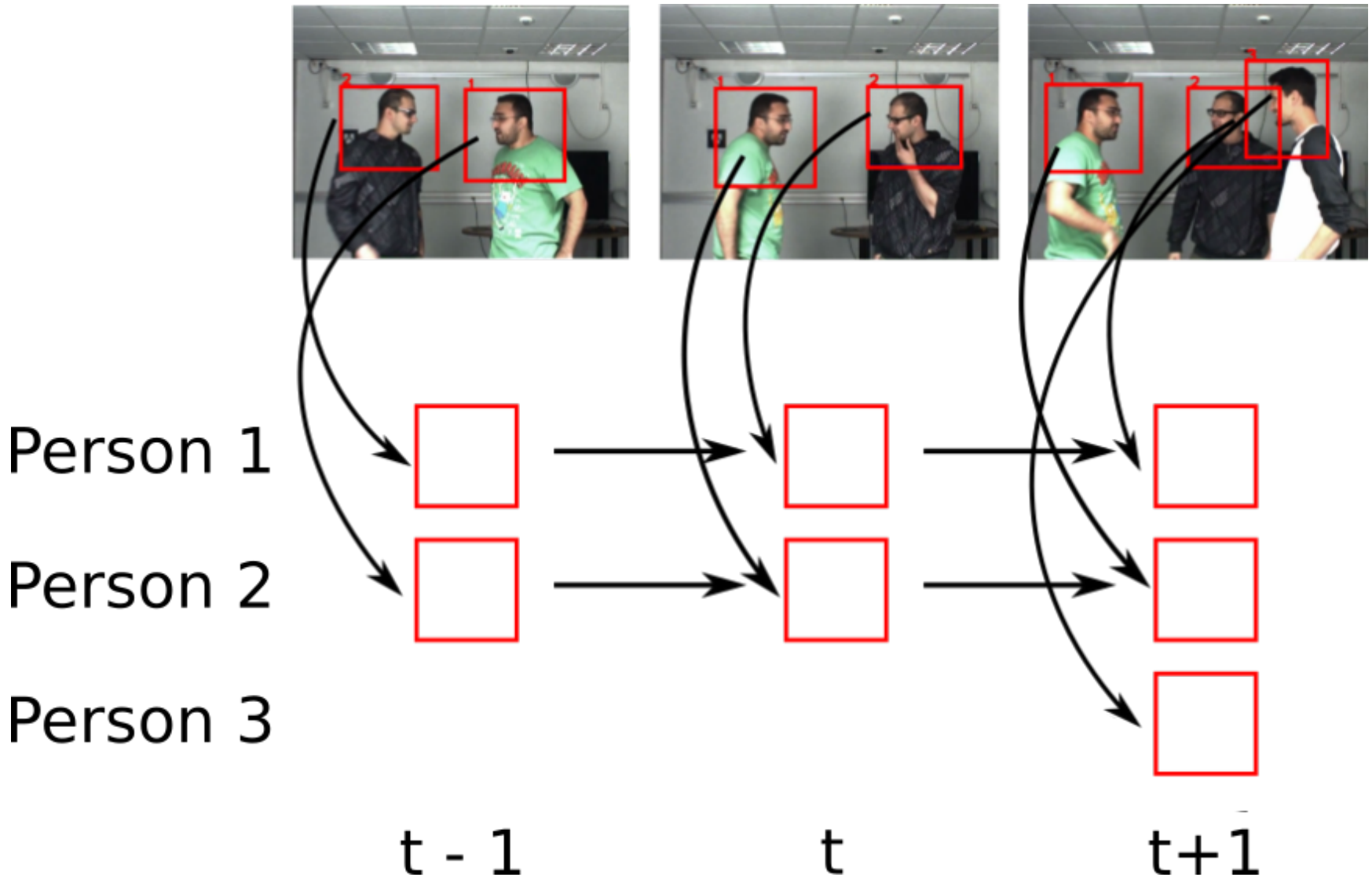
Data Challenge 2017

Guillaume Delorme - Stephane Lathuiliere

Perception - Inria

November 9, 2017

The tracking by detection framework



Data provided

- An input video
- Pose estimation extracted from each frame of the video (*detections.txt*) : coordinates (x,y) of the 18 joints (set to -1 if missed)
- Sound Source Localization (SSL) : for each frame, a heat map of possible positions of the sound source

The github repository is accessible at :
https://github.com/Stephlat/dataChallengePerception

Your goal

- An input video
- Pose estimation extracted from each frame of the video (*detections.txt*) : coordinates (x,y) of the 18 joints (set to -1 if missed)
- Sound Source Localization (SSL) : for each frame, a heat map of possible positions of the sound source

Goal : retrieve for each frame a matching **detection-identity**, and the **speaking activity** for each target.

Data organization

The dataset is constituted by:

- 20 annotated videos in different tracking situation (up to 4 people)
- Raw video
- Sound Source Localization (SSL)
- Pose detection
- Ground truth annotation

Data structure

- Sound Source Localization (SSL) in file *ssl.avi*
  Stored as movie file (.avi), use

```
capssl = cv2.VideoCapture(path+"/ssl.avi")
while not ret2:
        ret2, frameSSL = capssl.read()
        # frameSSL is an np array
        # corresponding to the current image
```

Data structure

- Joints detection in file *detections.txt*

time	Joint1_x	Joint1_y	$(\times 18 joints)$
1	340	450	...
1	580	400	...
2	582	403	...
3	584	402	...

Data structure

- Ground Truth in file *gt.txt*

time	person index	speaking	Joint1_x	Joint1_y	$(\times 18 \textit{joints})$
1	2	0	340	450	...
1	3	1	580	400	...
2	3	1	582	403	...
3	3	1	584	402	...

Code example

```
python visualizeObs.py videoDirectory/

python visualizePred.py videoDirectory/pred.txt
```

Tracking Metrics : MOTA

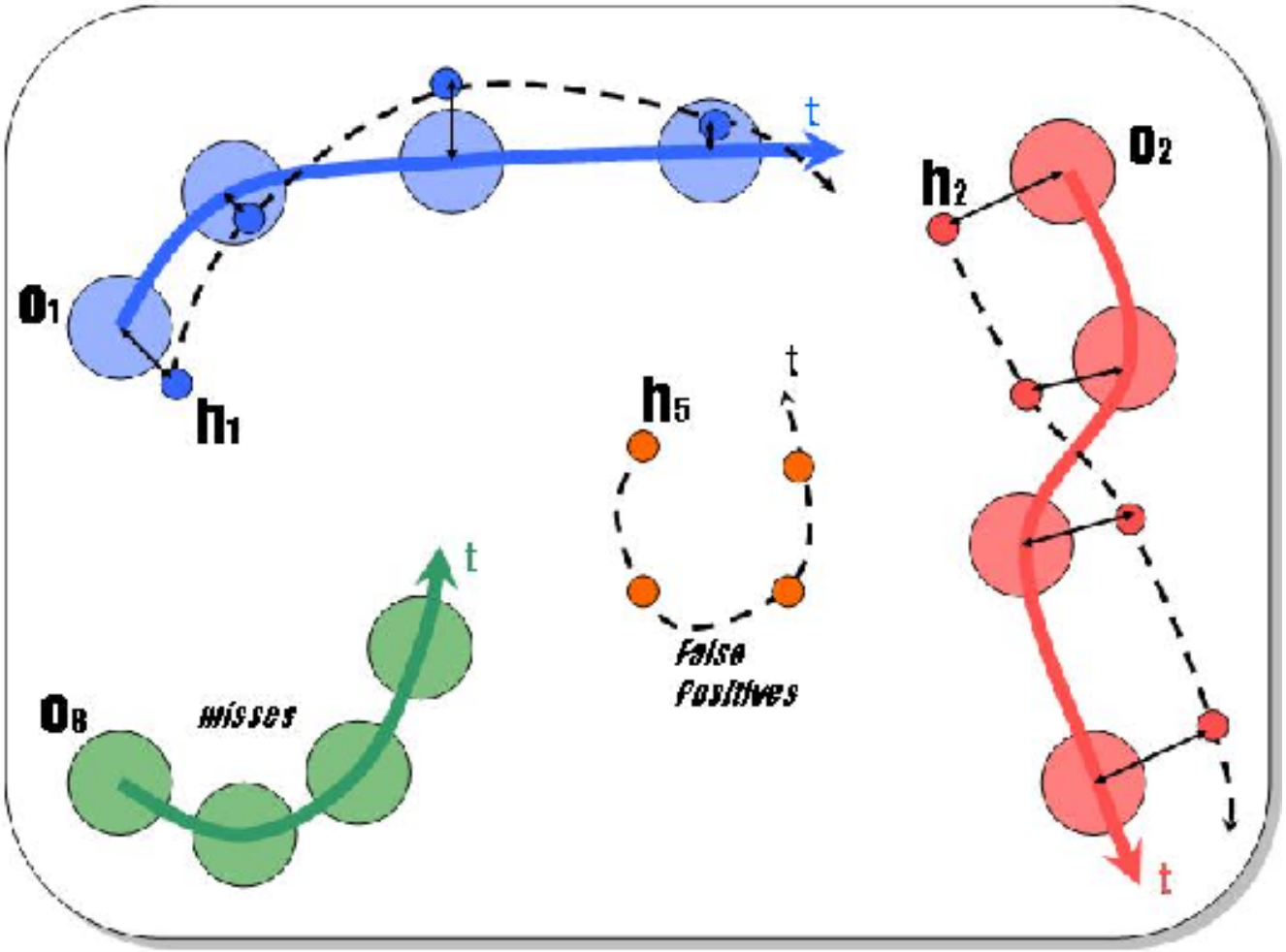
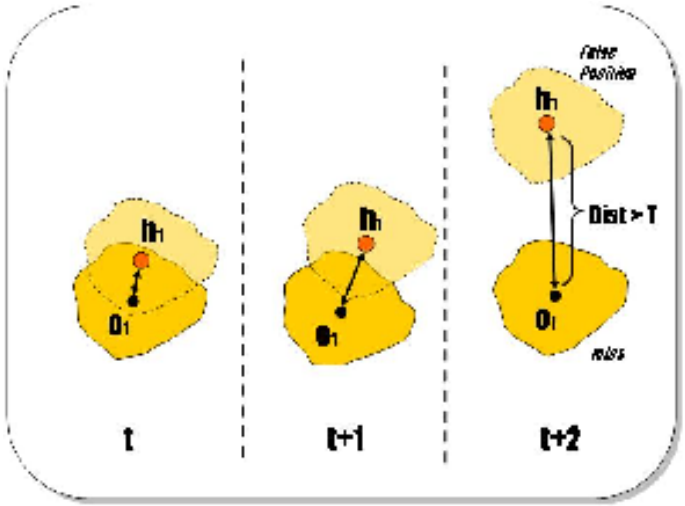
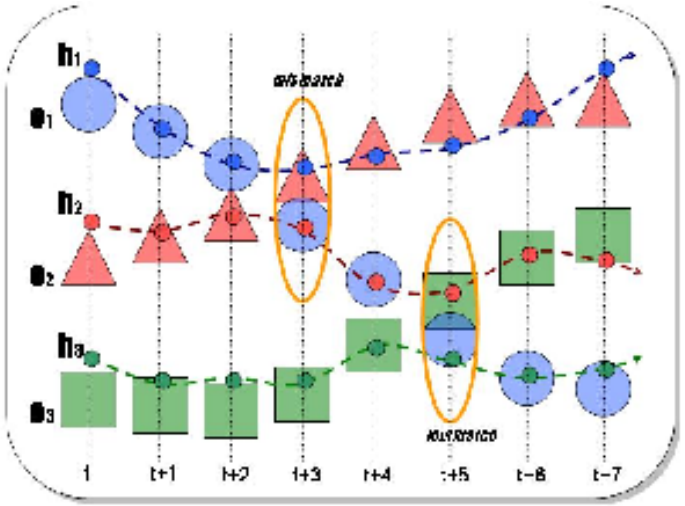


Figure: Bernardin, Keni et al. "Multiple Object Tracking Performance Metrics and Evaluation in a Smart Room Environment." (2006)

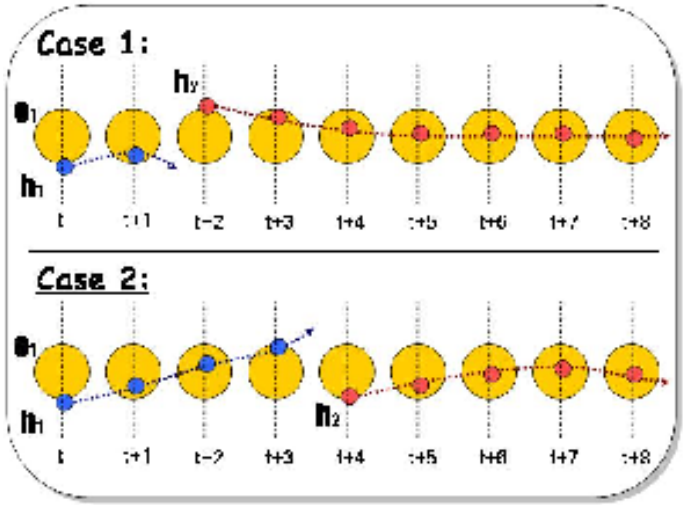
Tracking Metrics : MOTA



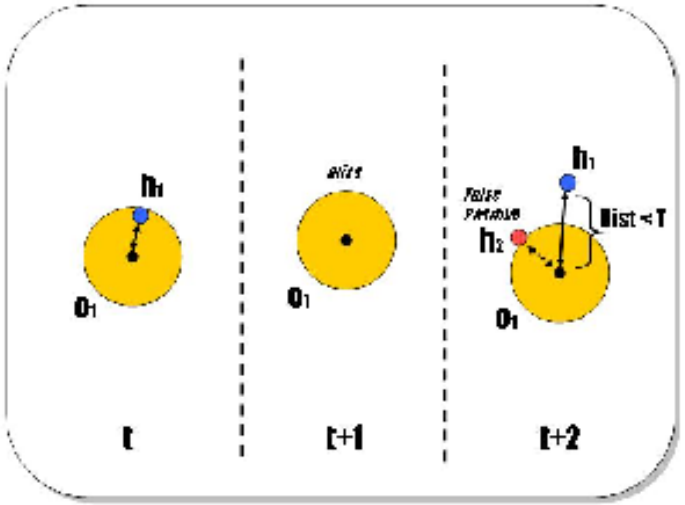
(a)



(b)



(c)



(d)

Tracking Metrics : MOTA

- False Positives at t : fp_t
- Misses (False Negatives) at t : m_t
- Ground truth (actual number of objects) at t : g_t
- Mismatch error : mme_t

$$MOTA = 1 - \frac{\sum_t (m_t + fp_t + mme_t)}{\sum_t g_t}$$