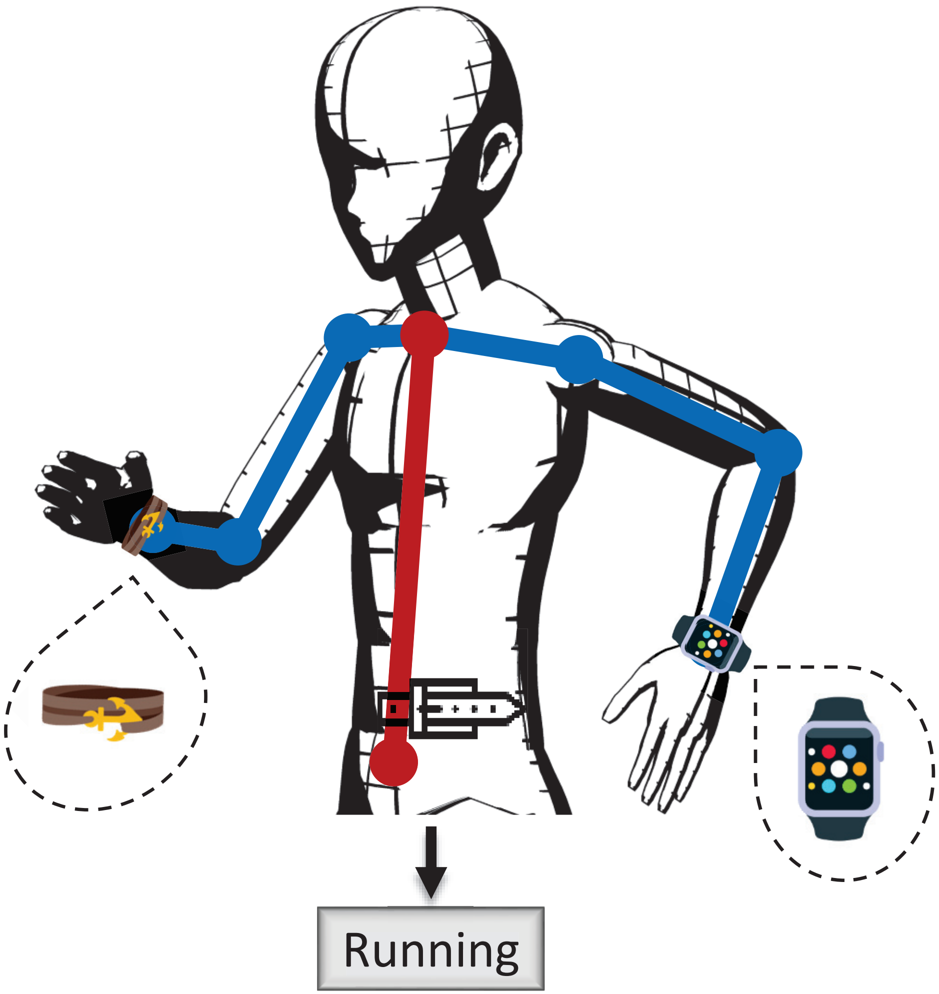
Original size

- **Time complexity:** $O\left(\left(\frac{\boxed{N}}{\boxed{n_1}}\right)^2 T + (n_1)^2 T\right)$

Size of region

- **Real-time**
- **Without impairing accuracy**

Understanding Human Arm Motions



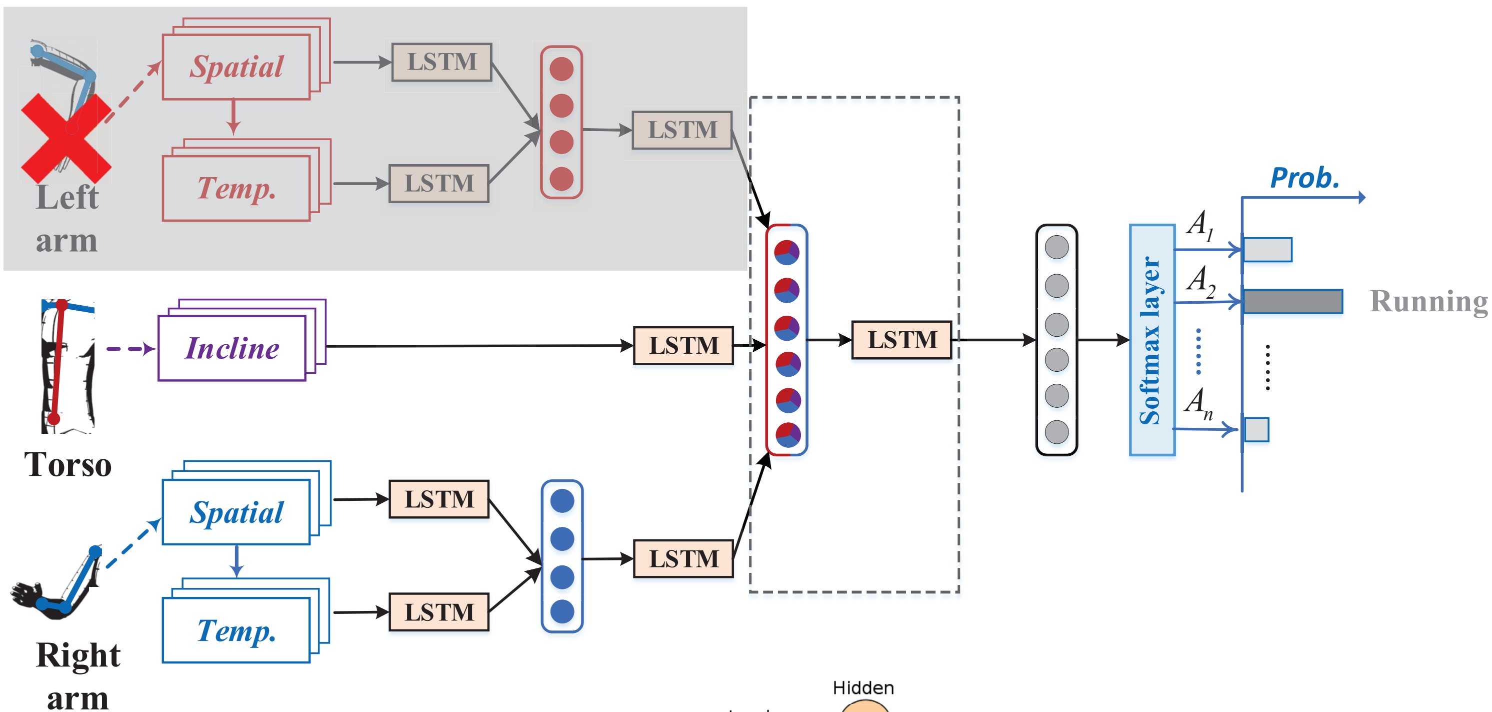
- **How is the arm moving?**

  Skeleton tracking

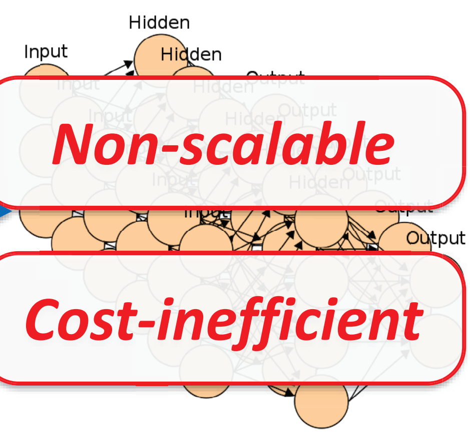
- **What is the meaning of this arm motion?**

  Motion inference

Motion Inference



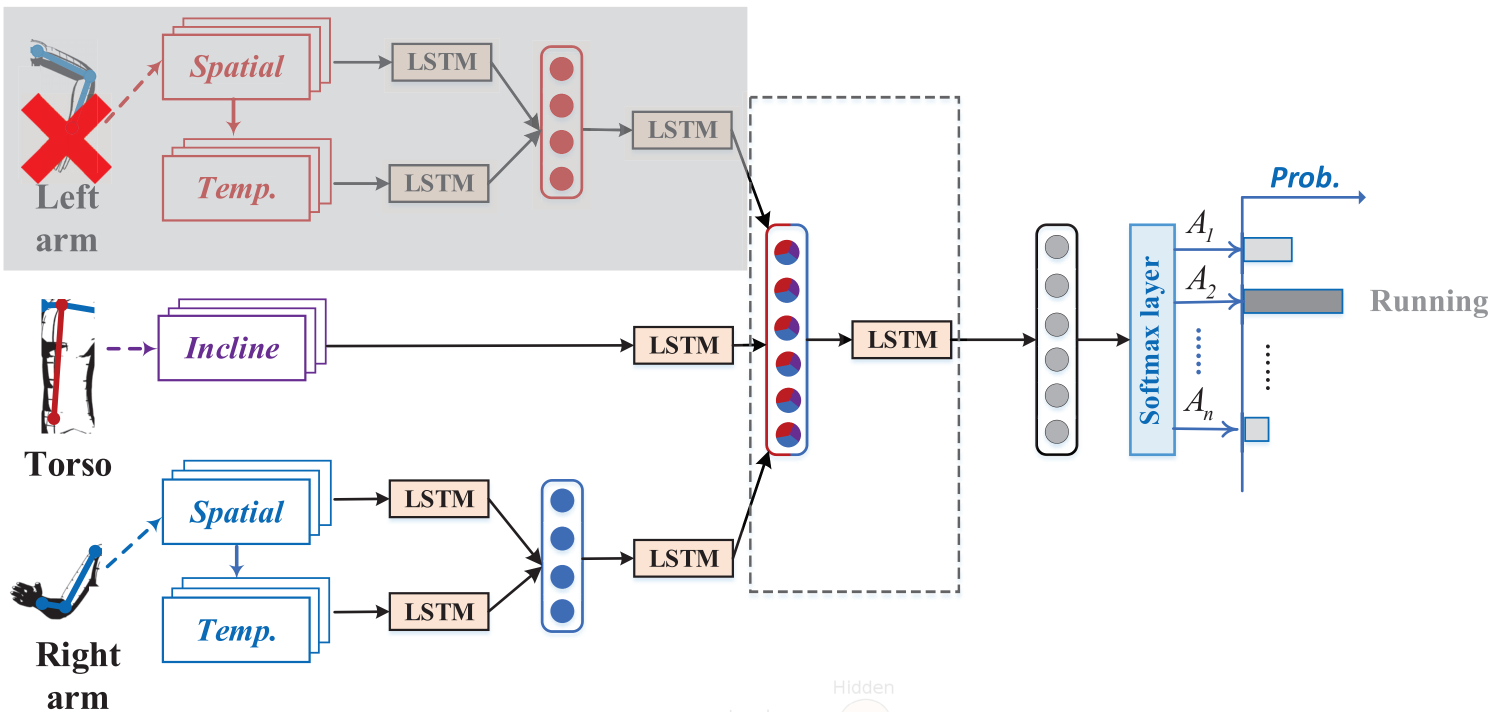
6 combinations of missing inputs



Non-scalable

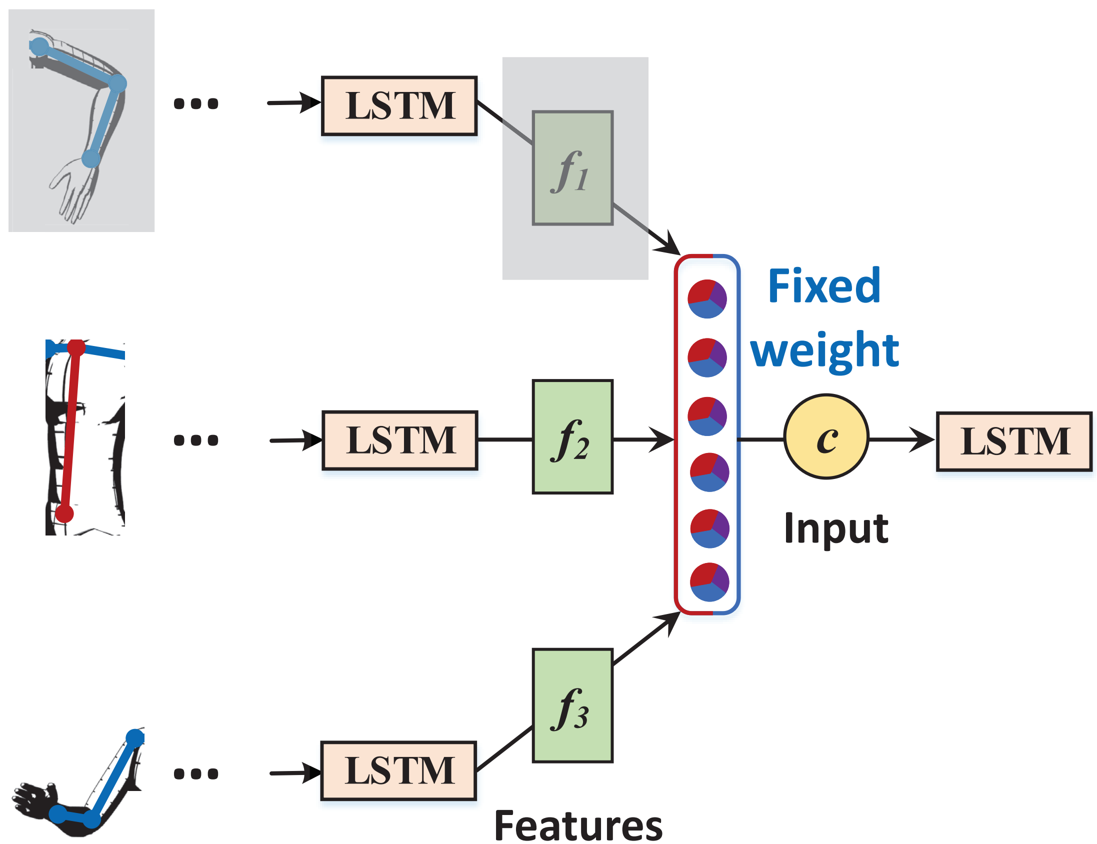
Cost-inefficient

Motion Inference

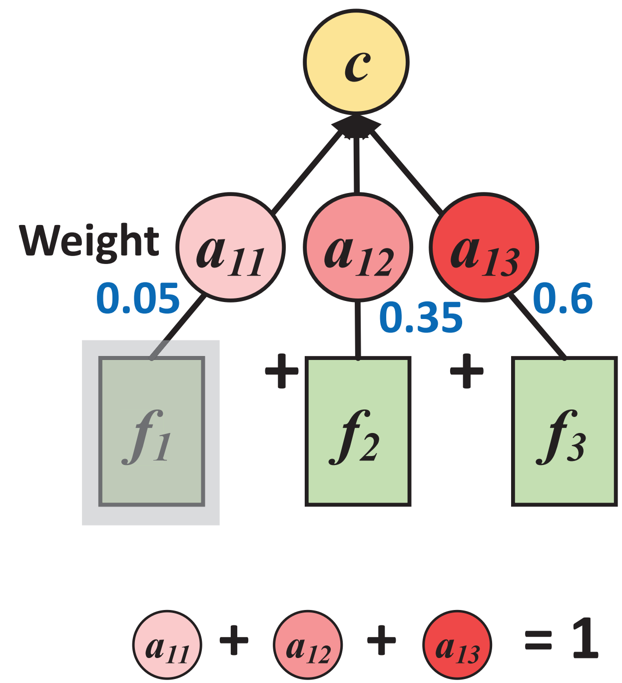


Handle **all** combinations using **one** network?

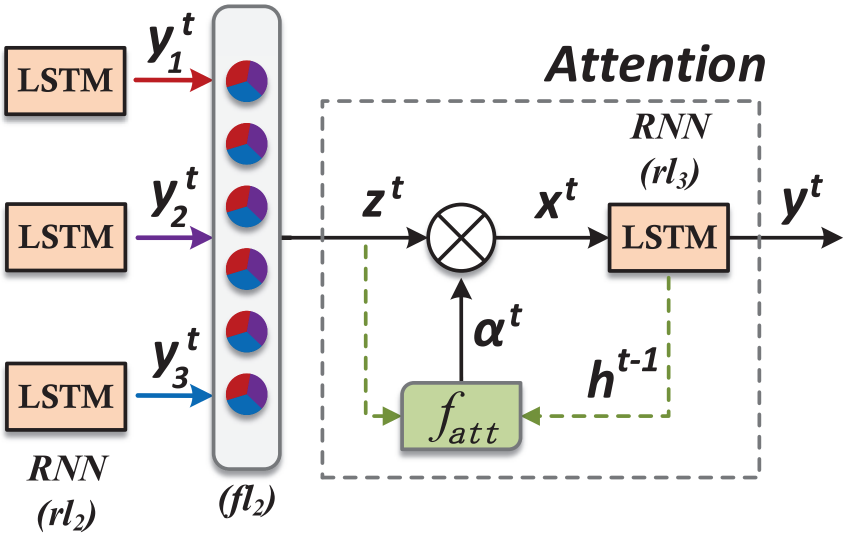
Our idea



- **Adaptive design**



Attention-based network adaption



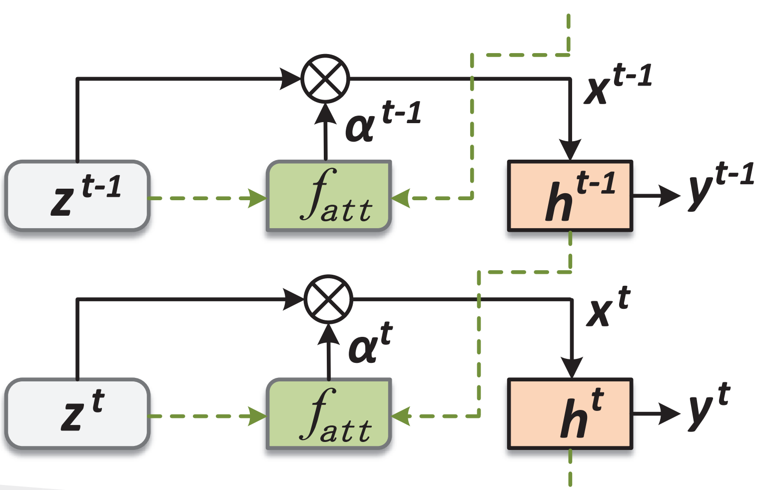
Features: $z^t = \{z_1^t, z_2^t, z_3^t\}$

Input: x^t

$z^t \to x^t$: Weighted fusion

$$x^t = \emptyset(\{z_r^t\}, \boxed{\{\alpha_r^t\}}), r = 1,2,3$$

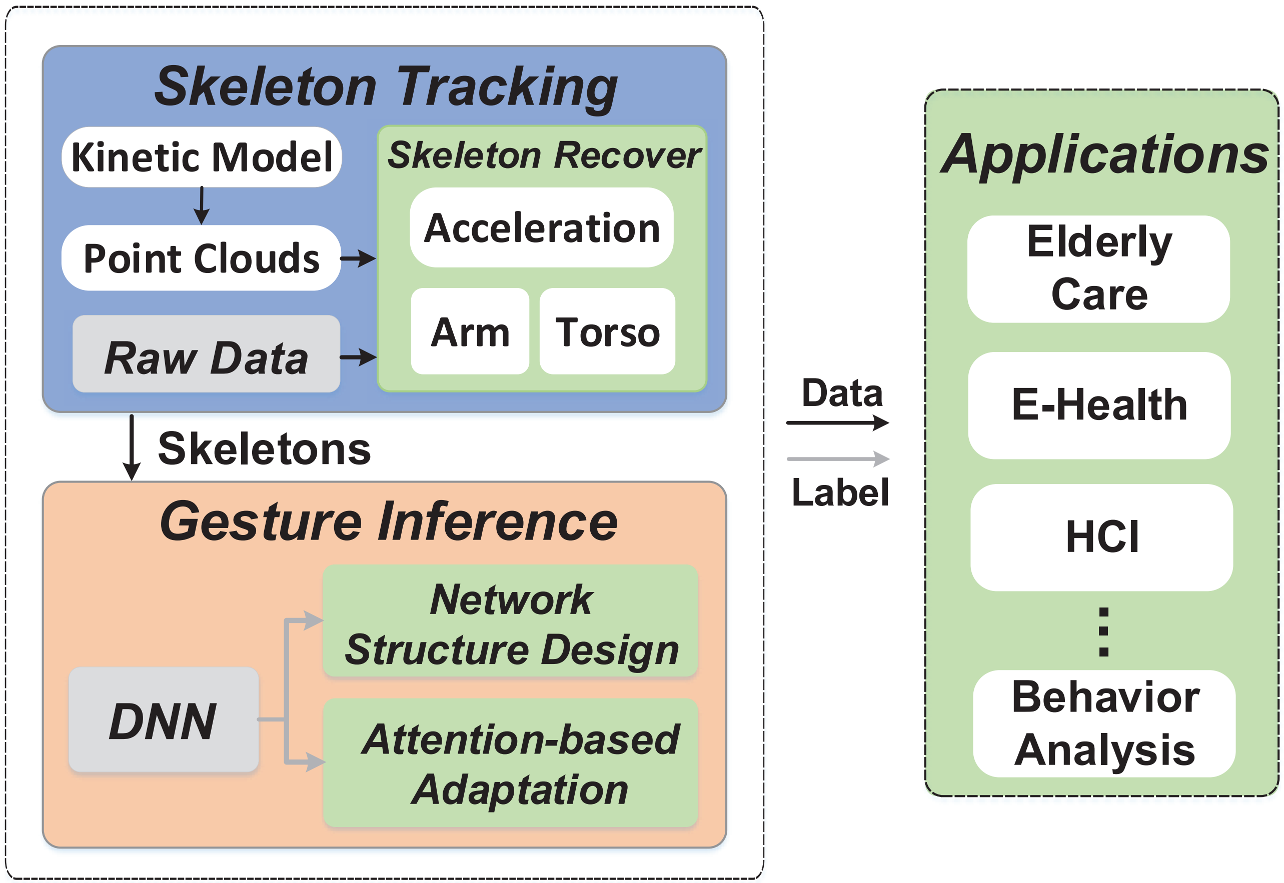
Updated weights



- **Weight update**
  - aligning with the activity descriptor h_{t-1}

$$\alpha^t = f_{att}(z^t, h^{t-1})$$
$$att^t = Relu(z^t + Wh^{t-1} + b)$$
$$\alpha^t = f_{soft-max}(Uatt^t)$$

ArmTroi Implementation



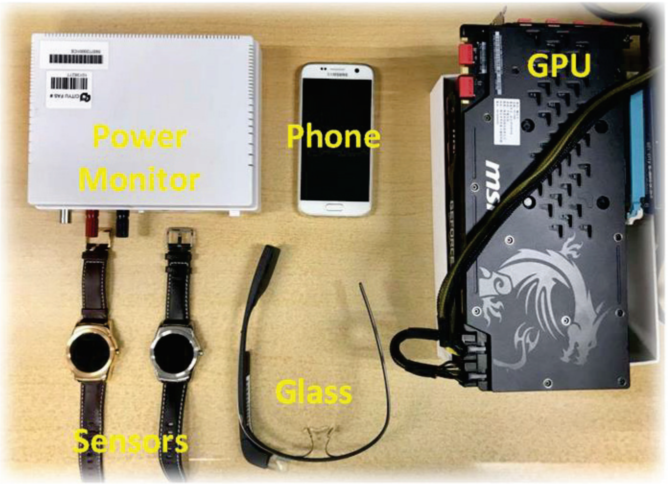
Experiment setup

- Participants: 7 volunteers
- Dataset:

Categories	Gestures
Daily gestures (4)	*shake hands, make a call, open a door, drink water*
Free-weight (10)	*front raise (a/p), biceps curl (a/p), bent over single arm, chest fly (i/s) bench press (i/s), lateral raise*
Customized (3)	*push, pull, circle*

Daily activities

Table 1: Targeted gestures in ArmTroi. The *a, p, i, s* stand for alternating, in parallel, incline and sitting, respectively.

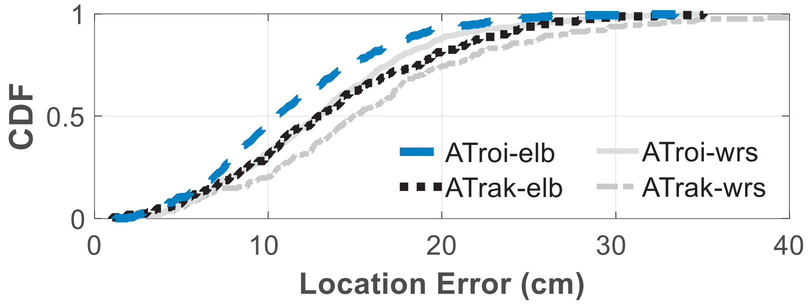


- Training: Intel i7-6700 CPU and Nvidia GTX 1080Ti GPU
- Running: SAMSUNG Galaxy S7

香港城市大學 City University of Hong Kong

Evaluation

- Skeleton tracking

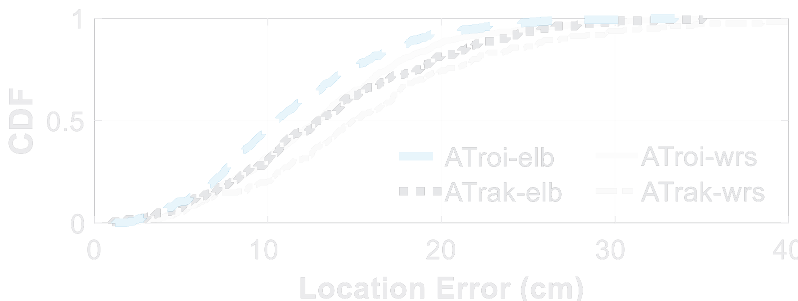


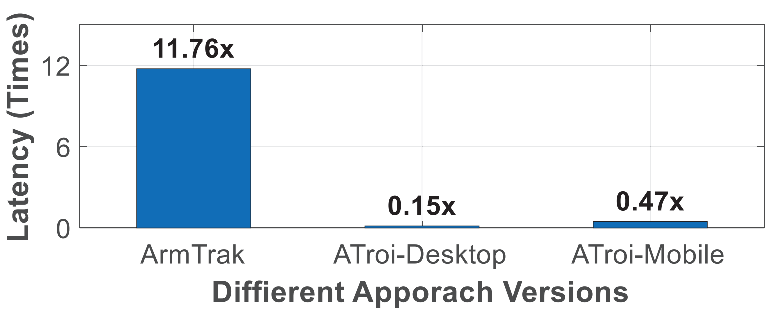
- **ArmTrak [1]**
  - Elbow: 12.94cm
  - Wrist: 14.91cm
- **ArmTroi**
  - Elbow: 10.53cm
  - Wrist: 12.94cm

[1] “I am a smartwatch and I can track my user’s arm”, in Proc. of ACM MobiSys, 2016.

Evaluation

- Skeleton tracking



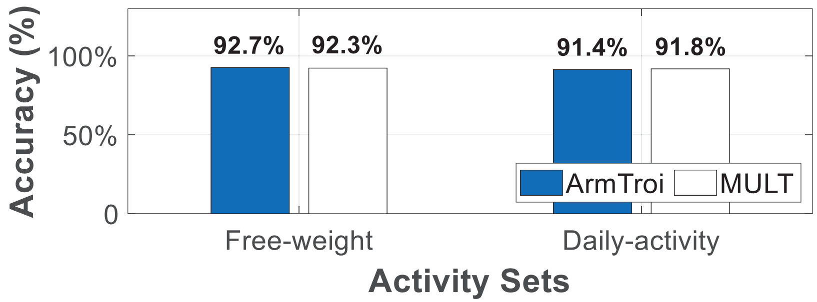


- **ArmTrak [1]**
  - Elbow: 12.94cm
  - Wrist: 14.91cm
- **ArmTroi**
  - Elbow: 10.53cm
  - Wrist: 12.94cm
- **Our latency**
  - Desktop: 0.15x
  - Phone: 0.47x

[1] "I am a smartwatch and I can track my user's arm", in Proc. of ACM MobiSys, 2016.

Evaluation

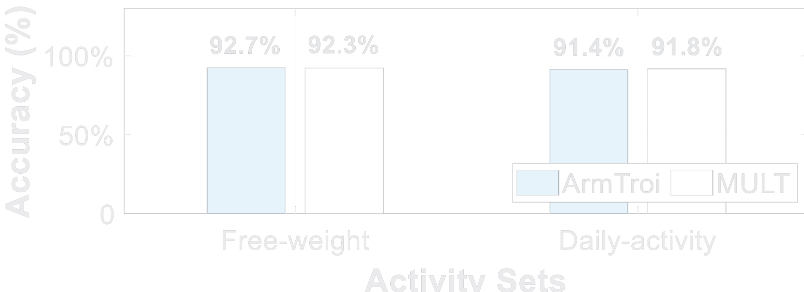
- Motion inference

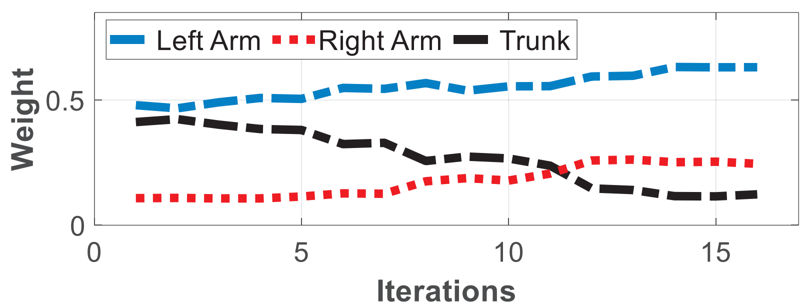


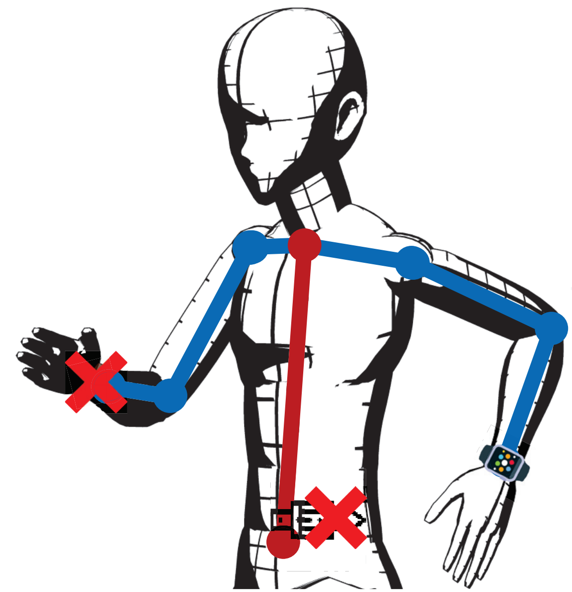
- **Baseline: MULT**
  - Each combination of missing input
- **Accuracy with full set**
  - FW: 92.7% vs 92.3%
  - DA: 91.4% vs 91.8%

Evaluation

- Motion inference







- **Weight updating**
  - Available input: Left Arm
  - LA’s weight increases

Conclusion 1, 2, 3

1. **One goal:**
    - Understanding **human arm motions**
2. **Two aspects:**
    - **Real-time** tracking
    - Motion inference **tolerant to missing inputs**
3. **Three techniques:**
    - HMM state **reorganization**
    - **Hierarchical** search
    - Attention-based network **adaption**