CIS centre for intelligent sensing

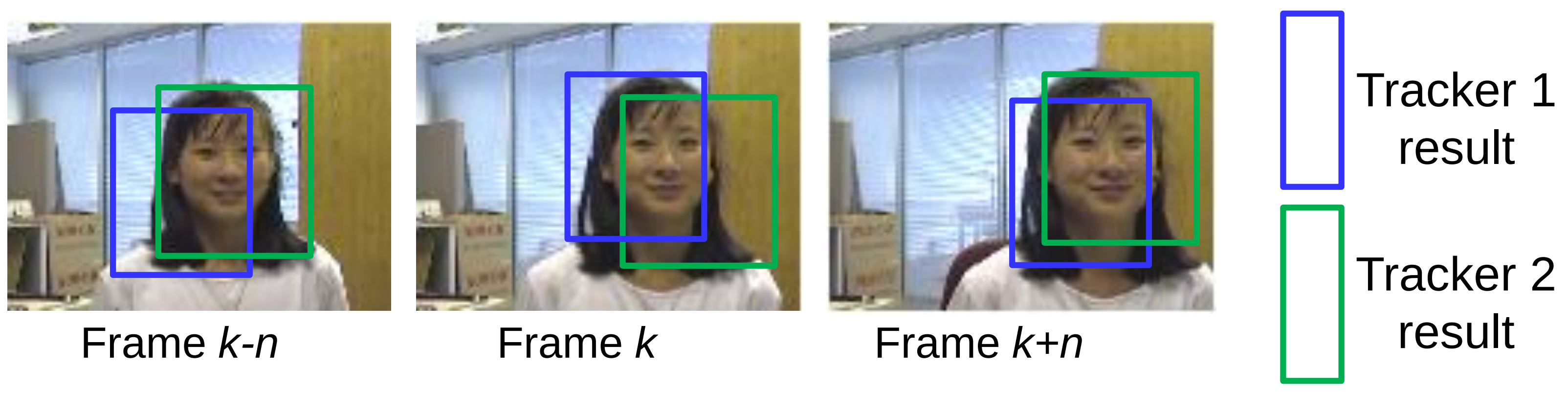
Queen Mary University of London

# ASSESSING TRACKING ASSESSMENT MEASURES

Tahir Nawaz, Fabio Poiesi, Andrea Cavallaro

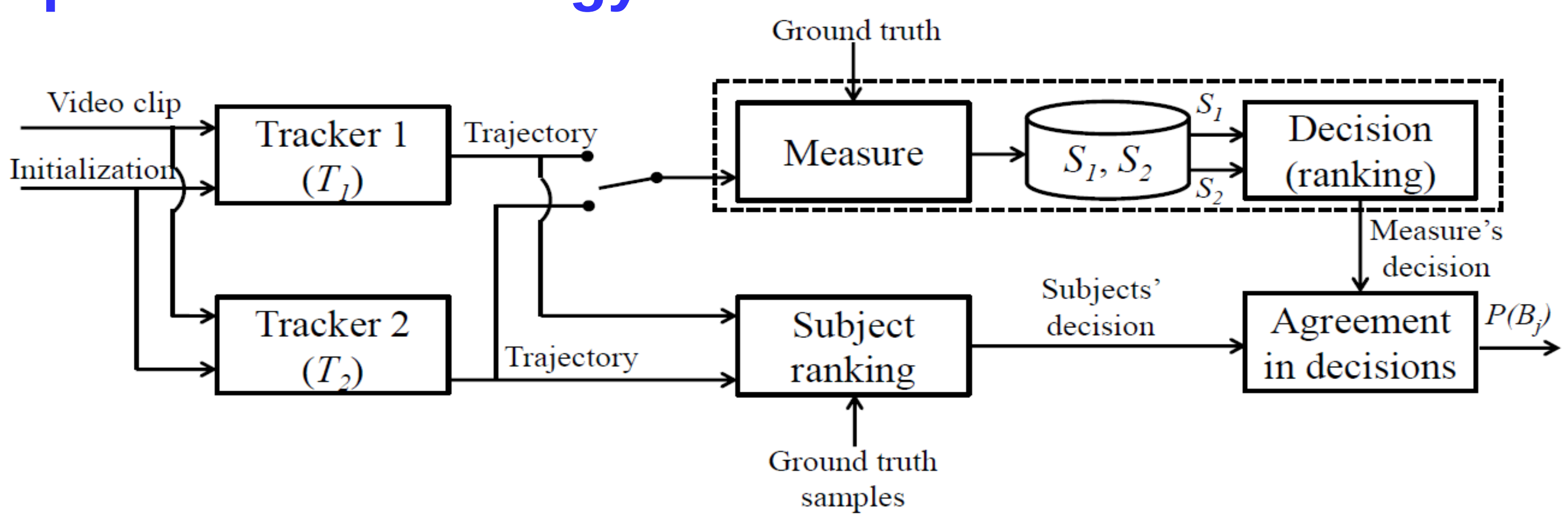
{tahir.nawaz,fabio.poiesi,andrea.cavallaro}@eecs.qmul.ac.uk

## 1. Motivation



- Measure A: tracker 1 performs better than tracker 2
- Measure B: tracker 2 performs better than tracker 1
- Measure C: tracker 1 and tracker 2 perform the same
- How to quantitatively assess performance of measures?

## 2. Proposed methodology



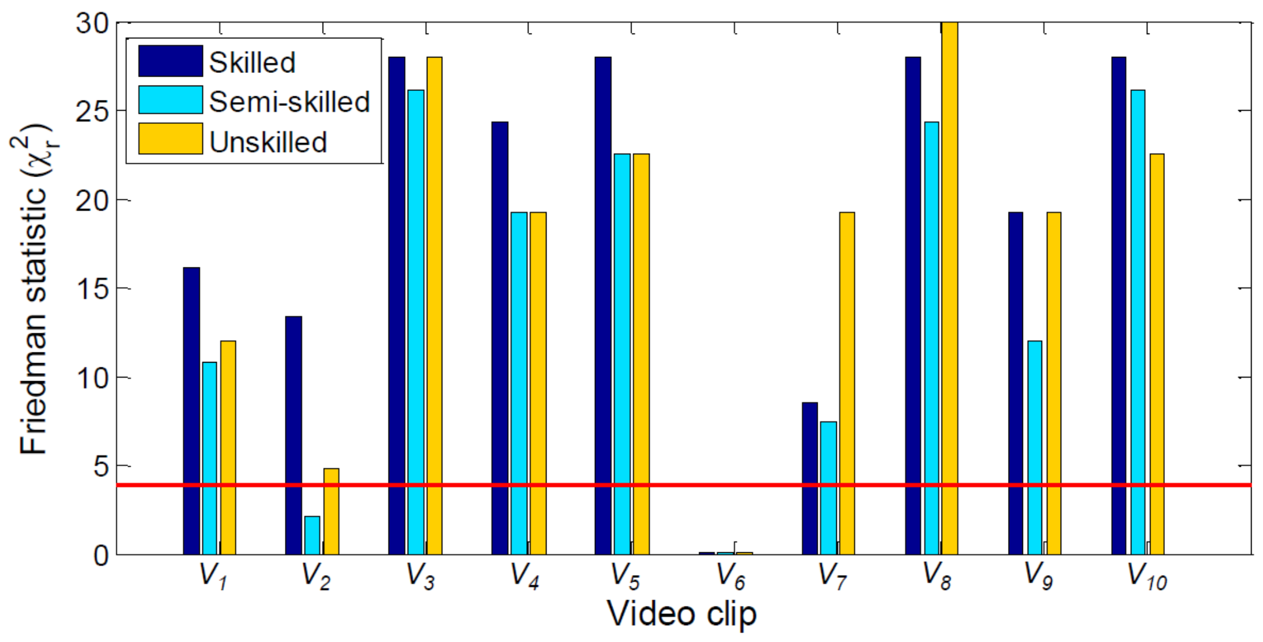
- $S_1$: evaluation score of tracker 1 using the measure
- $P(B_j)$: agreement of measure's decision w.r.t. decisions of human subjects

## 3. Subjective evaluation

- Judgements of (skilled, semi-skilled, unskilled) of human subjects on ranking tracker pairs collected on ten video clips ($V_1$, …, $V_{10}$)
- Statistical significance testing using Friedman's test:

$$\chi^2 = \frac{12}{NF(F+1)} \sum_{f=1}^{F} \left( \sum_{l=1}^{N} \hat{R}_{il}(f) \right)^2 - 3N(F+1)$$

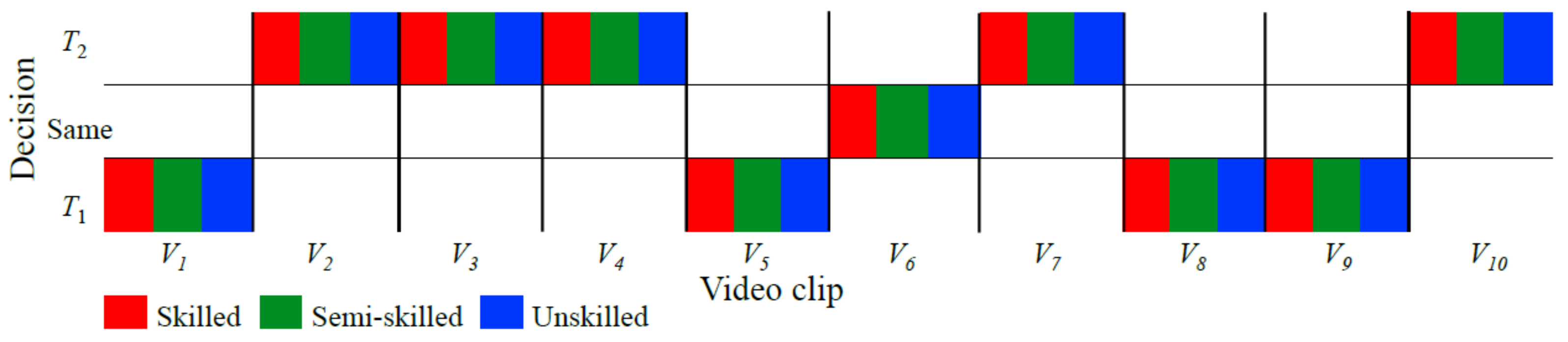
*N*: number of (human) judges; *F*: number of trackers; $\hat{R}_{il}(f)$: rank assigned to tracker $T_f$



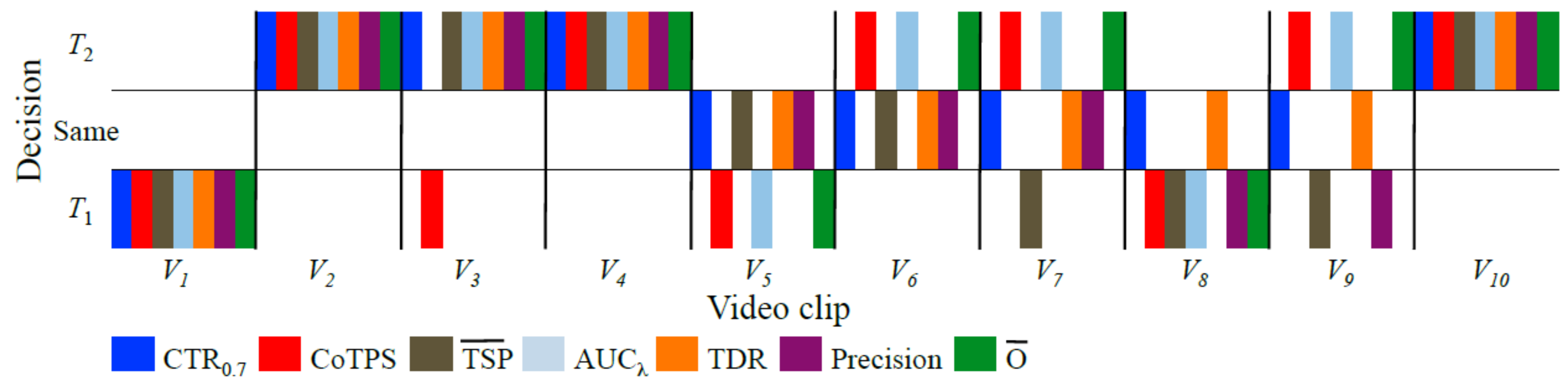
Statistical significance is achieved when the value is above the red line.

## 5. Measure-subject agreement

- **Decision (ranking) of subjects for tracker pairs ($T_1$, $T_2$) on $V_1$, …, $V_{10}$**



- **Decision (ranking) of measures for tracker pairs ($T_1$, $T_2$) on $V_1$, …, $V_{10}$**



- **Amount of agreement ($P(B_j)$) between decisions of a measure and decisions of subjects on *M=10* clips**

$$P(B_j) = \frac{1}{M} \sum_{i=1}^{M} \sum_{r=1}^{3} P(B_j^i | E_r^i) P(E_r^i)$$

The events ($E^i_r$) of a sample of subjects (skilled, semi-skilled, unskilled) where the symbol $\succ$ indicates the preference and $\equiv$ means the two results are indistinguishable.

$$E_1^i = \{T_1(V_i) \succ T_2(V_i)\};\ E_2^i = \{T_2(V_i) \succ T_1(V_i)\};\ E_3^i = \{T_1(V_i) \equiv T_2(V_i)\}$$

$B^i_j$: event of measure *j* with the same probability space as $E^i_r$

| Measure | $\overline{\text{TSP}}$ | $\hat{P}$ | $CTR_{0.7}$ | CoTPS | $AUC_\lambda$ | $\bar{O}$ | TDR |
|---|---|---|---|---|---|---|---|
| Skilled | 0.74 | 0.74 | 0.58 | 0.61 | 0.71 | 0.71 | 0.58 |
| Semi-skilled | 0.68 | 0.67 | 0.52 | 0.57 | 0.66 | 0.66 | 0.52 |
| Unskilled | 0.70 | 0.71 | 0.53 | 0.61 | 0.70 | 0.70 | 0.53 |

## 4. Measures

- Mean Overlap ($\bar{O}$)

$$O_k = \frac{|\hat{A}_{ik} \cap A_{ik}|}{|\hat{A}_{ik} \cup A_{ik}|}$$

$A_{ik}$: area (bounding box) information of the estimation

$\hat{A}_{ik}$: area (bounding box) information of the ground truth

- Precision ($\hat{P}$)

$$\hat{P} = \frac{|TP|}{|TP| + |FP|}$$

$|TP|$: number of true positives

$|FP|$: number of false positives

- Track Detection Rate (TDR) [1]

$$\text{TDR} = \frac{|TC|}{\bar{K}_i}$$

$|TC|$: number of true positive coincidences

$\bar{K}_i$: number of ground-truth points

- Area under the lost-track ratio curve ($AUC_\lambda$) [2]

$$AUC_\lambda = \Delta\tau_2 \sum_{\tau_2=0}^{1} \lambda(\tau_2)$$

$\lambda(\tau_2)$: lost-track ratio corresponding to $\tau_2$

- Combined Tracking Performance Score (CoTPS) [3]

$$\text{CoTPS} = \beta\Omega + (1-\beta)\lambda_0$$

$\Omega$: tracking accuracy

$\lambda_0$: tracking failure

$\beta$: adaptive weighting factor

- Tracking Success Probability ($\overline{\text{TSP}}$) [4]

$$\text{TSP}_k = \frac{\exp(\nu \cdot a(\hat{A}_{ik}, A_{ik}))}{1 + \exp(\nu \cdot a(\hat{A}_{ik}, A_{ik}))}$$

$a(\hat{A}_{ik}, A_{ik})$: amount of overlap

$\nu$: fixed parameter

- Correct Track Ratio ($CTR_{0.7}$) [5]

Dice score: $D_k = \frac{2|\hat{A}_{ik} \cap A_{ik}|}{|\hat{A}_{ik}| + |A_{ik}|}$

*CTR*: %age of frames with $D_k$ > threshold

$CTR_{0.7}$: CTR value corresponding Mean $D_k$ (*MD*) of atleast 0.7 in *MD* vs *CTR* plot [5]

## References

[1] Black *et al.*, *A novel method for video tracking performance evaluation*, in Proc. of VS-PETS Workshop, 2003.

[2] Nawaz and Cavallaro, *PFT: a protocol for evaluating video trackers*, in Proc. of IEEE ICIP, 2011.

[3] Nawaz and Cavallaro, *A protocol for evaluating video trackers under real-world conditions*, IEEE Trans. on IP, 22(4), 2013.

[4] Li *et al.*, *Real-time visual tracking using compressive sensing*, in Proc. of CVPR, 2011.

[5] Salti *et al.*, *Adaptive appearance modeling for video tracking: Survey and evaluation*, IEEE Trans. on IP, 21(10), 2012.

## Acknowledgement

This work was supported in part by the Artemis JU and in part by the UK Technology Strategy Board through COPCAMS Project under Grant 332913. Tahir Nawaz was supported by the Erasmus Mundus Joint Doctorate in Interactive and Cognitive Environments, which is funded by the Education, Audiovisual & Culture Executive Agency (FPA no 2010-0012).

**Code available:** http://www.eecs.qmul.ac.uk/~andrea/pft2/
http://www.eecs.qmul.ac.uk/~andrea/mtte.html