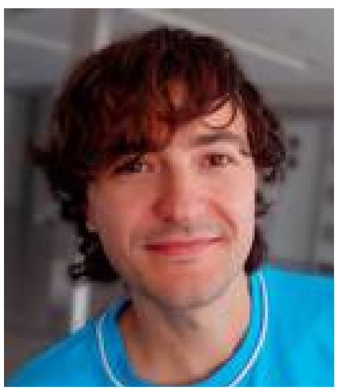
Privacy Preserving Multi-target Tracking




Anton Milan   Stefan Roth   Konrad Schindler   Mineichi Kudo

THE UNIVERSITY of ADELAIDE   TECHNISCHE UNIVERSITÄT DARMSTADT   ETH   北海道大学 HOKKAIDO UNIVERSITY

Visual People Tracking

Applications and Benefits

- ✔ CCTV: Increased safety
- ✔ Automated video analysis
- ✔ Crowd motion estimation
- ✔ Robotic navigation

Visual People Tracking

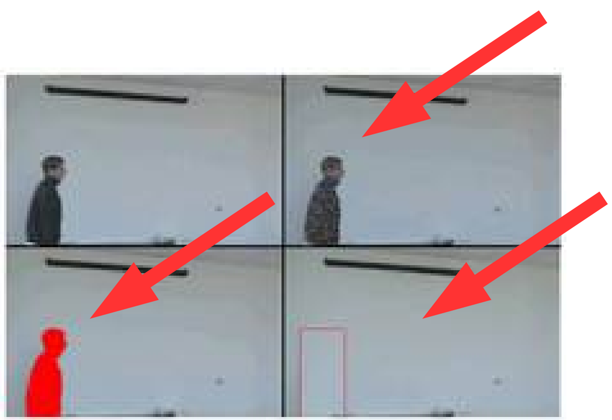
Applications and Benefits

- ✔ CCTV: Increased safety
- ✔ Automated video analysis
- ✔ Crowd motion estimation
- ✔ Robotic navigation

<u>Drawback:</u>
Heavy intrusion of privacy

Existing Alternatives

[Schiff et al., 2009]

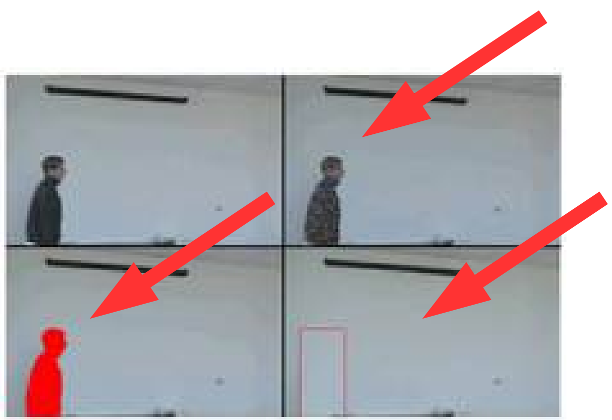


[Wickramasuriya et al., 2005]

[Spindler et al., 2006]

Existing Alternatives

[Schiff et al., 2009]



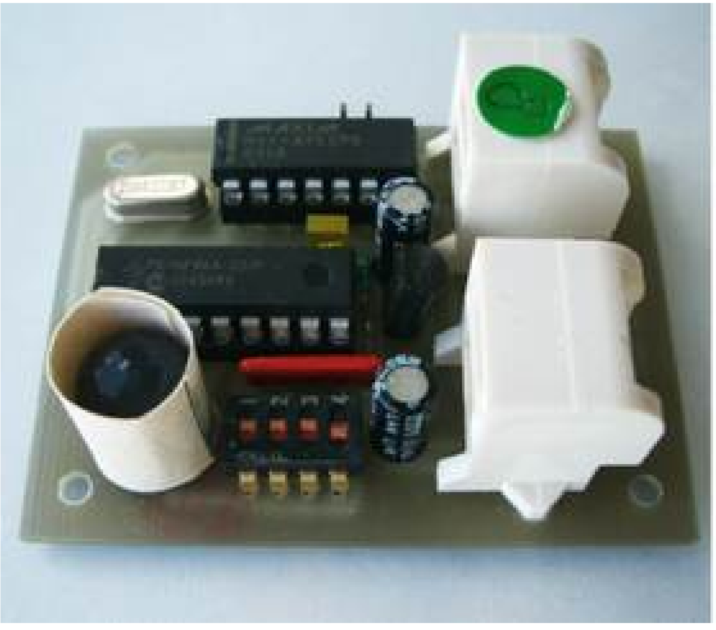
[Wickramasuriya et al., 2005]

[Spindler et al., 2006]

Such systems may fail (or be switched off)

Our Approach

- A different sensor modality
- Existing multi-target tracking techniques

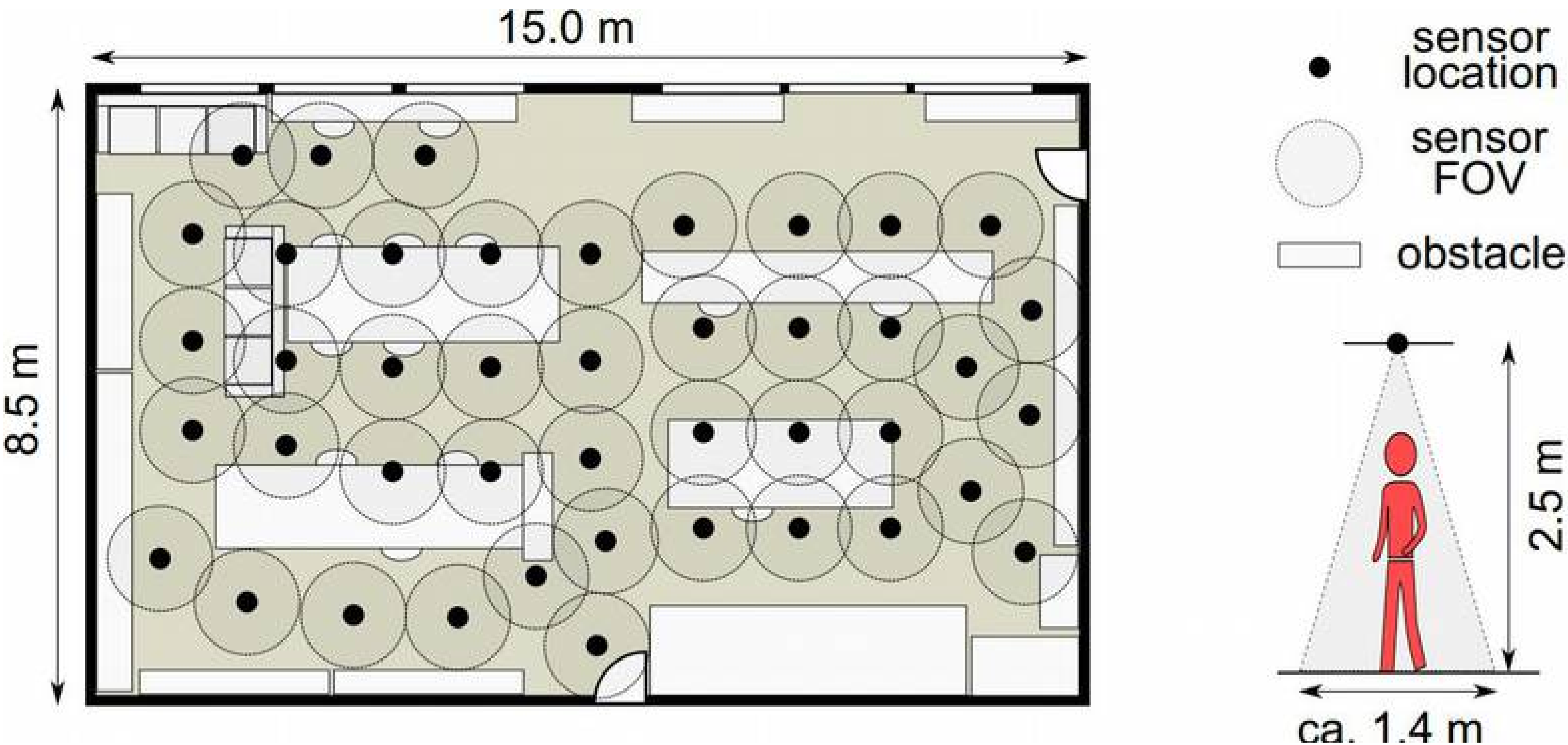


Pyroelectric infrared sensors* ...

...mounted on a ceiling

*Also known as: Infrared motion sensors

The Setup



43 nodes, ca. 3m stride. Total cost: ≈ $100 USD.

Tracking with Infrared Sensors

A mostly unexplored research area!

[Luo et al., 2009]

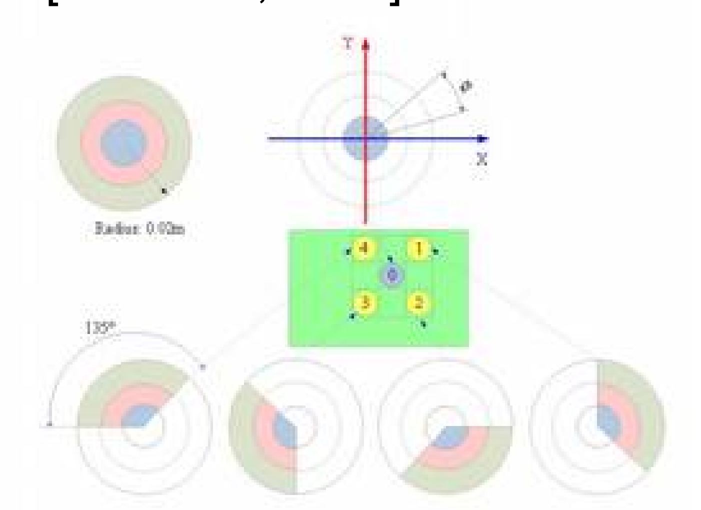
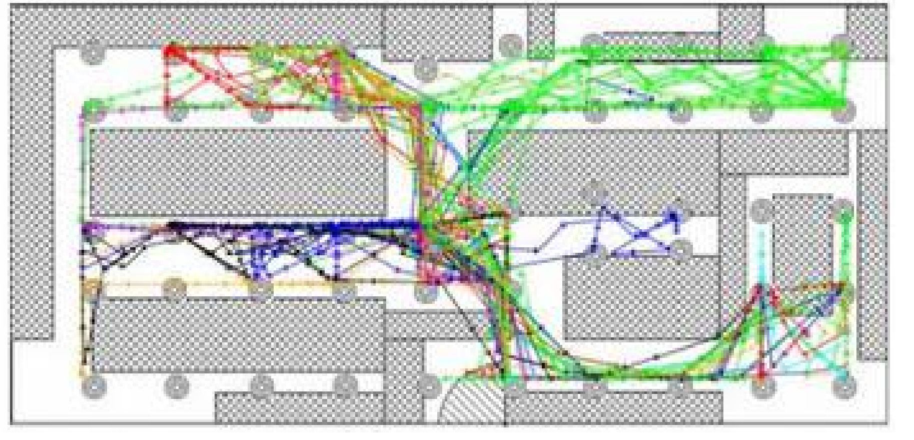


Fig. 2. Sensor Module

[Hosokawa et al., 2009]



- Expensive sensor array with Fresnel lenses

- Limited state space
- Ad hoc algorithm for data association

Benefits

- Individal identification impossible
  - Respects privacy
- Insensitive to lighting conditions
- Low cost

Limitations

- Indoor application only
- Less flexible

Main Challenges

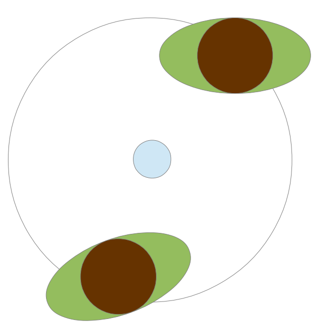
- Extremely low resolution (43 sensors)
- A binary response at 2 Hz per sensor
- No visual (appearance) information
- Poor localization + sensor noise / delay
- Activation by several people
- Multiple measurements by one person

Main Challenges

- Extremely low resolution (43 sensors)
- A binary response at 2 Hz per sensor
- No visual (appearance) information
- Poor localization + sensor noise / delay
- Activation by several people
- Multiple measurements by one person

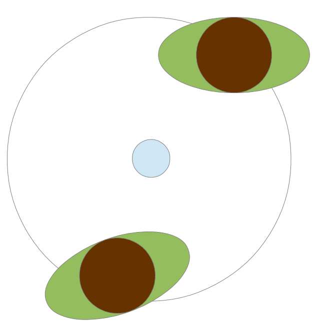
Main Challenges

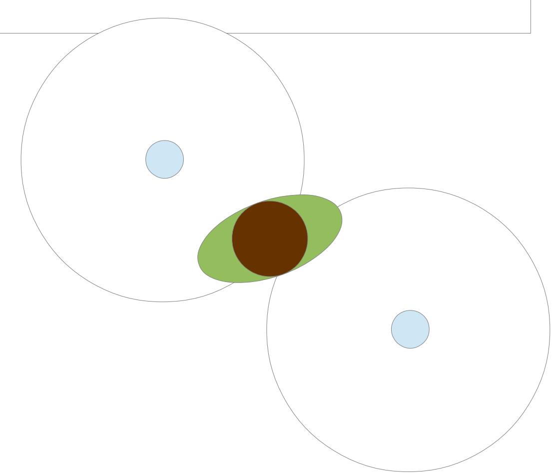
- Extremely low resolution (43 sensors)
- A binary response at 2 Hz per sensor
- No visual (appearance) information
- Poor localization + sensor noise / delay
- Activation by multiple people
- Multiple measurements by one person



Main Challenges

- Extremely low resolution (43 sensors)
- A binary response at 2 Hz per sensor
- No visual (appearance) information
- Poor localization + sensor noise / delay
- Activation by several people
- Multiple measurements by one person

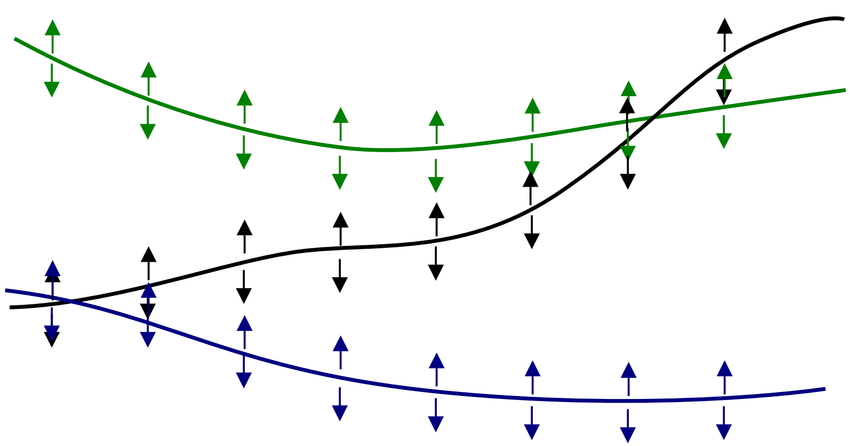




Continuous Energy Minimization

$$E(\boldsymbol{X}) = E_{\text{obs}} + E_{\text{dyn}} + E_{\text{exc}} + E_{\text{per}} + E_{\text{reg}}$$

State vector: *X*, *Y*-locations of **all** targets at **all** frames $\boldsymbol{X} \in \mathbb{R}^d, d \approx 2000$



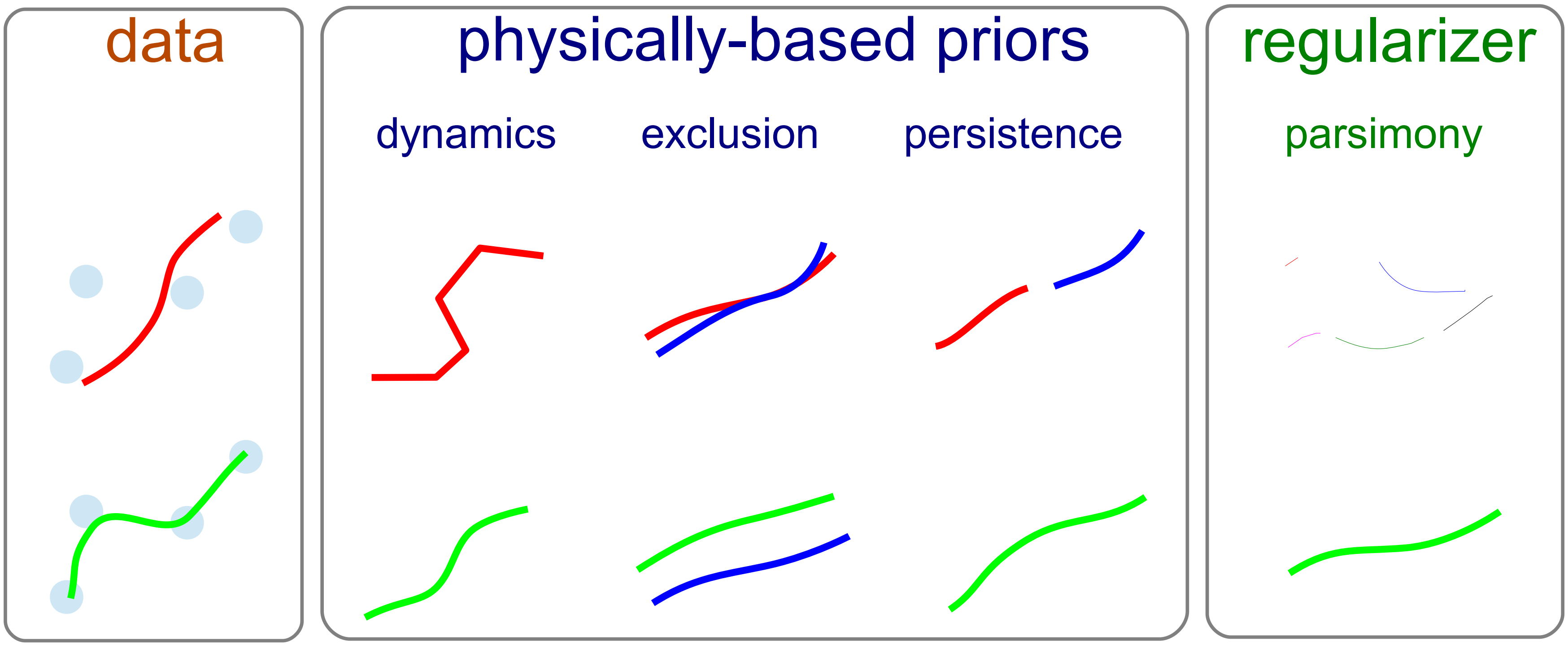
[Milan et al., PAMI 2014]

Why Continuous Energy?

- Continuous trajectories
  - low sensor resolution not an issue
- No implicit data association
  - multiple assignments possible
- Provides best results
  - Measured by standard tracking metrics

The Energy

$$E = E_{\text{obs}} + aE_{\text{dyn}} + bE_{\text{exc}} + cE_{\text{per}} + dE_{\text{reg}}$$



$$-\sum_g \|\boldsymbol{X}_i - \boldsymbol{D}_g\|^{-2}$$

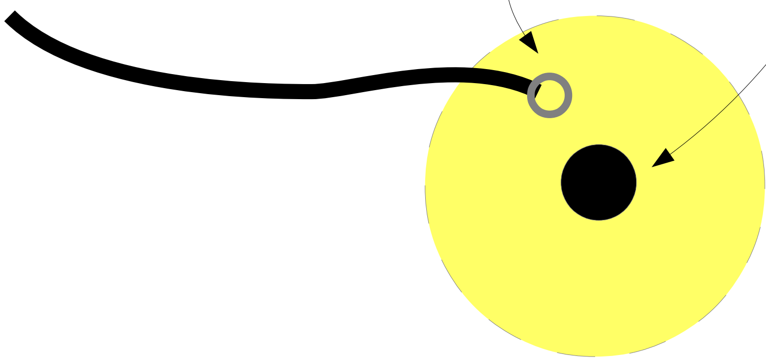
$$\sum_i \|\mathbf{v}_i^t - \mathbf{v}_i^{t+1}\|^2 \qquad \sum_{i \neq j} \|\boldsymbol{X}_i - \boldsymbol{X}_j\|^{-2} \qquad \sum_i \left(1 + \exp\left(1 - b\left(\boldsymbol{X}_i\right)\right)\right)^{-1}$$

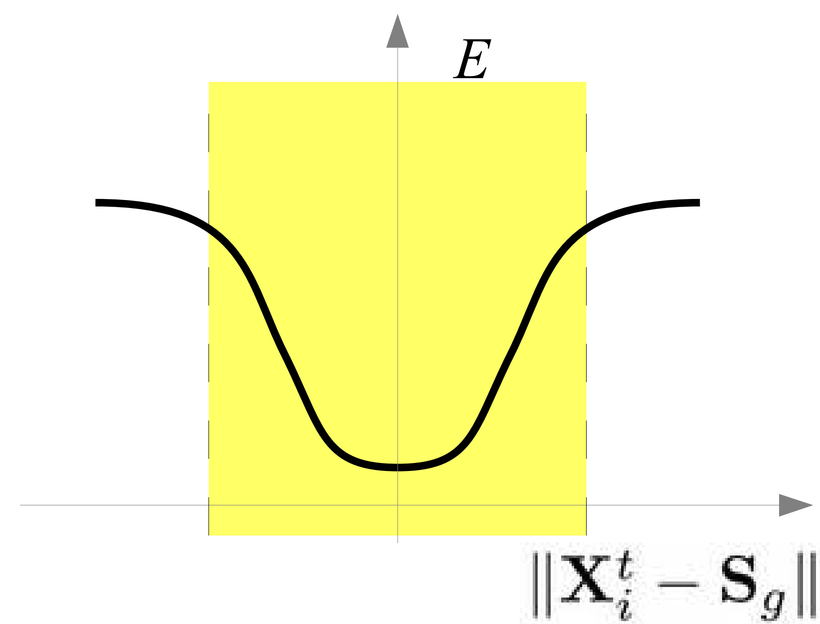
$$N + \sum_i 1/length_i$$

Data Term

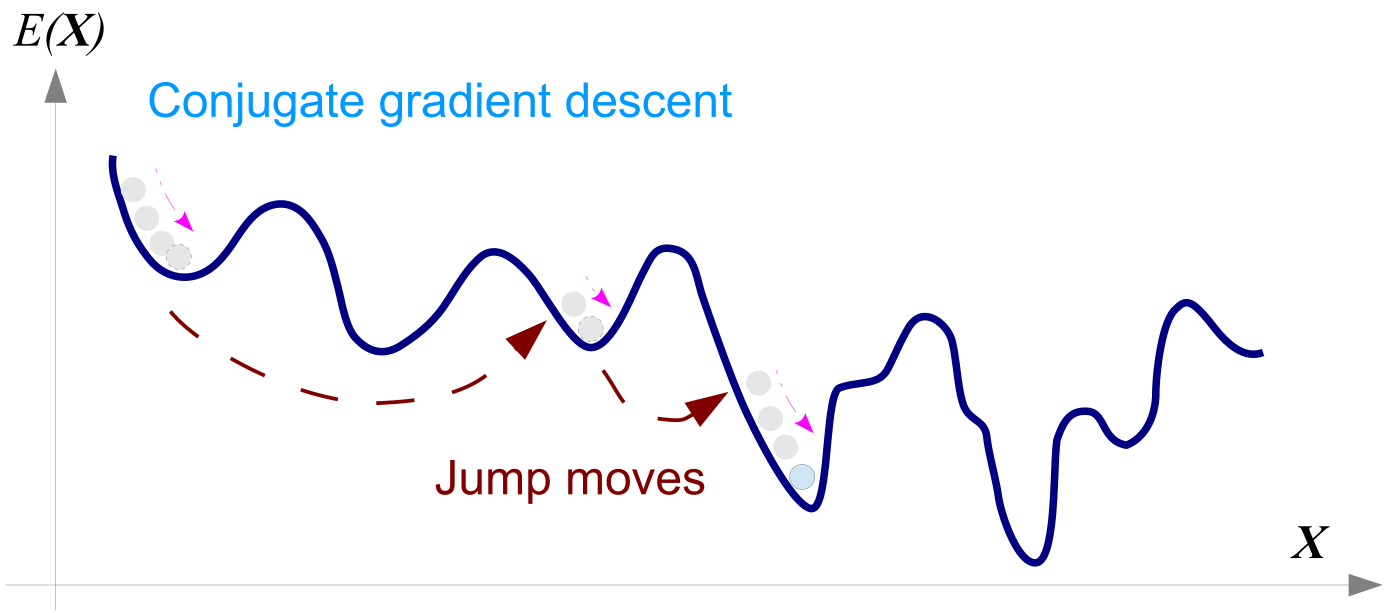
$$E_{\text{det}}(\mathbf{X}) = \sum_{i=1}^{N} \sum_{t=s_i}^{e_i} \left[\lambda - \sum_{g \in G(t)} \frac{s^2}{\|\mathbf{X}_i^t - \mathbf{S}_g\|^2 + s^2} \right]$$

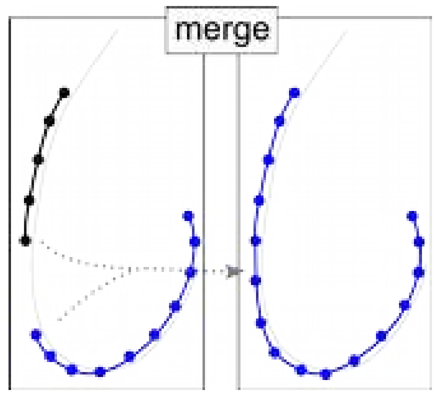
lobe size





Optimization





Merge – Split
Grow – Shrink
Add – Remove

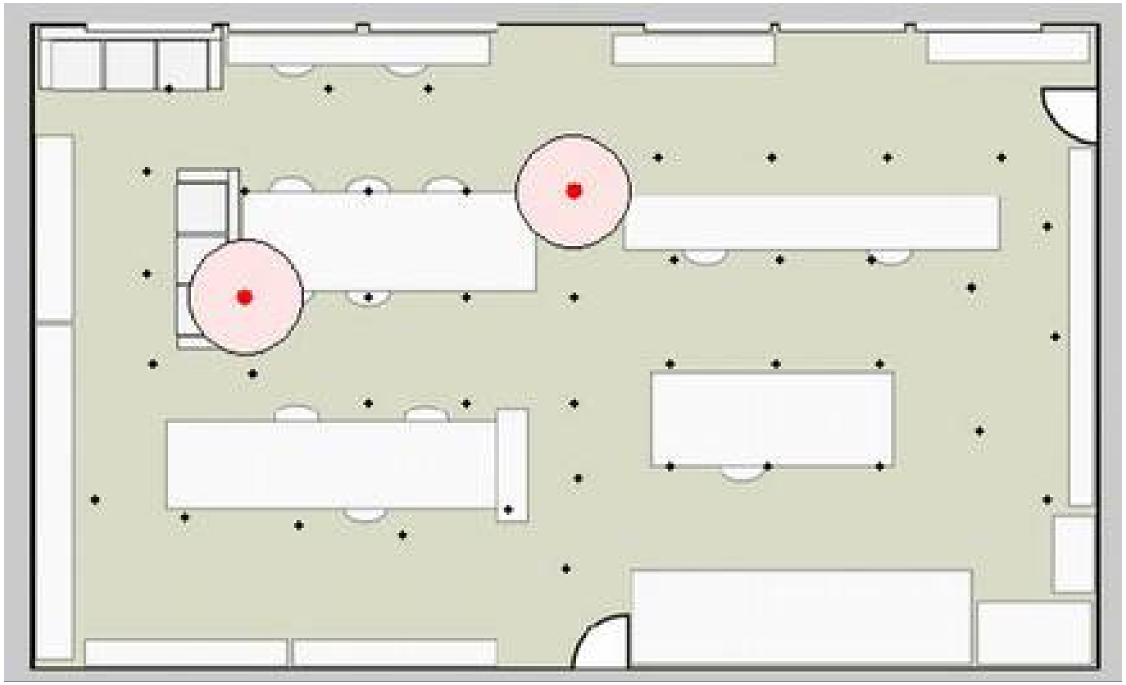
- conjugate gradient descent for **local optimization**
- discontinuous jumps to determine **dimensionality (number of targets)**

Experiments

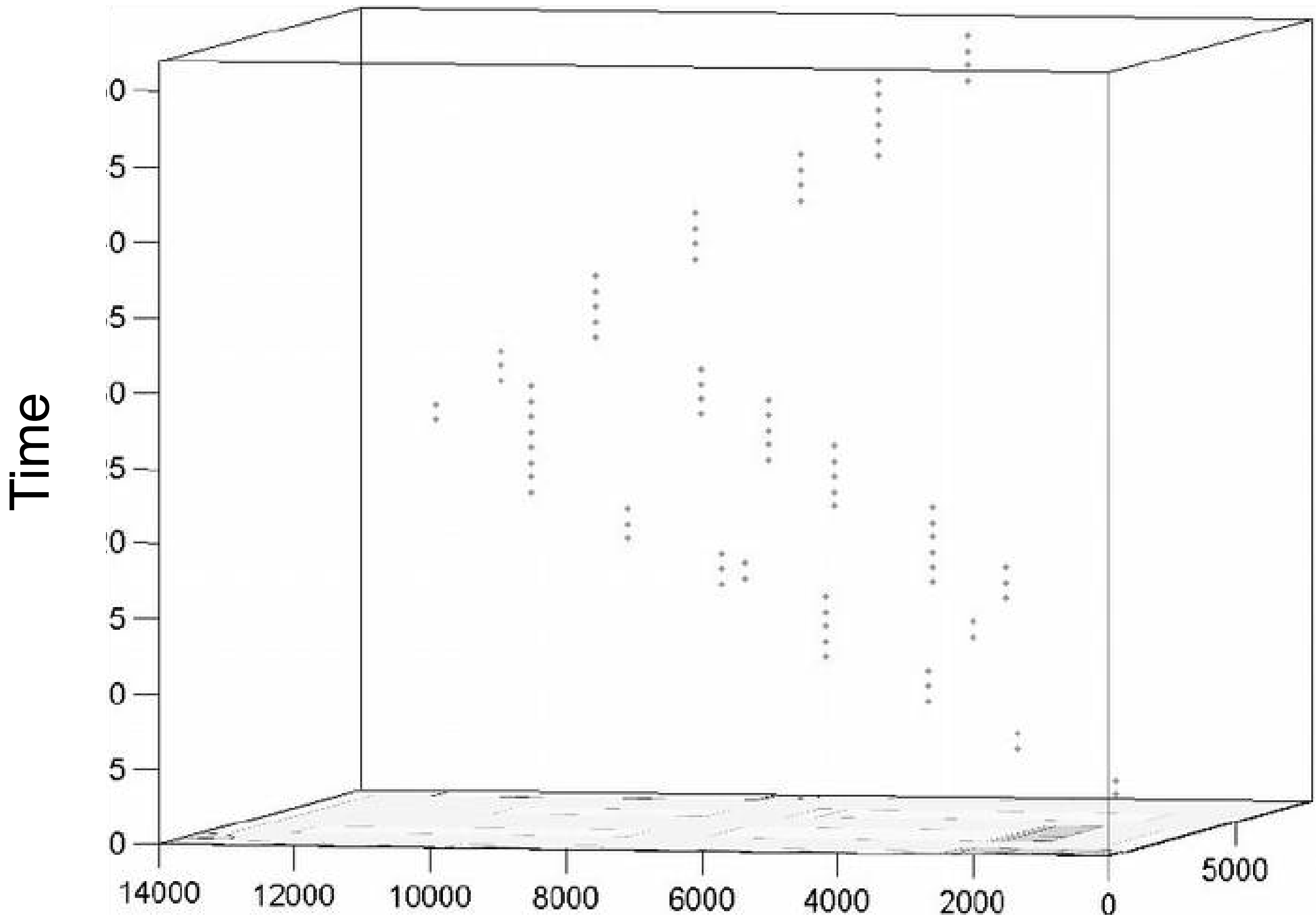
Synthetic Data

- Manual assignment of keyframes
- Interpolation and sensor simulation

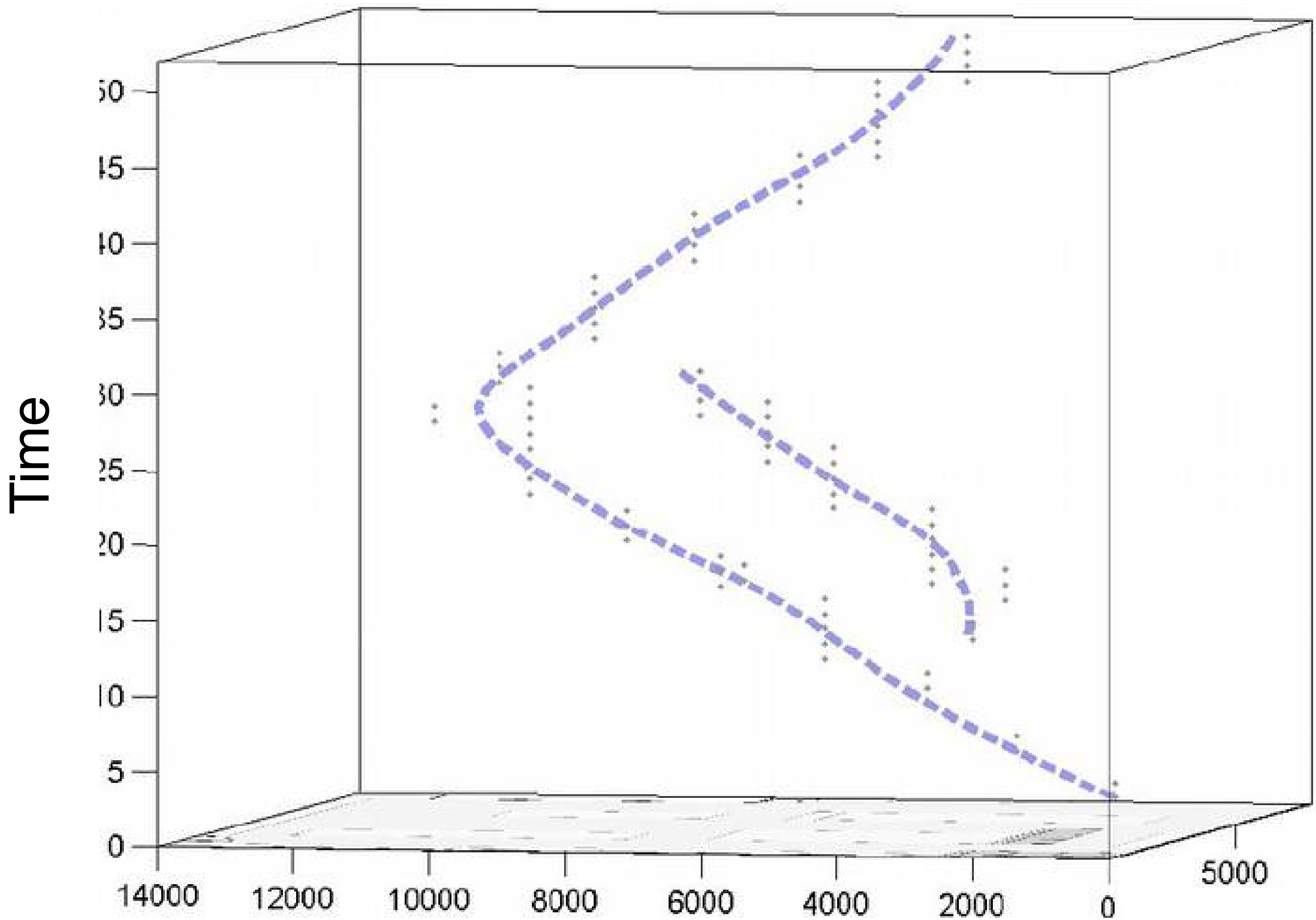
Measurements



Measurements

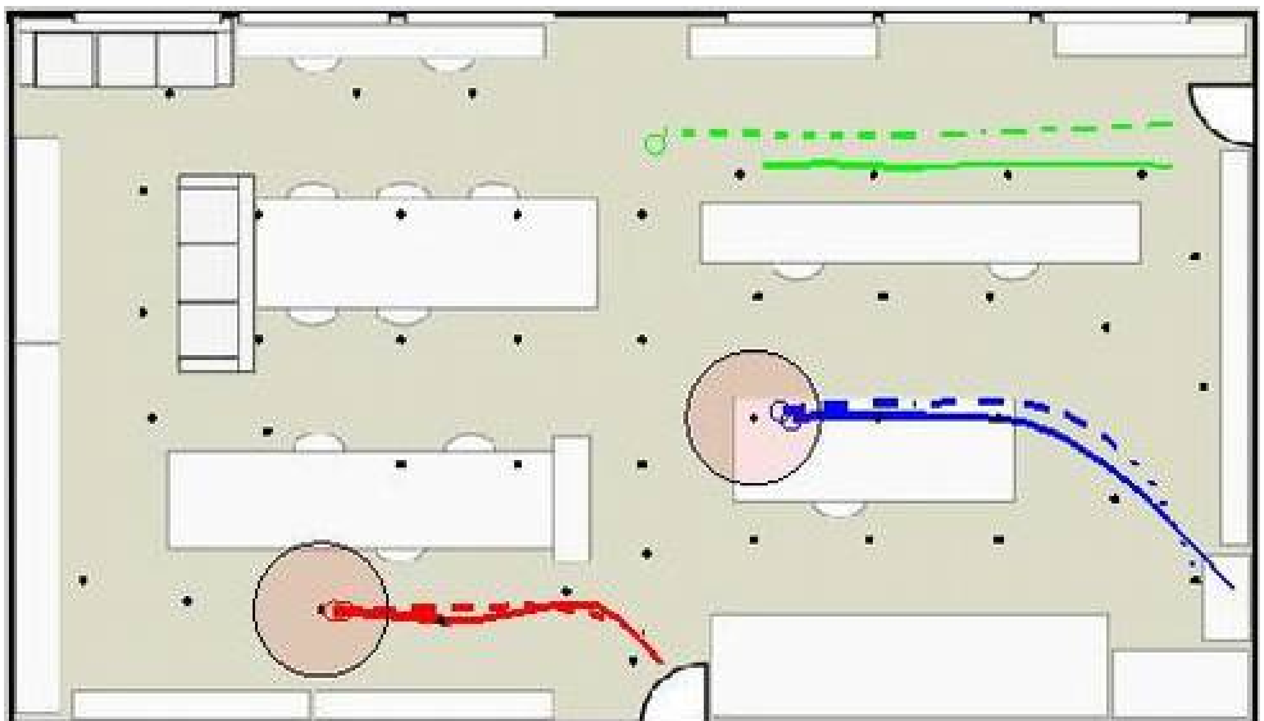


Ground Truth



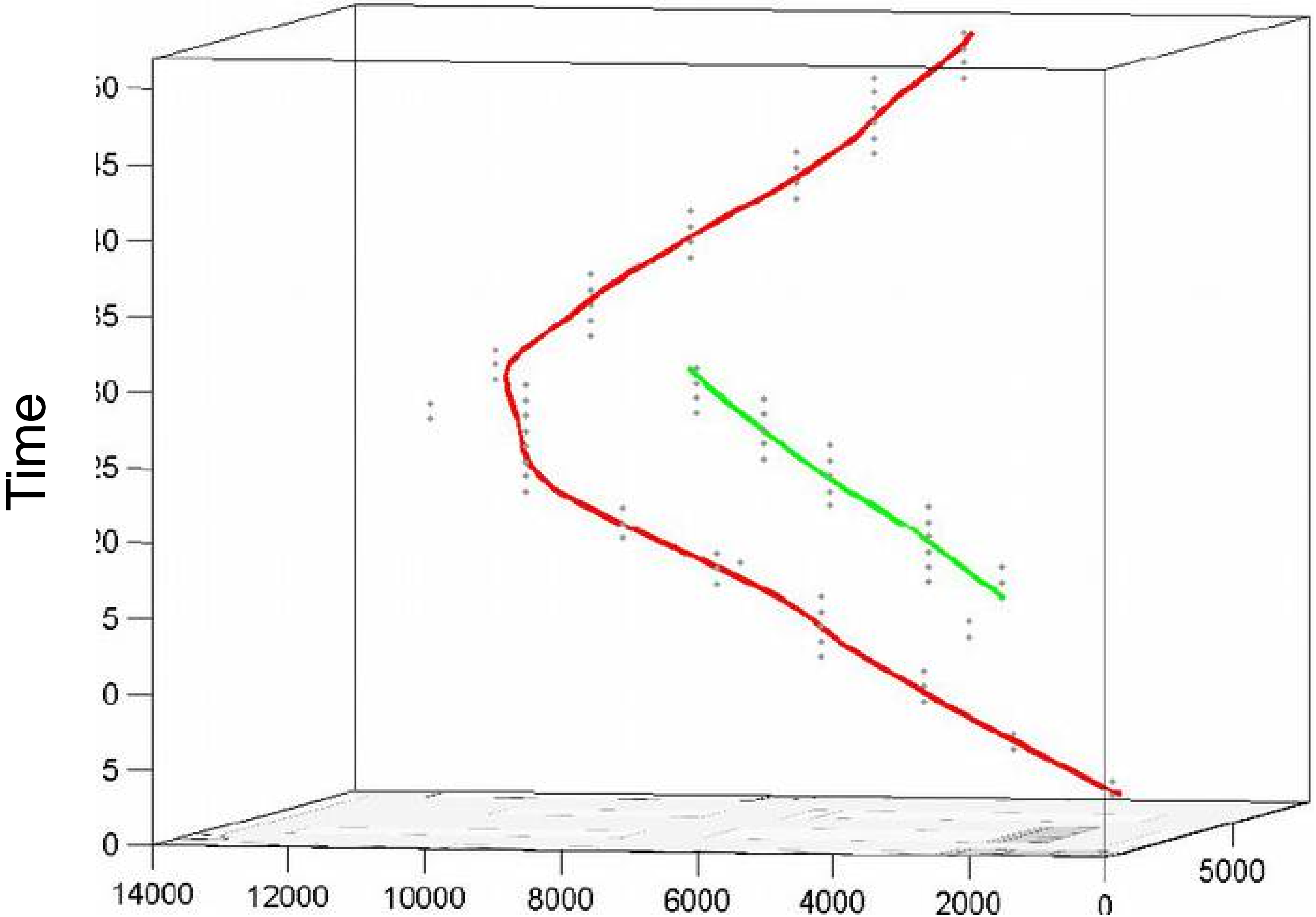
Experiments
Synthetic Data

- Manual assignment of keyframes
- Interpolation and sensor simulation

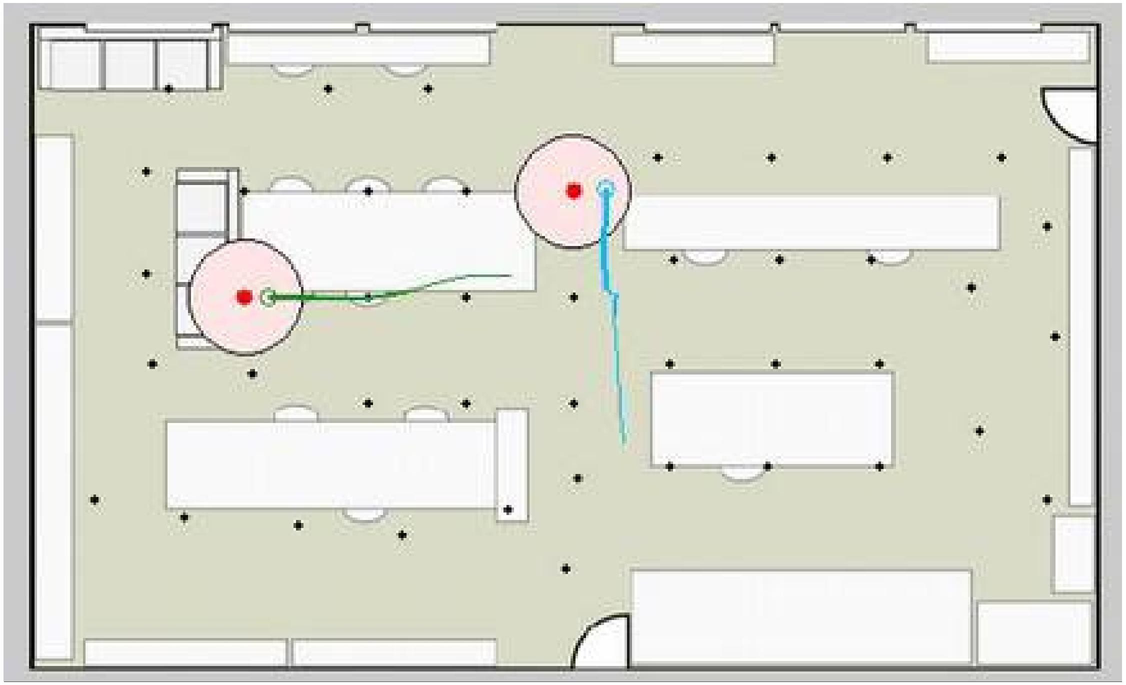


Result **GT**

Result

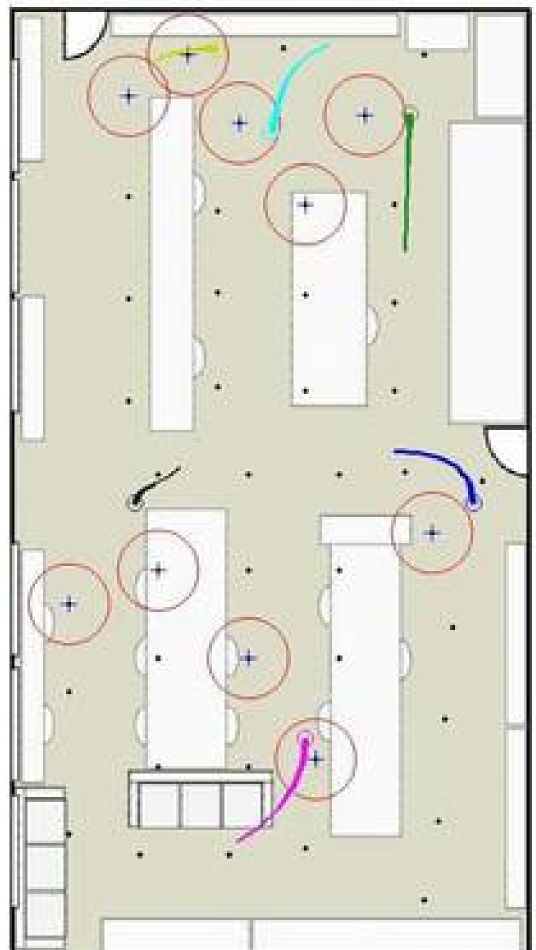


Result



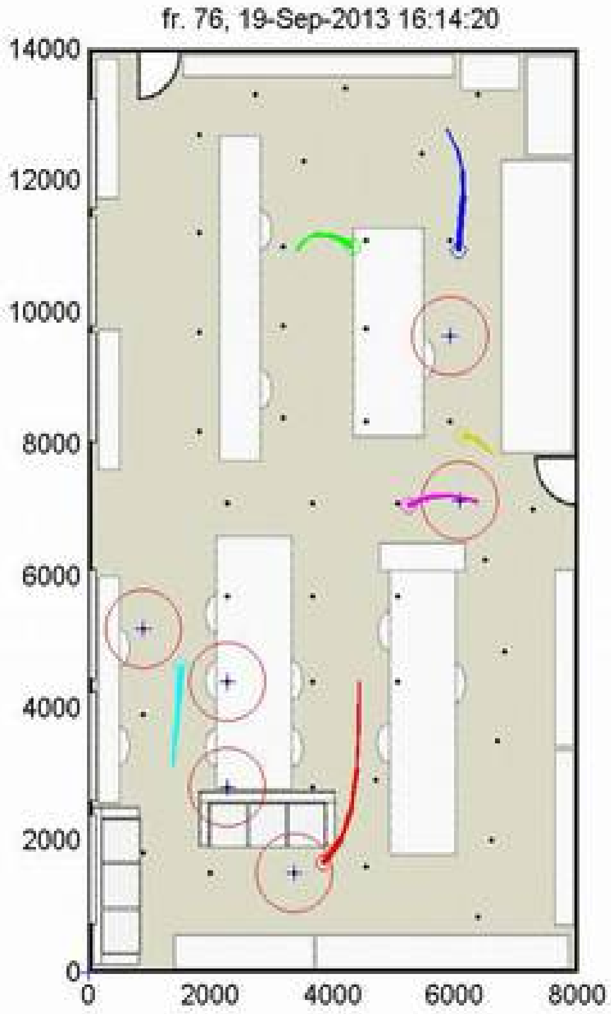
Real Data

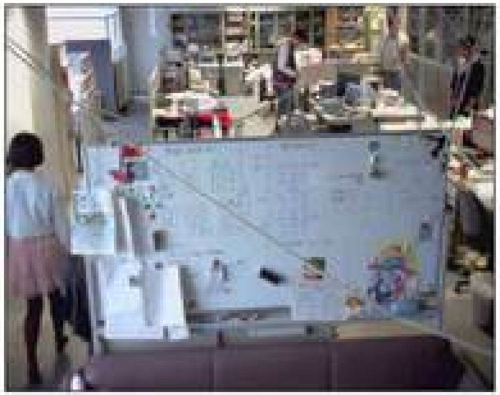
- Up to six people in a large lab
- Two cameras (2 Hz)
- Temporal alignment
- Annotation of key frames (very approximate)



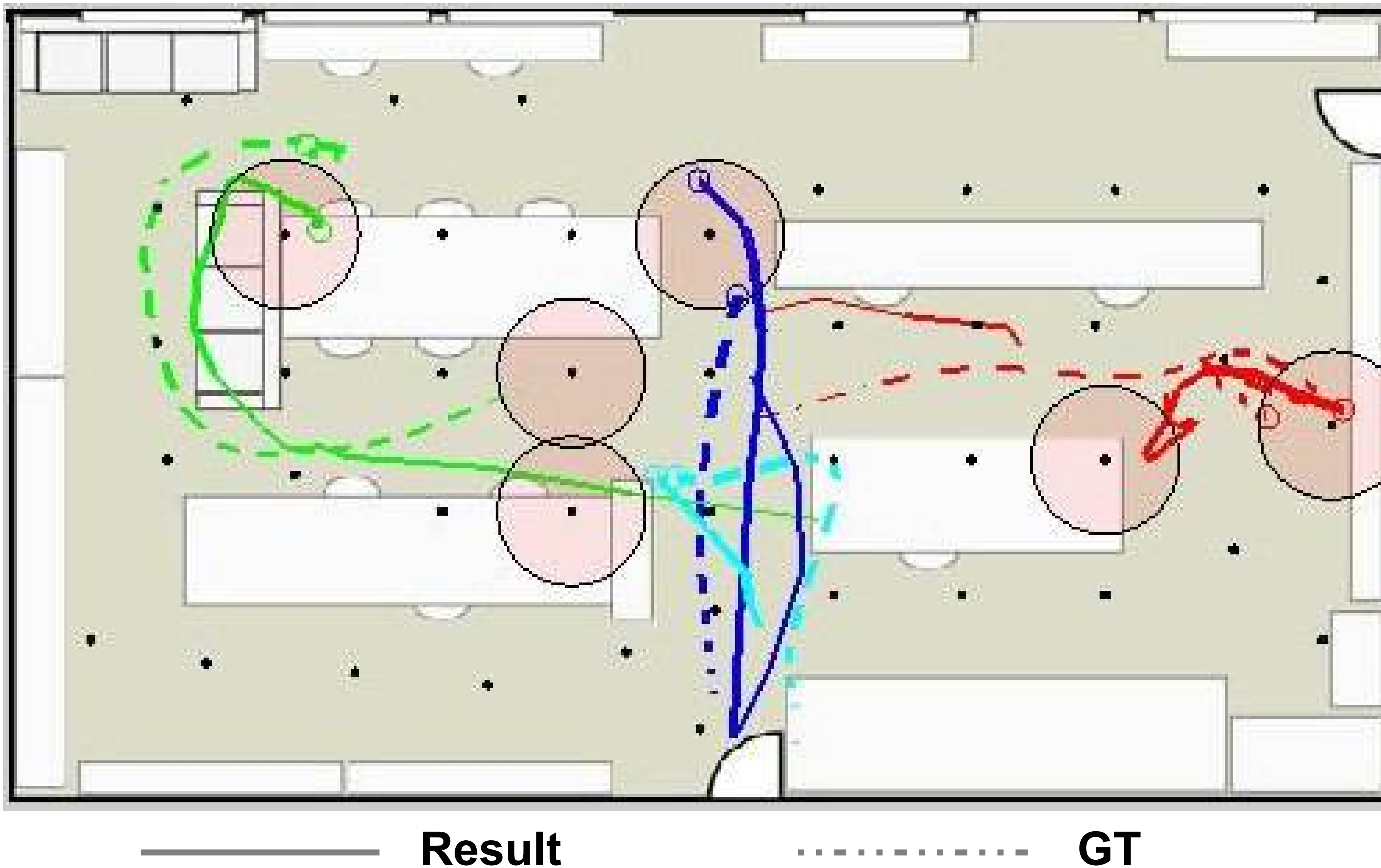
Real Data

- Up to six people in a large lab
- Two cameras (2 Hz)
- Temporal alignment
- Annotation of key frames (very approximate)

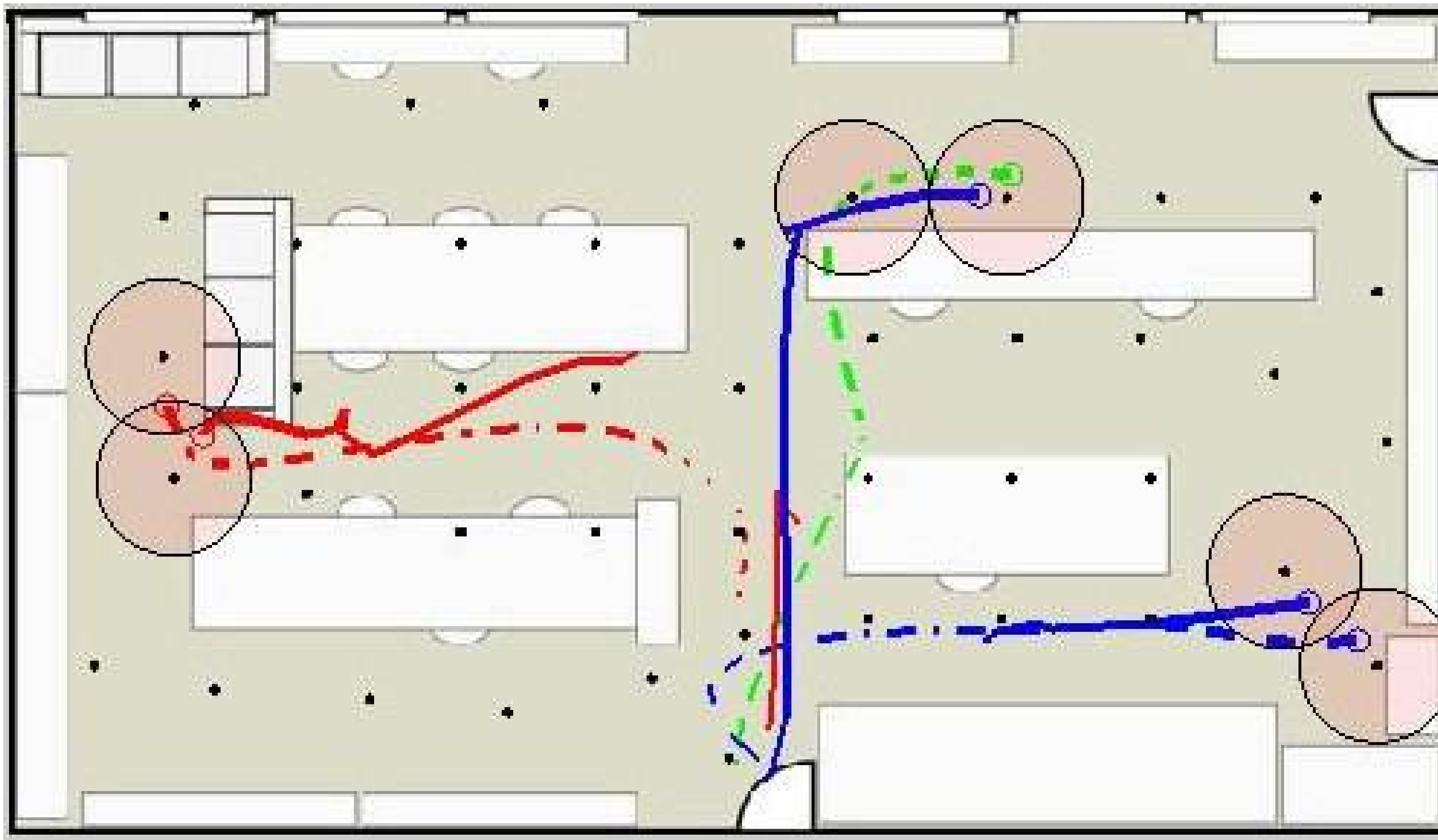




Results (real)

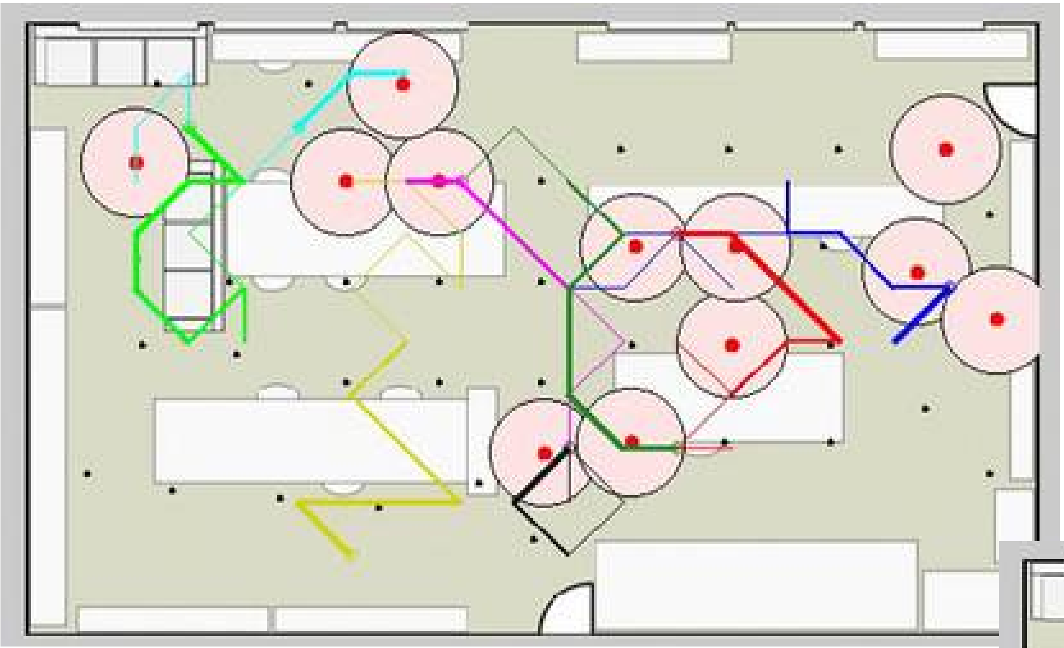


Results (real)



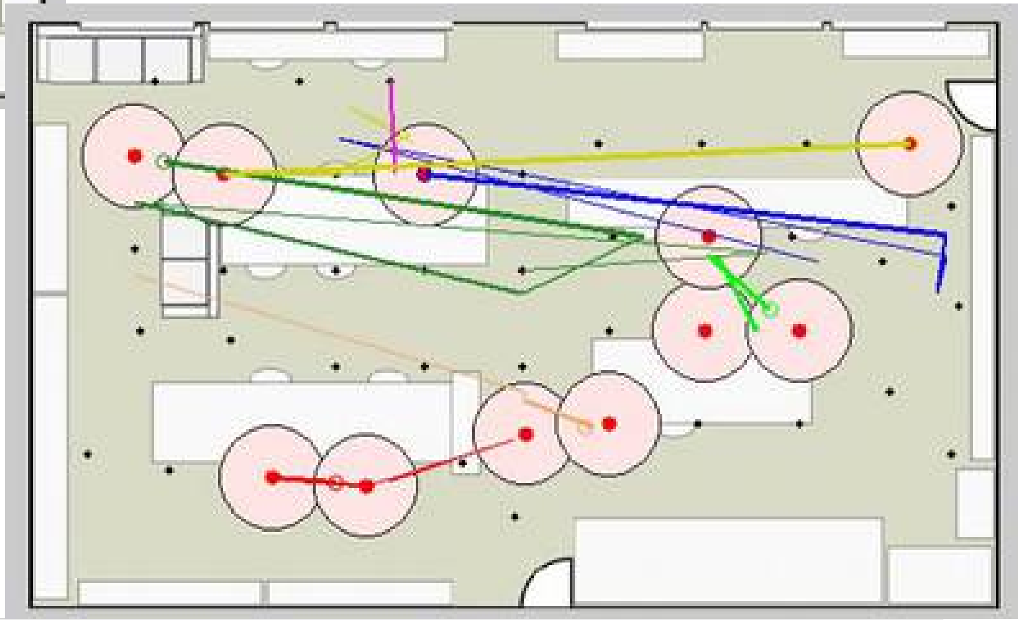
Result GT

Other Approaches



[Berclaz et al., PAMI 2011]

[Tao et al., Sensors 2012]



Quantitative Results

MOTA = normalized error count

Dataset	Method	MOTA [%]	MOTP [%]	ID sw	#Targets (MAE)
synthetic	**Ours**	**76.0**	**73.6**	13	**0.54**
Real data	**Ours**	**55.3**	**54.6**	**43**	**0.76**

MOTP = localization error (73% ≈ 35 cm)

Quantitative Results

Dataset	Method	MOTA [%]	MOTP [%]	ID sw	#Targets (MAE)
synthetic	**Ours**	**76.0**	**73.6**	13	**0.54**
	Linear DA [1]	66.6	64.6	58	0.57
	DP [2]	55.9	65.3	57	0.62
	KSP [3]	75.5	67.5	**6**	1.52
Real data	**Ours**	**55.3**	**54.6**	**43**	**0.76**
	Linear DA [1]	9.3	50.1	252	1.00
	DP [2]	9.6	47.3	128	1.25
	KSP [3]	31.1	48.3	48	1.52

[1] Tao et al., *Sensors 2012*
[2] Pirsiavsah et al., *CVPR 2011*
[3] Berclaz et al., *PAMI 2014*

Advertisement

Multiple Object Tracking Benchmark

home | data | results | submit | people | login | sign up

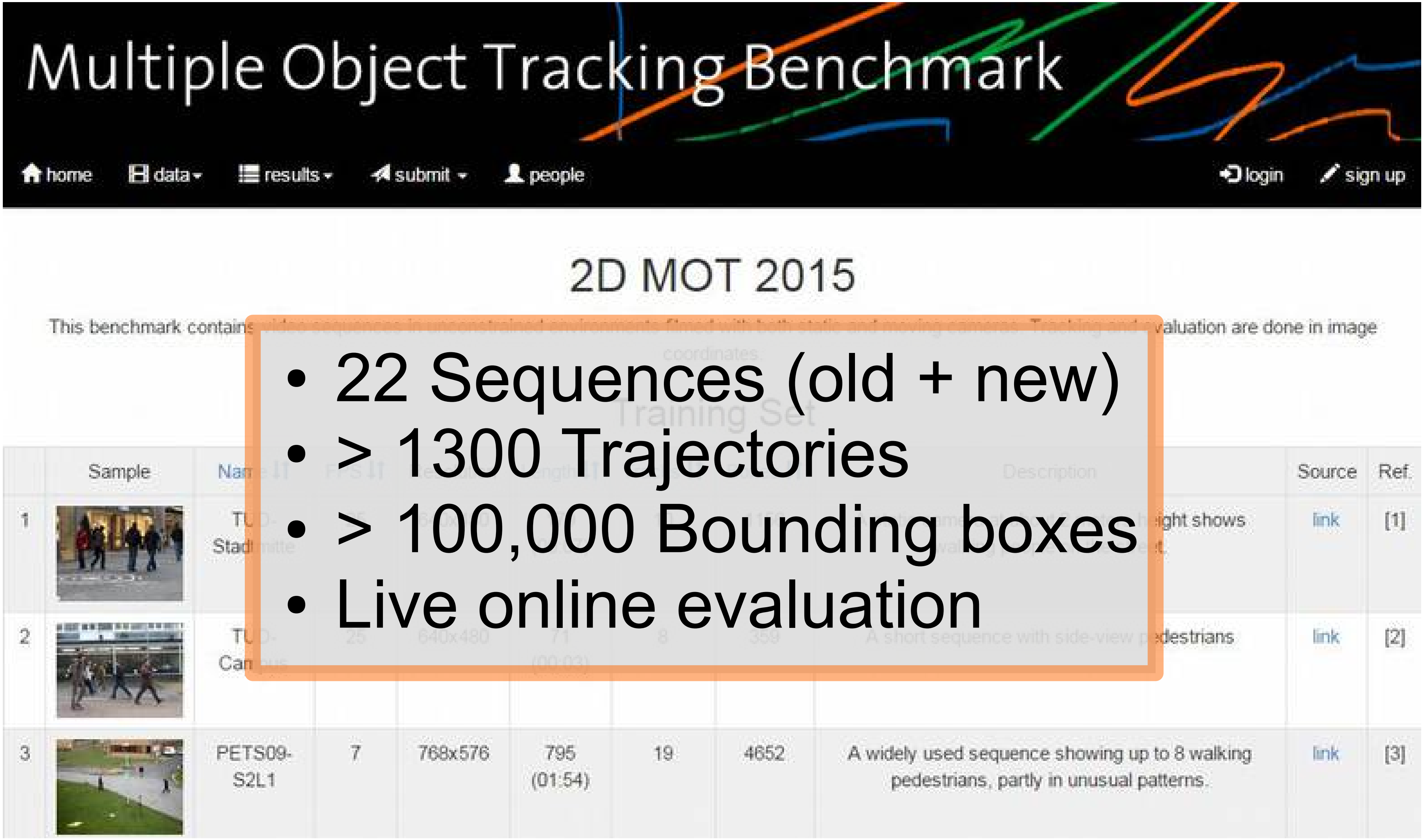
2D MOT 2015

This benchmark contains video sequences in unconstrained environments filmed with both static and moving cameras. Tracking and evaluation are done in image coordinates.

Training Set

	Sample	Name	FPS	Resolution	Length	Tracks	Boxes	Description	Source	Ref.
1		TUD-Stadtmitte	25	640x480	179 (00:07)	10	1156	A static camera at about 2 meters height shows walking people on the street.	link	[1]
2		TUD-Campus	25	640x480	71 (00:03)	8	359	A short sequence with side-view pedestrians	link	[2]
3		PETS09-S2L1	7	768x576	795 (01:54)	19	4652	A widely used sequence showing up to 8 walking pedestrians, partly in unusual patterns.	link	[3]

Advertisement



- 22 Sequences (old + new)
- > 1300 Trajectories
- > 100,000 Bounding boxes
- Live online evaluation

Advertisement

Multiple Object Tracking Benchmark

home | data | results | submit | people | login | sign up

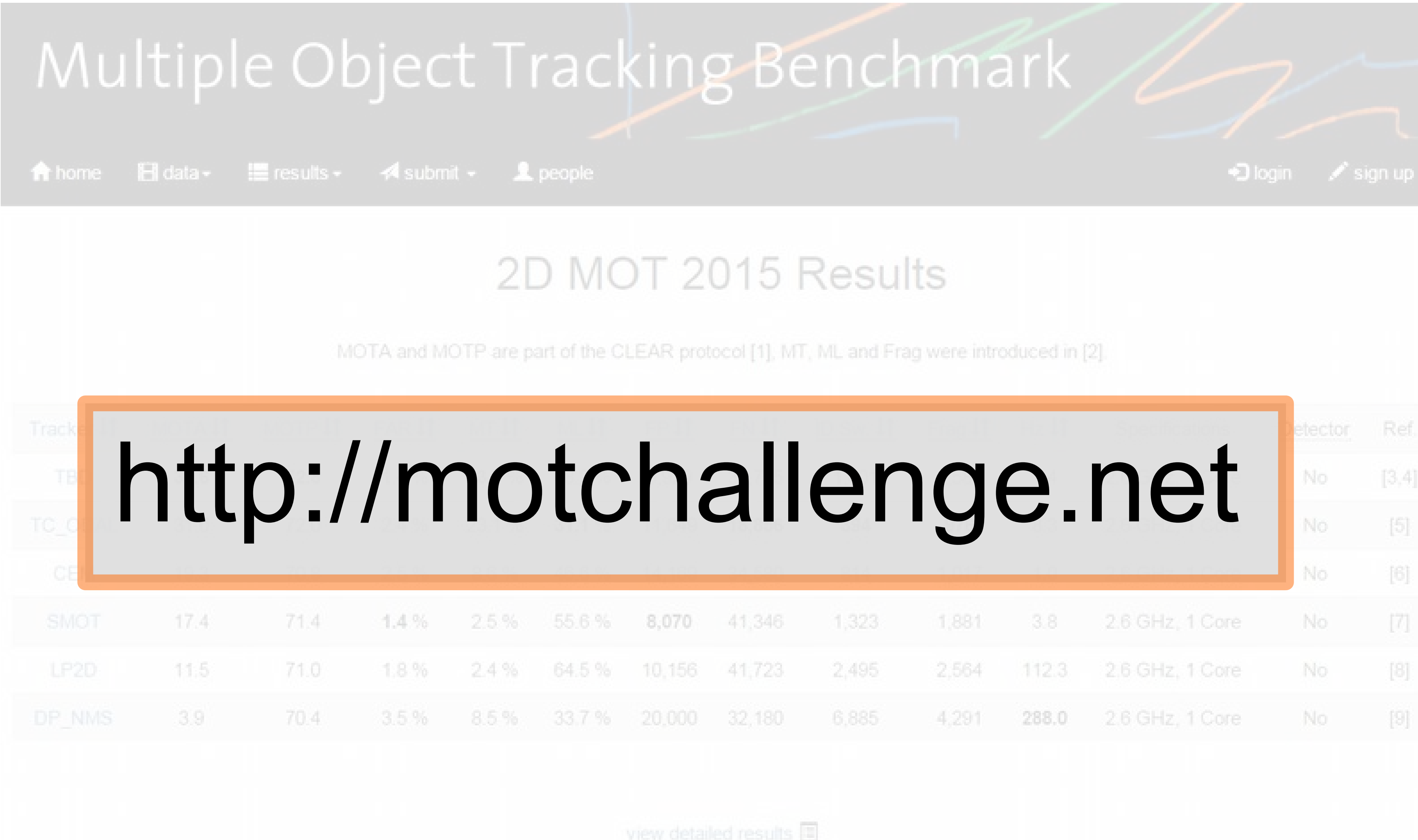
2D MOT 2015 Results

MOTA and MOTP are part of the CLEAR protocol [1], MT, ML and Frag were introduced in [2].

Tracker	MOTA	MOTP	FAR	MT	ML	FP	FN	ID Sw.	Frag	Hz	Specifications	Detector	Ref.
TBD	**31.6**	**72.8**	1.4 %	**13.7** %	48.2 %	9,929	30,735	1,801	1,583	1.4	2.6 GHz, 1 Core	No	[3,4]
TC_ODAL	31.0	72.0	2.4 %	23.1 %	**31.1** %	11,649	**18,856**	**394**	**976**	8.3	2.6 GHz, 1 Core	No	[5]
CEM	19.3	70.8	2.5 %	8.6 %	46.6 %	14,169	34,580	814	1,017	1.0	2.6 GHz, 1 Core	No	[6]
SMOT	17.4	71.4	**1.4** %	2.5 %	55.6 %	**8,070**	41,346	1,323	1,881	3.8	2.6 GHz, 1 Core	No	[7]
LP2D	11.5	71.0	1.8 %	2.4 %	64.5 %	10,156	41,723	2,495	2,564	112.3	2.6 GHz, 1 Core	No	[8]
DP_NMS	3.9	70.4	3.5 %	8.5 %	33.7 %	20,000	32,180	6,885	4,291	**288.0**	2.6 GHz, 1 Core	No	[9]

view detailed results

Advertisement



http://motchallenge.net

Summary

- A principled alternative to preserve privacy
- Continuous energy with soft assignments
- Still a very challenging problem
- Data + Code online

http://research.milanton.net/irtracking/