INTELLIGO-GRAIL
GRAphics and Interaction Lab
UNIVERSITÀ di VERONA
Department of ENGINEERING FOR INNOVATION MEDICINE

Gestural Interaction in Extended Reality

Marco Emporio

Why gestures matter in HCI

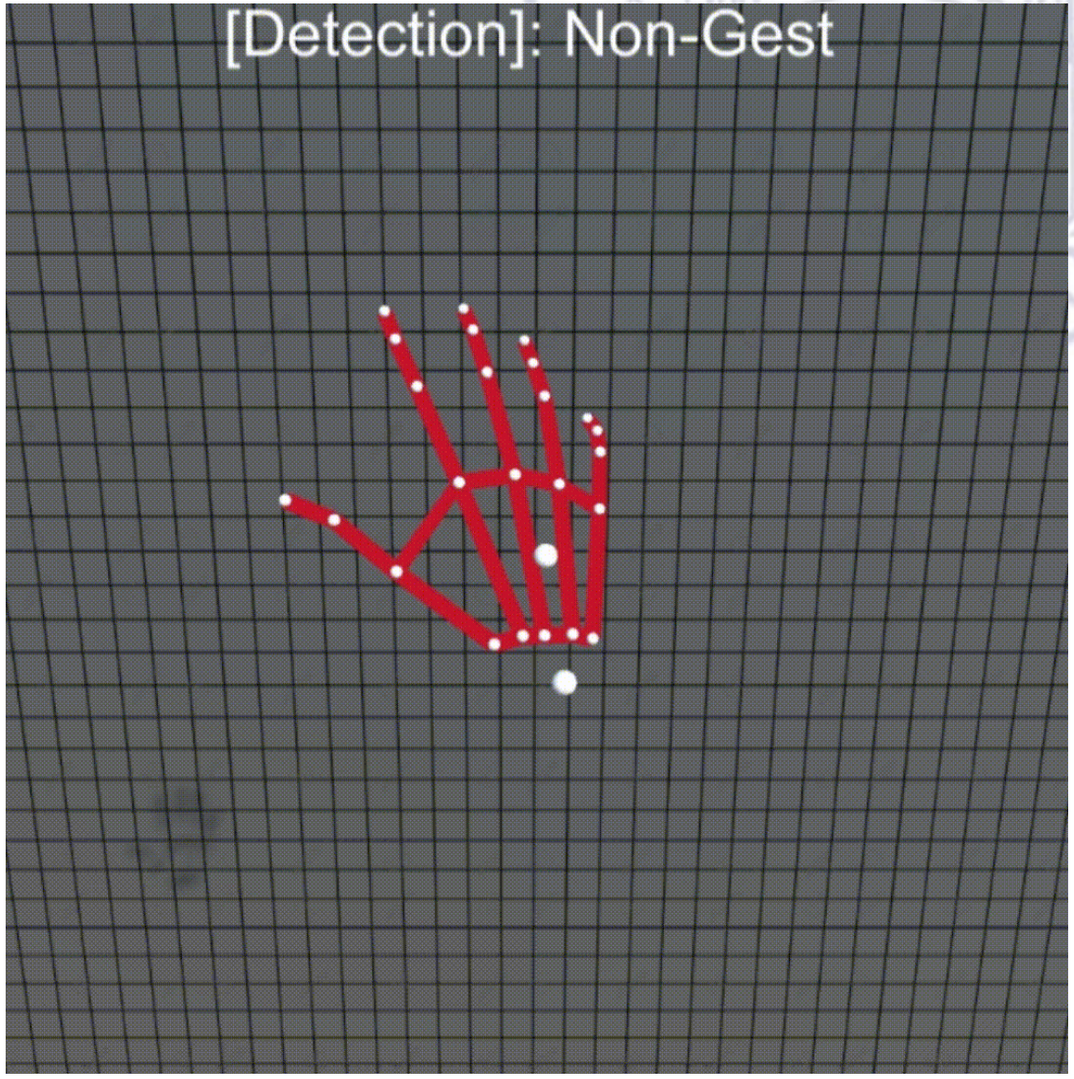
- **Gestures in daily life:**
  - Used for communication and object manipulation.
  - Natural, intuitive, and expressive.

Challenges in Gesture Recognition

- **Hand movements tracking**
- **Variability in gestures:**
  - Different users, styles, speeds and variations due to fatigue or environment.
- **Gestures don’t happen in isolation**
- **Fast recognition time**
- **Minimizing false positives and misclassifications.**

Taxonomy of Hand Gestures

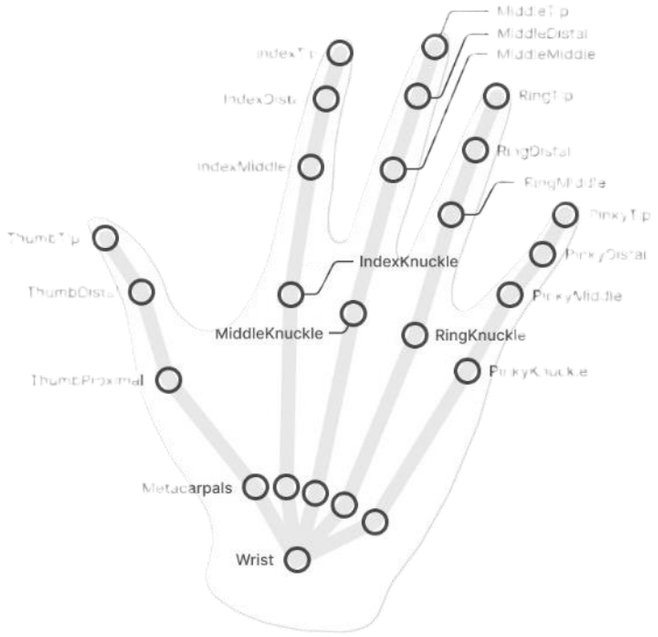
- **Functional Classification**:
  - Semaphoric / Language gestures
  - Gesticulation / Non-gesture movements
- **Temporal Classification**:
  - Static Gestures
  - Dynamic Gestures
    - Coarse , fine and periodic

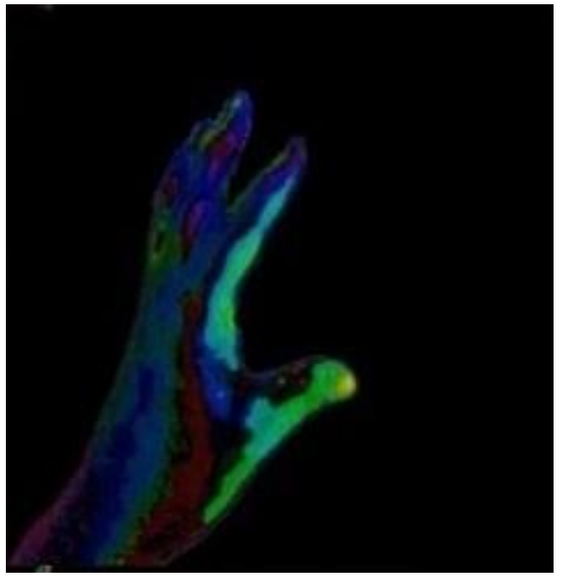


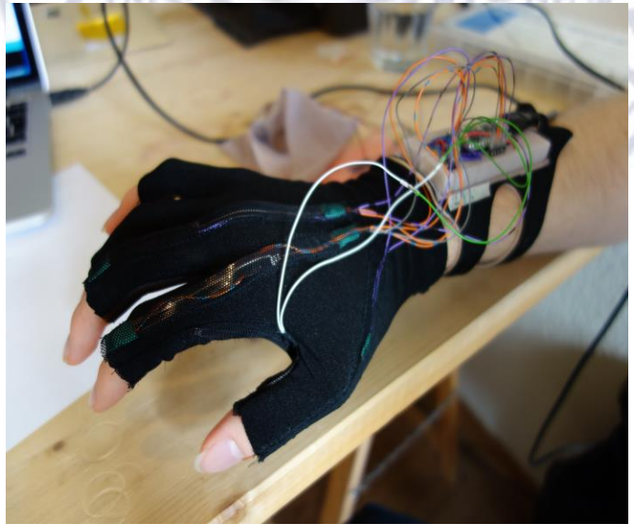
Gesture Invariance

- **Position-Invariant**:
  - Gesture meaning remains the same regardless of translation.
- **Orientation-Invariant**:
  - Gesture meaning is unchanged despite hand rotation.
- **Direction-Invariant**:
  - Gesture meaning is unaffected by movement direction.

Input Stream Encoding for Gesture







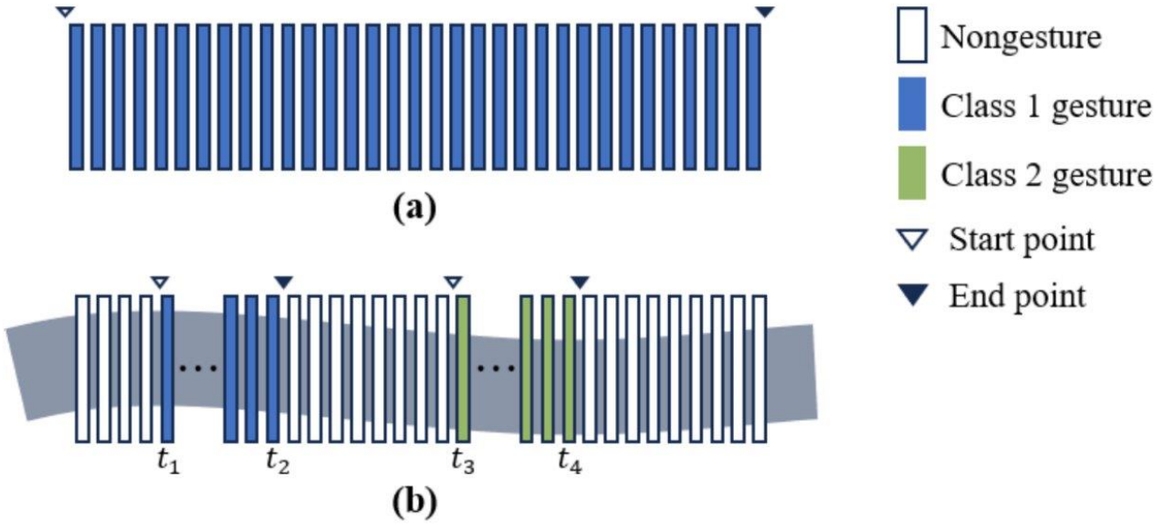
Video data

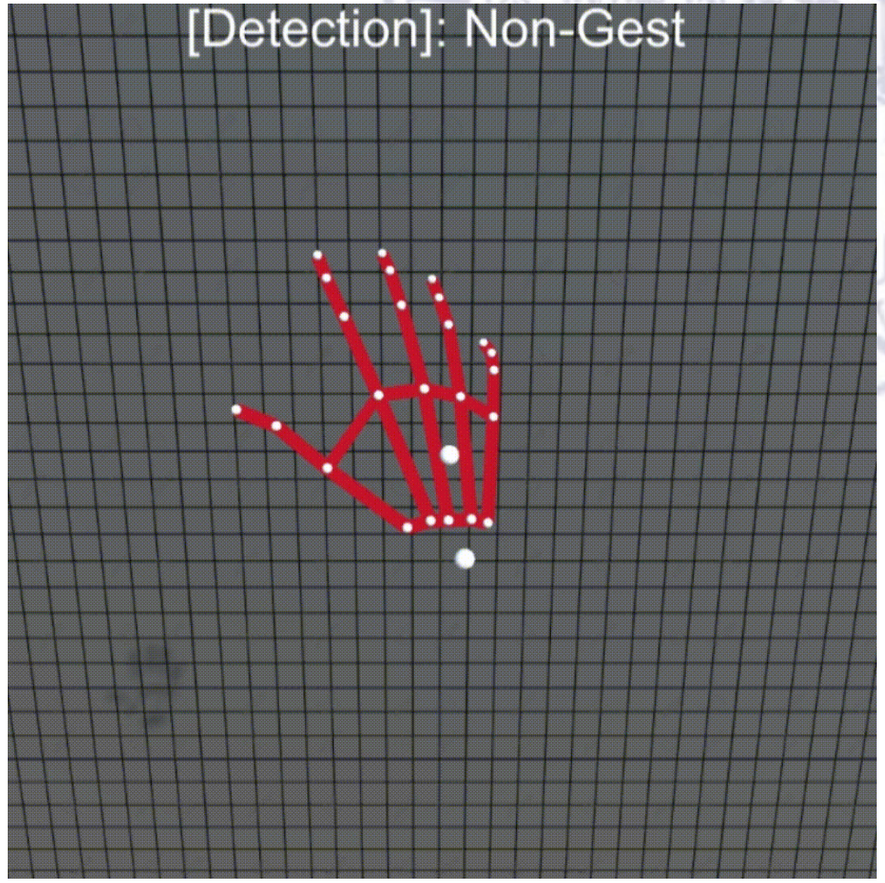
Skeleton data

Image feature maps

Non-visual sensors data

Gesture Recognition

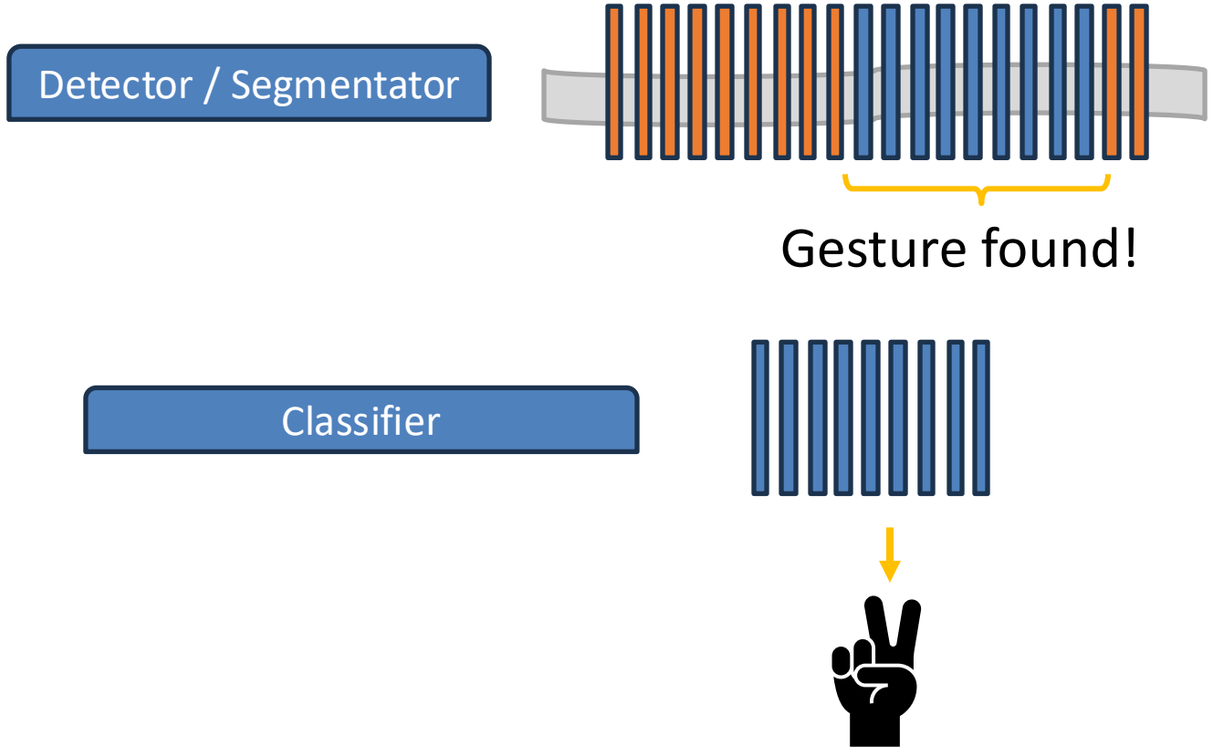




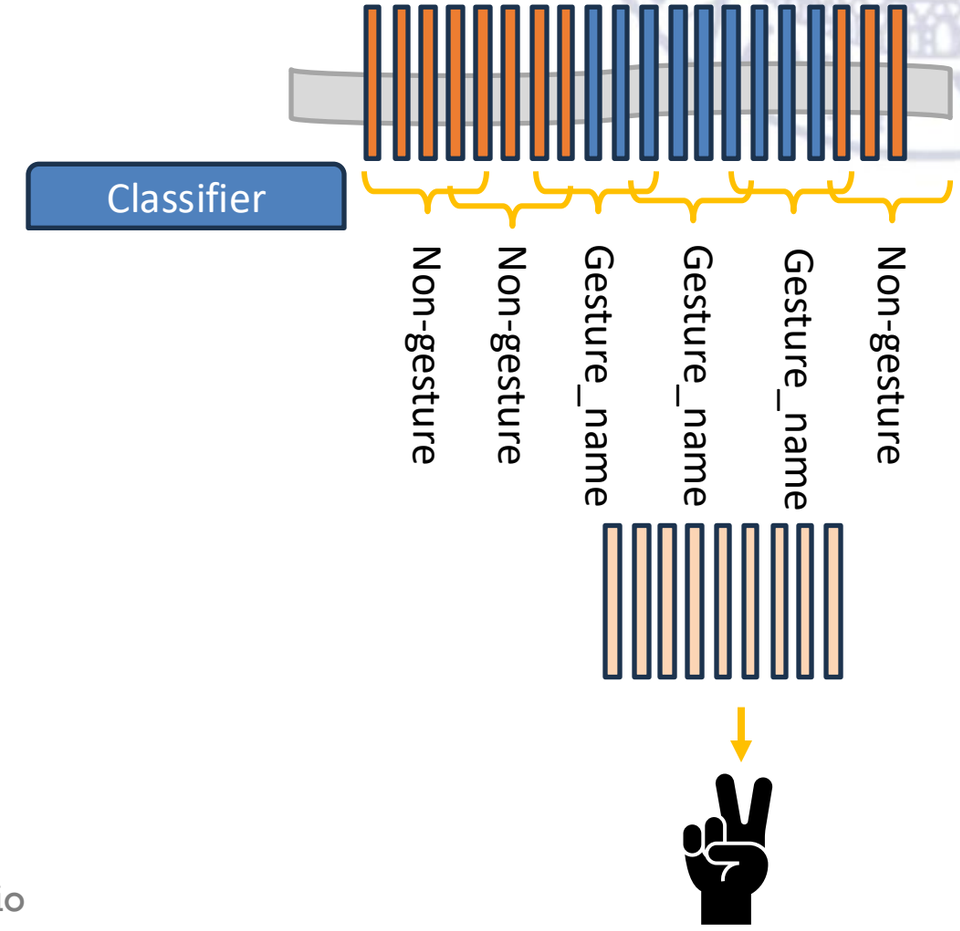
Frame number	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
Ground Truth	ng	ng	ng	deny	deny	deny	deny	ng	ng	ng	two	two	two	two	ng	ng	ng	ng	ng	knob	knob	knob	ng	ng	three	three	ng
Prediction	ng	three	ng	deny	deny	deny	deny	deny	ng	ng	ng	two	two	ng	ng	ng	ng	ng	knob	knob	knob	knob	ng	ng	three	three	ng

Handling Non-gesture

Two Modules

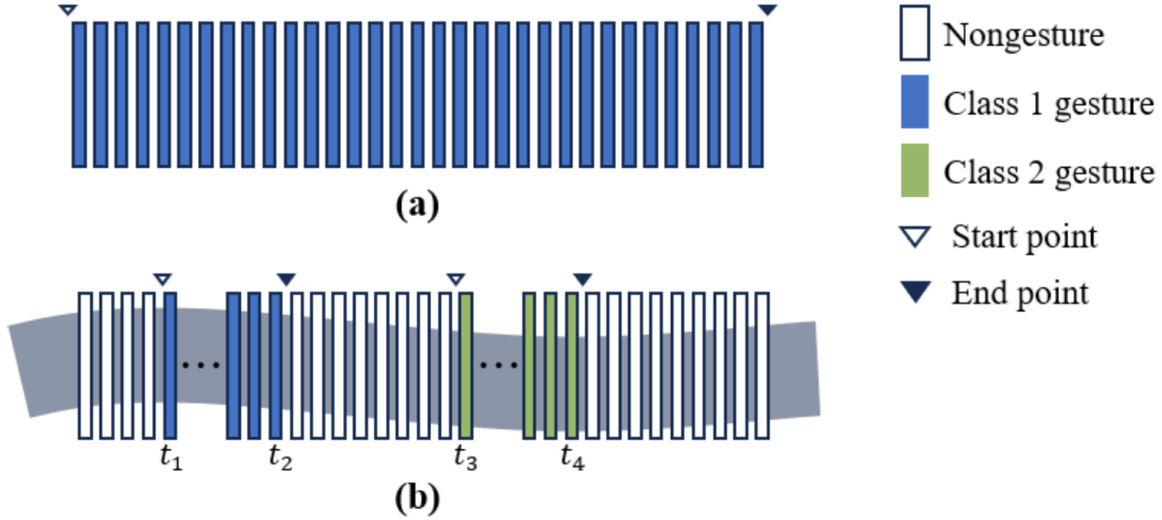


Sliding Window



Evaluation metrics

- **Evaluation metrics**:
  - Frame-Based : treat gesture as labels **sequences** for comparison
  - Event-Based: treat gestures as **events** (start time, end time, label)
  - Efficiency-Based: memory requirements, computational complexity, real-time performance



Frame-Based Metrics

- **True positives (TP)**: correctly recognized gesture frame
- **True Negatives (TN)**: correctly recognized non-gesture frame
- **False Positives (FP)**: misclassified gesture's frame
- **False Negative (FN)**: misclassified non-gesture frame

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

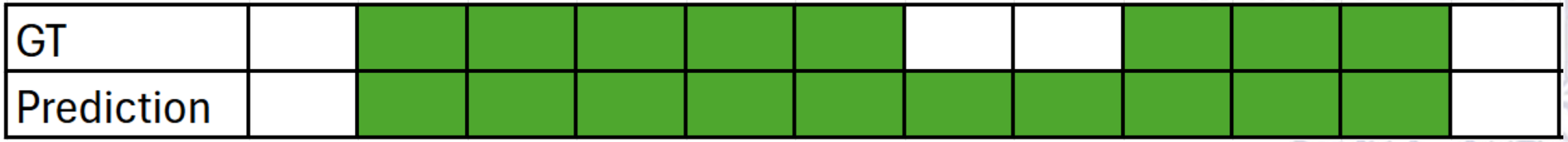
$$Precision = \frac{TP}{(TP + FP)}$$

$$Recall = \frac{TP}{(TP + FN)}$$

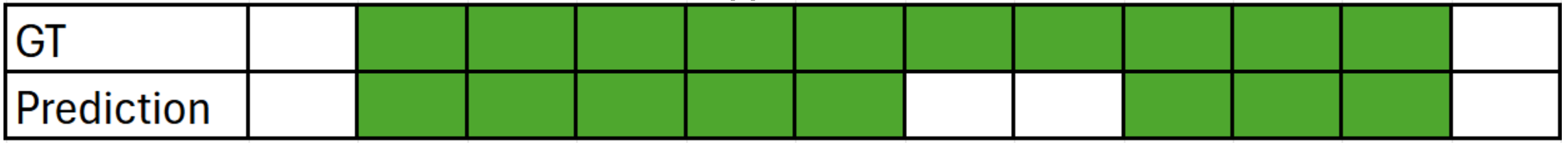
$$F - Score = \frac{2 \cdot (Precision \cdot Recall)}{(Precision + Recall)}$$

Frame-Based Problems

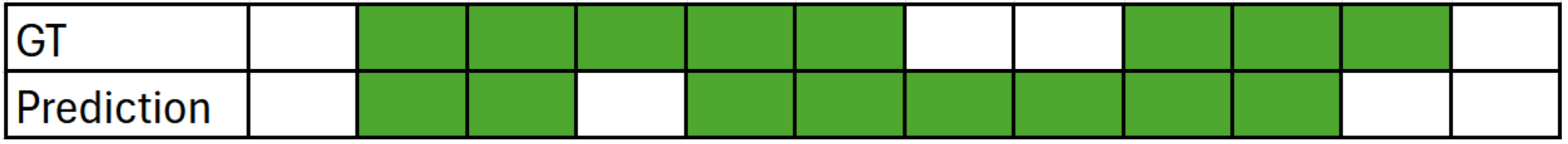
Merged



Fragmented



Merged and Fragmented



Event-Based Metrics

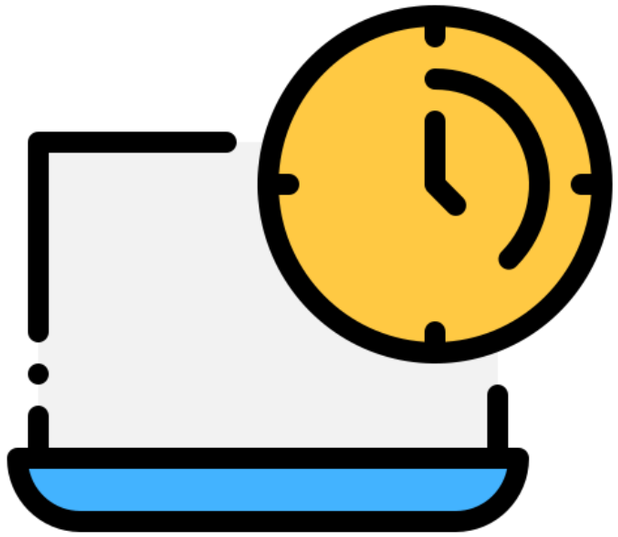
A gesture is considered **correctly detected (CD)** if the start (and end) time is closer than a threshold from the ground truth one, if it has the same label as the ground truth annotation, and its time window has an intersection with the annotated one larger than half the ground truth length. **False predictions (FP)** and ground truth **occurence number (ON)**

$$DetectionRate = \frac{CD}{ON}$$

$$FalsePositiveScore = \frac{FP}{ON}$$

Hardaware Metrics

- Classification time
- Network Delay
- Number of network parameters



Benchmarks use different metrics

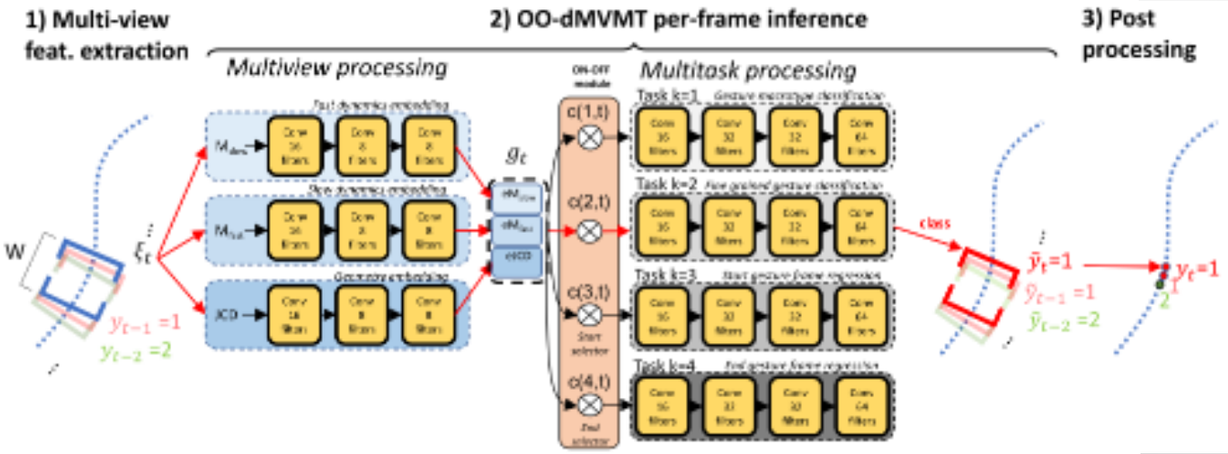
Table 9. Performances of the best methods on the EgoGesture benchmark. (*) metric from Köpüklü et al. (2020).

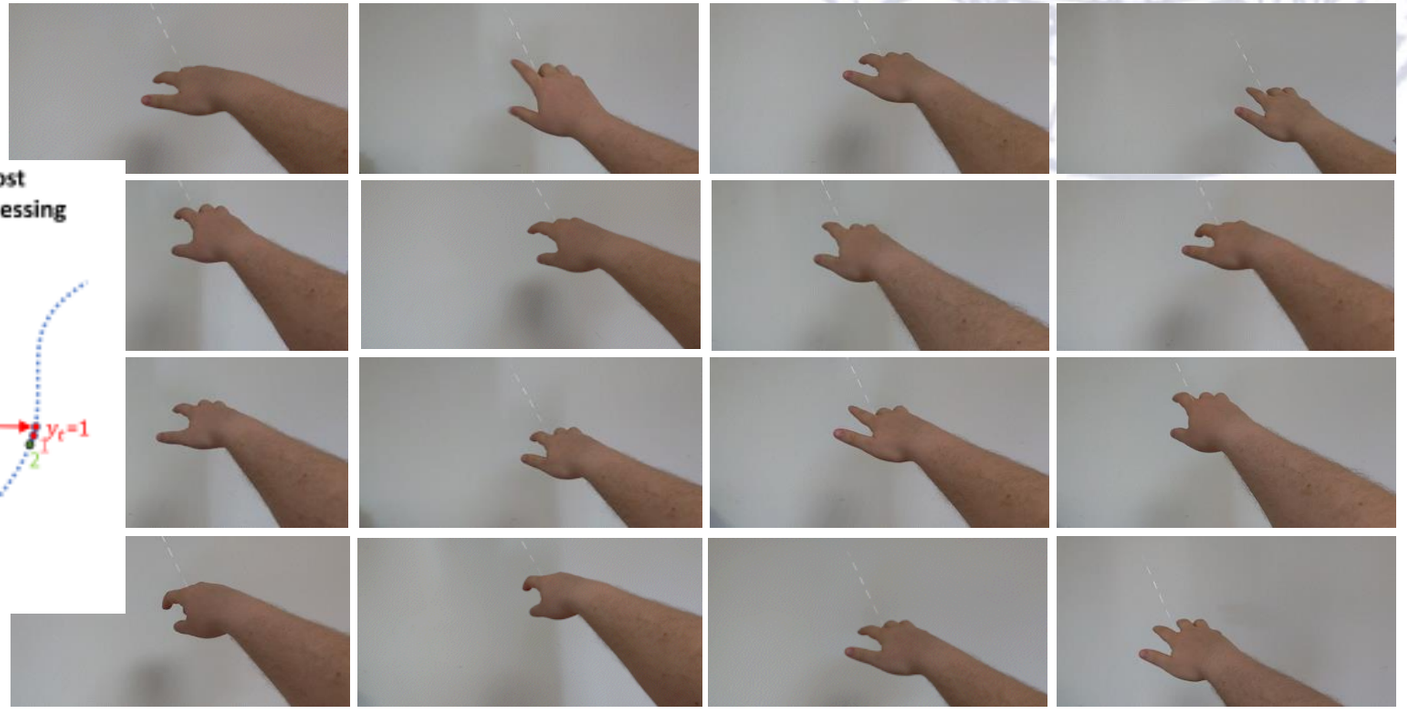
Method	Date	JI (%)	Runtime (FPS)	LA (%)*
TMMF	2021	80.3	NA	NA
C3D	2015	71.8	112	
ResNet+ResNeXt	2020	NA	NA	

Table 8. Performances of the best methods on the NVGesture benchmark. (*) metric from Köpüklü et al. (2020).

Method	Date	τ	AUC	TPR (%)	FPR (%)	NTtD	LA (%)*
R3DCNN	2015	0.3	0.93	88	15	0.56	NA
HSTV	2022	0.16	0.81	85	17	0.2158	NA
ResNet+ResNeXt	2020	0.15	NA	NA	NA	NA	77

OO-dMVMT: A Deep Multi-view Multi-task Classification Framework for Real-time 3D Hand Gesture Classification and Segmentation





Integration of Extended Reality with a Cyber-Physical Factory Environment

- **XR Integration in Industry 4.0/5.0**
  - Real-World Use Cases & Gesture-Based Interaction
  - Synchronizing Data Between Physical and Digital Twins
- **Cyber-Physical Factory & Digital Twin**
  - Enhancing Worker Productivity & Safety
  - Real-Time Monitoring and Control