Advances in video tracking:

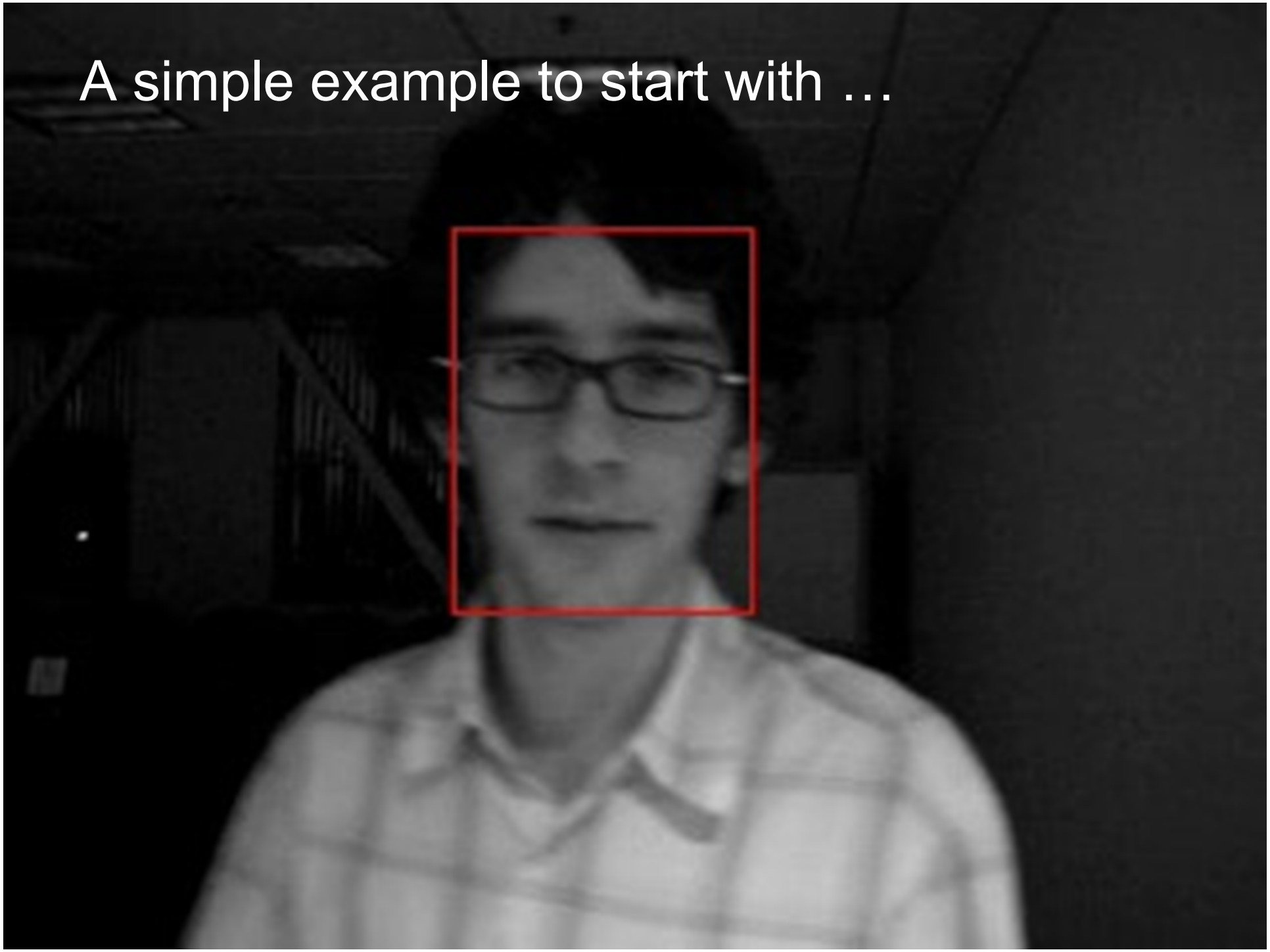
adaptation, self-evaluation & context learning

Andrea Cavallaro

CIS centre for intelligent sensing

Queen Mary University of London

A simple example to start with …



Multi-target tracking on confidence maps: an application to people tracking

F. Poiesi, R. Mazzon, A. Cavallaro

Computer Vision and Image Understanding, Vol. 117, Issue 10, pp. 1257-1272, October 2013

Multi-target tracking on confidence maps: an application to people tracking

F. Poiesi, R. Mazzon, A. Cavallaro

Computer Vision and Image Understanding, Vol. 117, Issue 10, pp. 1257-1272, October 2013

… and how about tracking these targets?

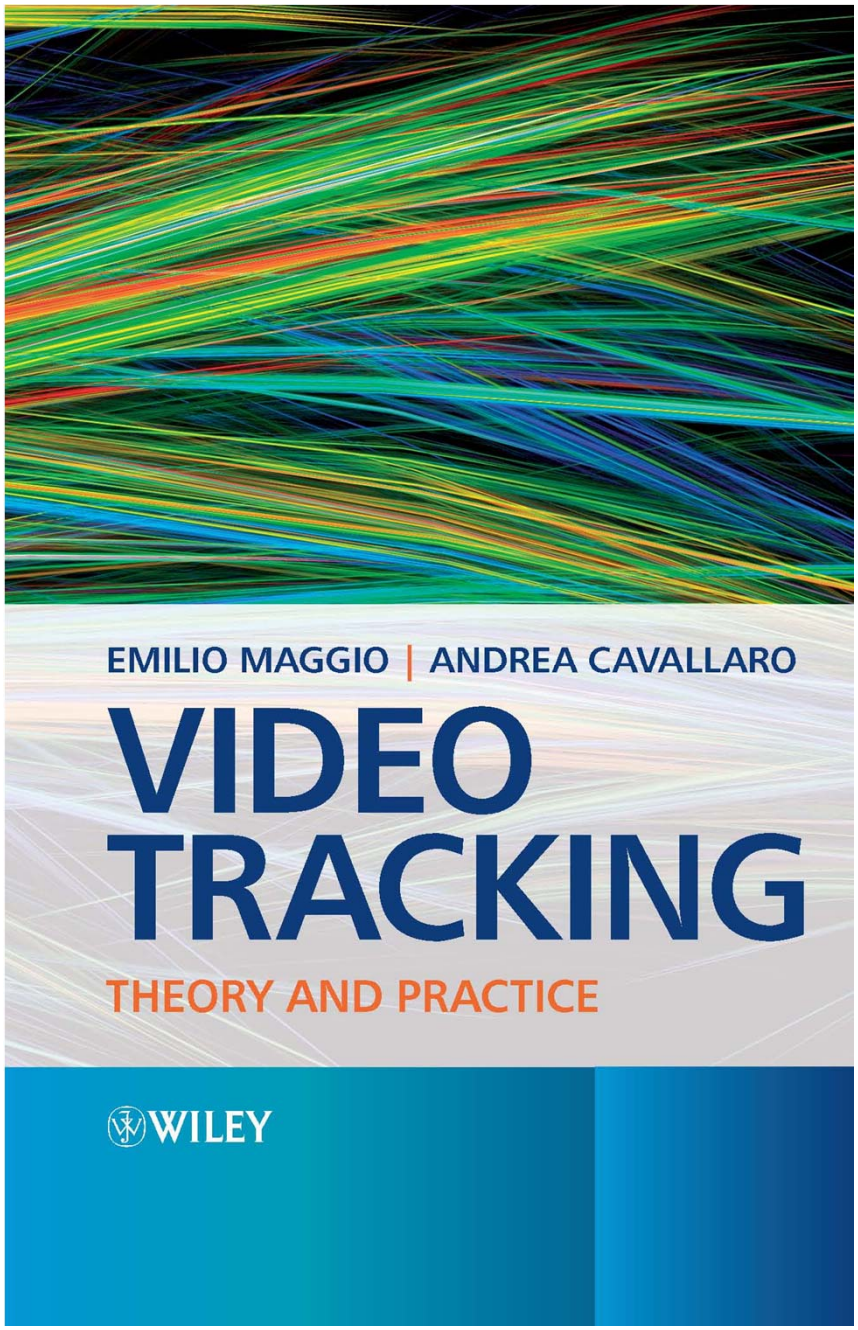
Reference

Book

Video Tracking

Theory and Practice

Wiley and Sons, 2011
ISBN: 978-0-470-74964-7



Website

www.videotracking.org

Advances in video tracking

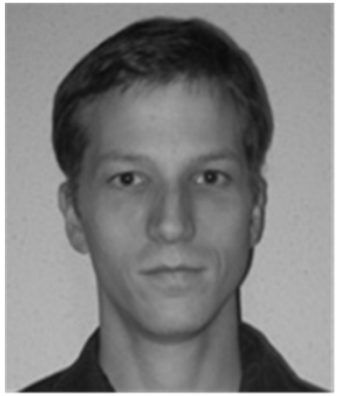
adaptation **self-evaluation** **context learning**

Acknowledgements

adaptation

Samuele Salti

self-evaluation

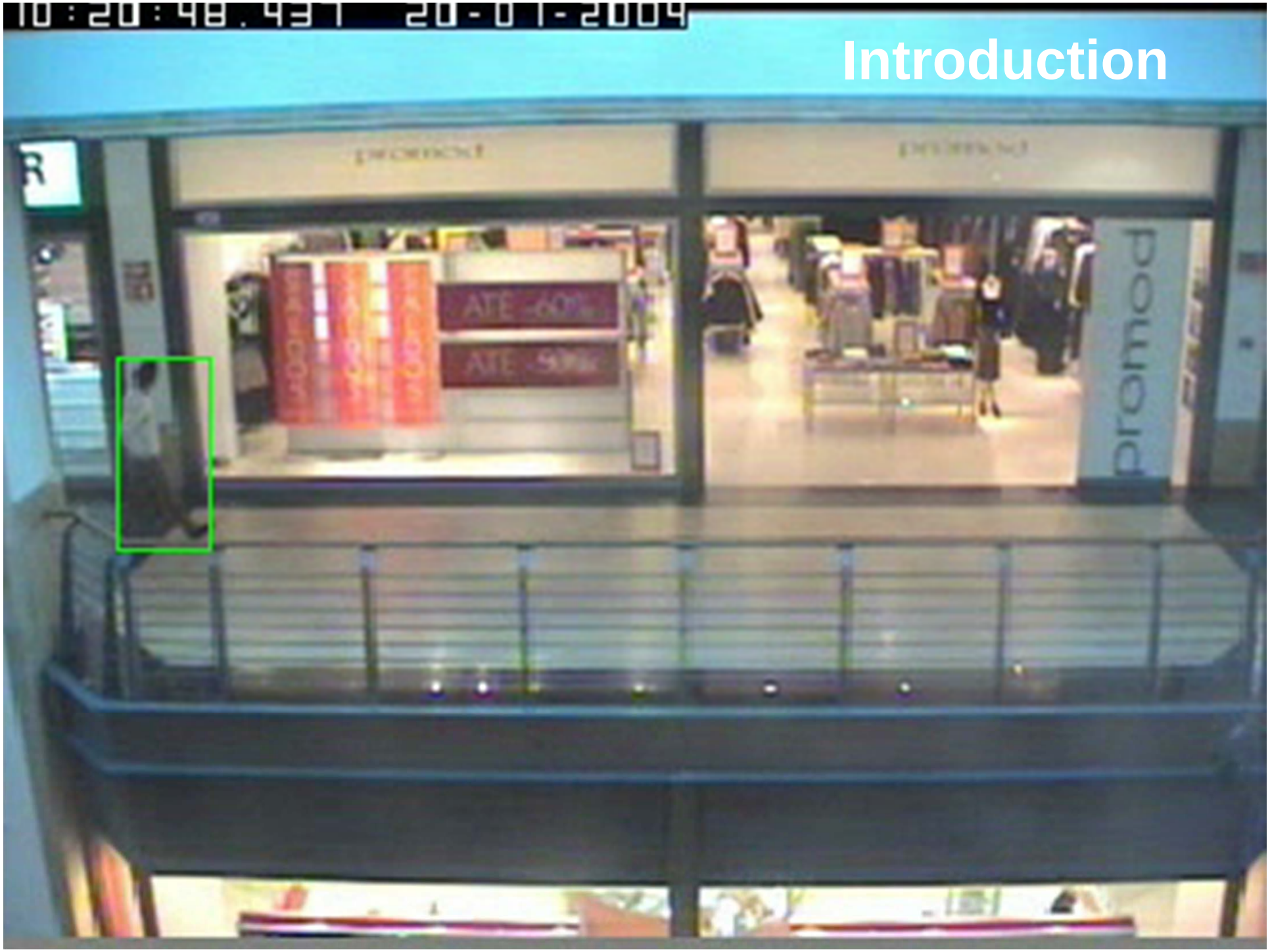


Juan Carlos SanMiguel

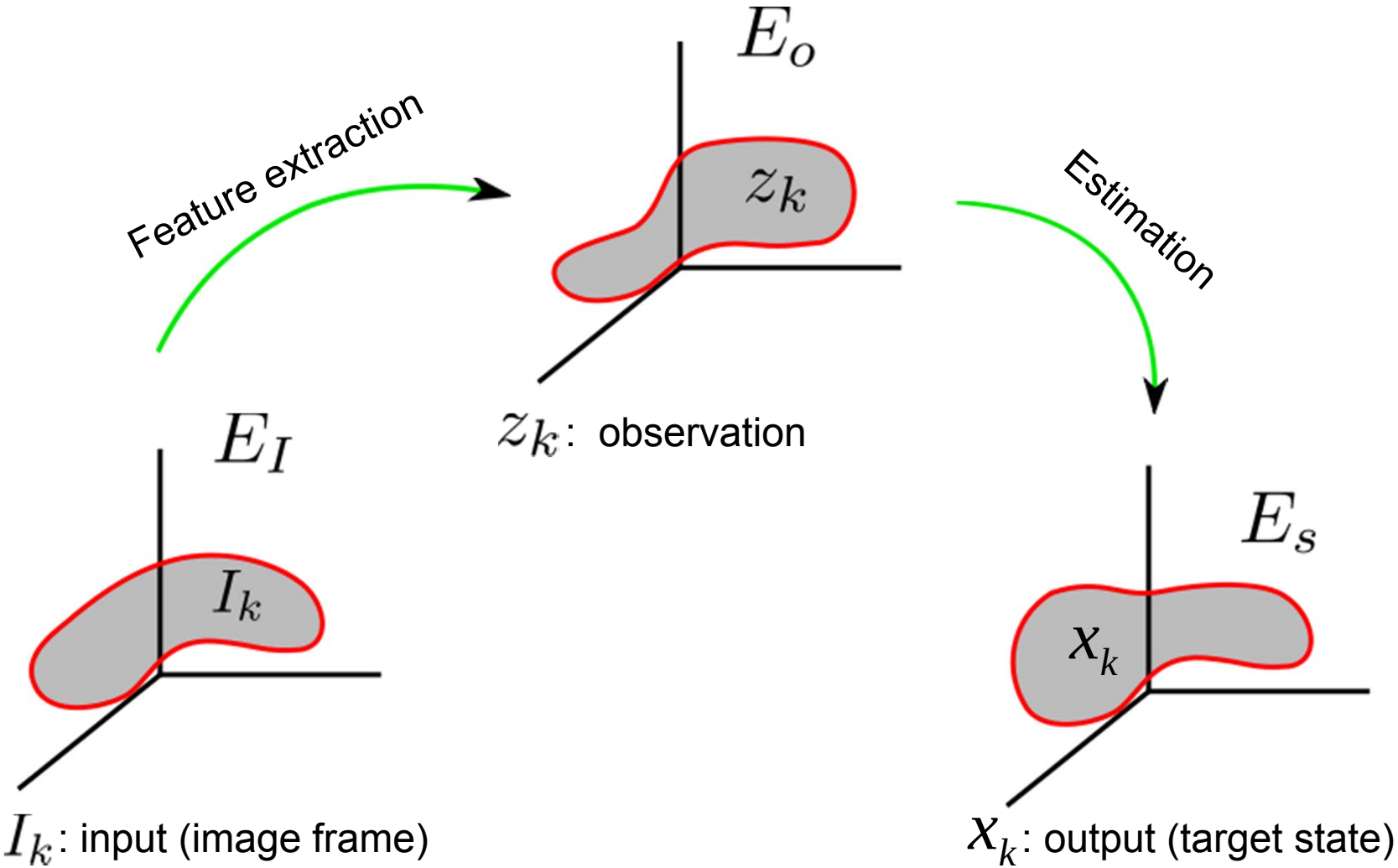
context learning

Emilio Maggio

Introduction

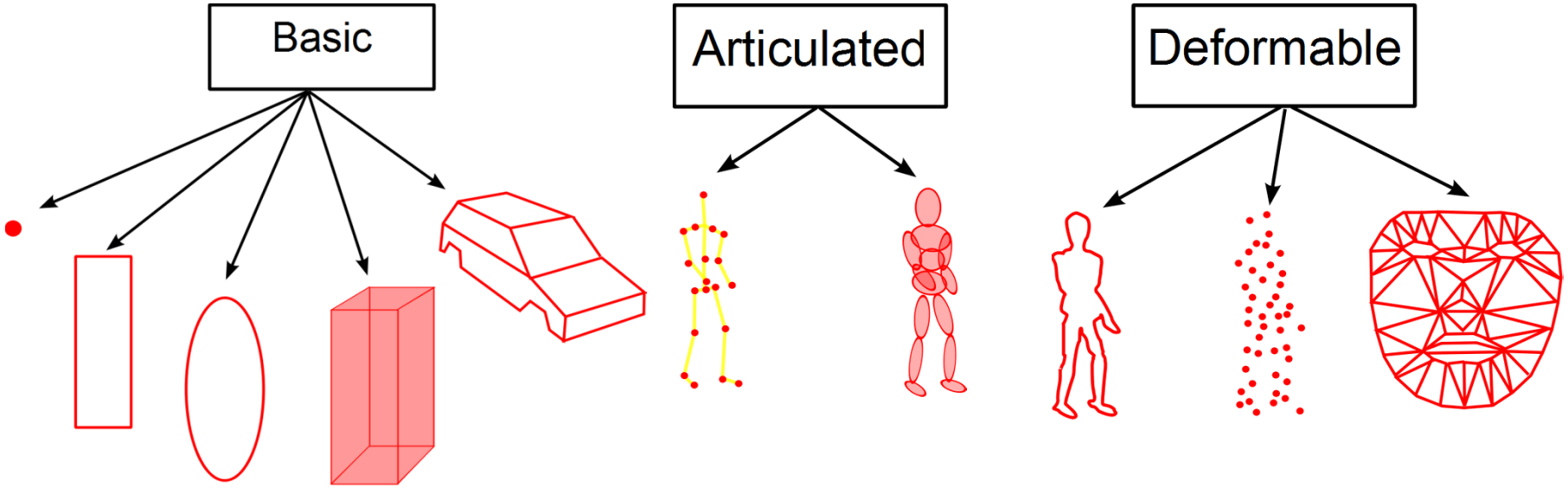


Tracking: information flow



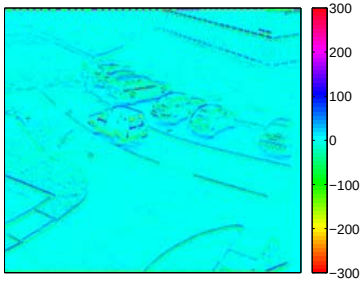
Target state

- The state x_k is a vector of object parameters
  - Shape and position
  - Appearance
  - Higher order moments (e.g. speed, acceleration)

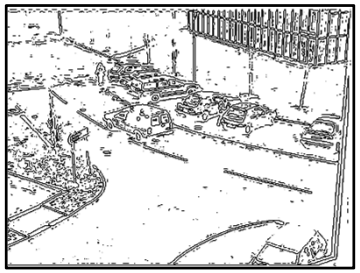


Observation

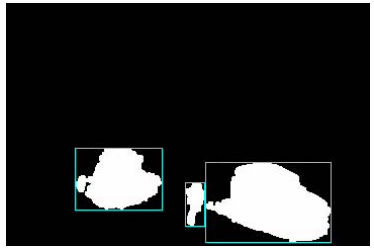
- The observation vector z_k represents object information extracted from the image
  - Low-level information
    - e.g. pixel colour, gradient



  - Mid-level features
    - e.g. location of edges



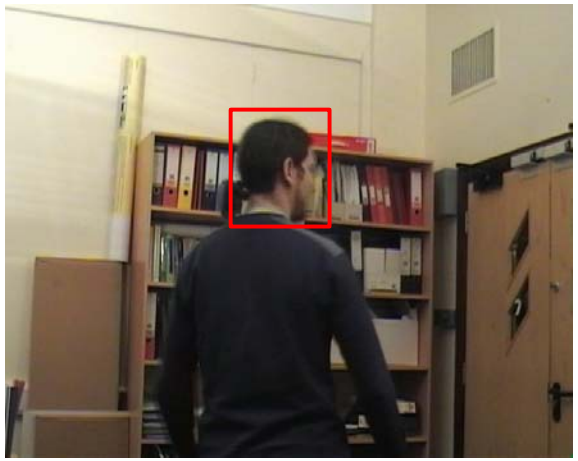
  - High-level features
    - e.g. object detection



  - Combinations of the above

Challenges: appearance variations

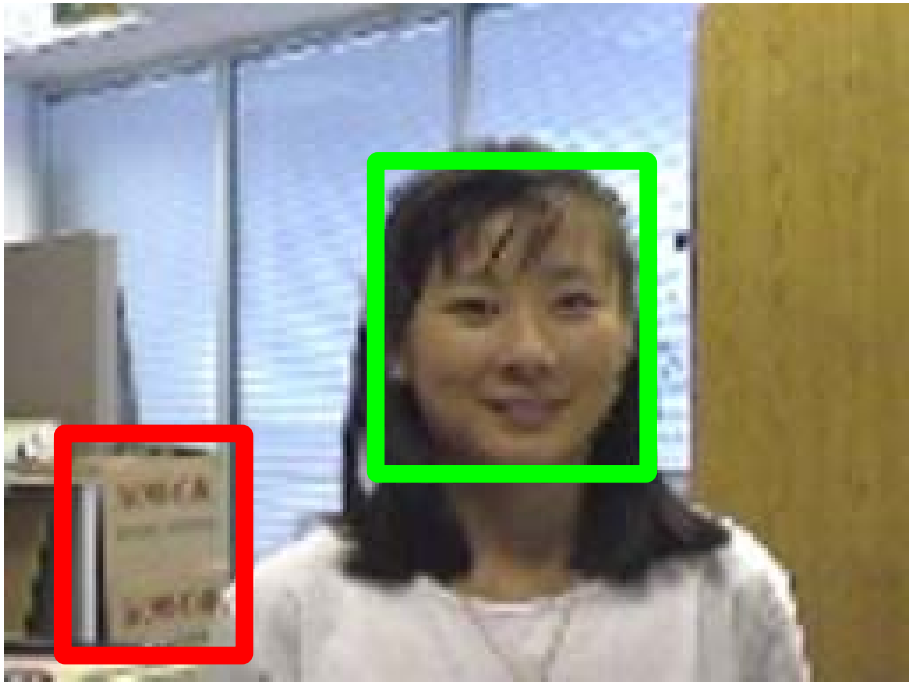
- Pose variations

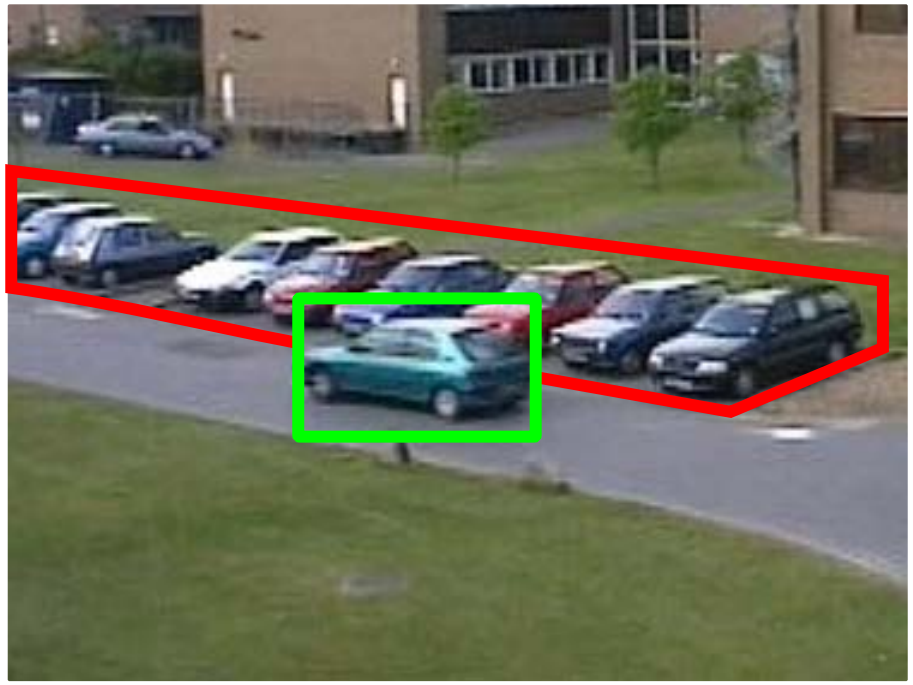


- Occlusions

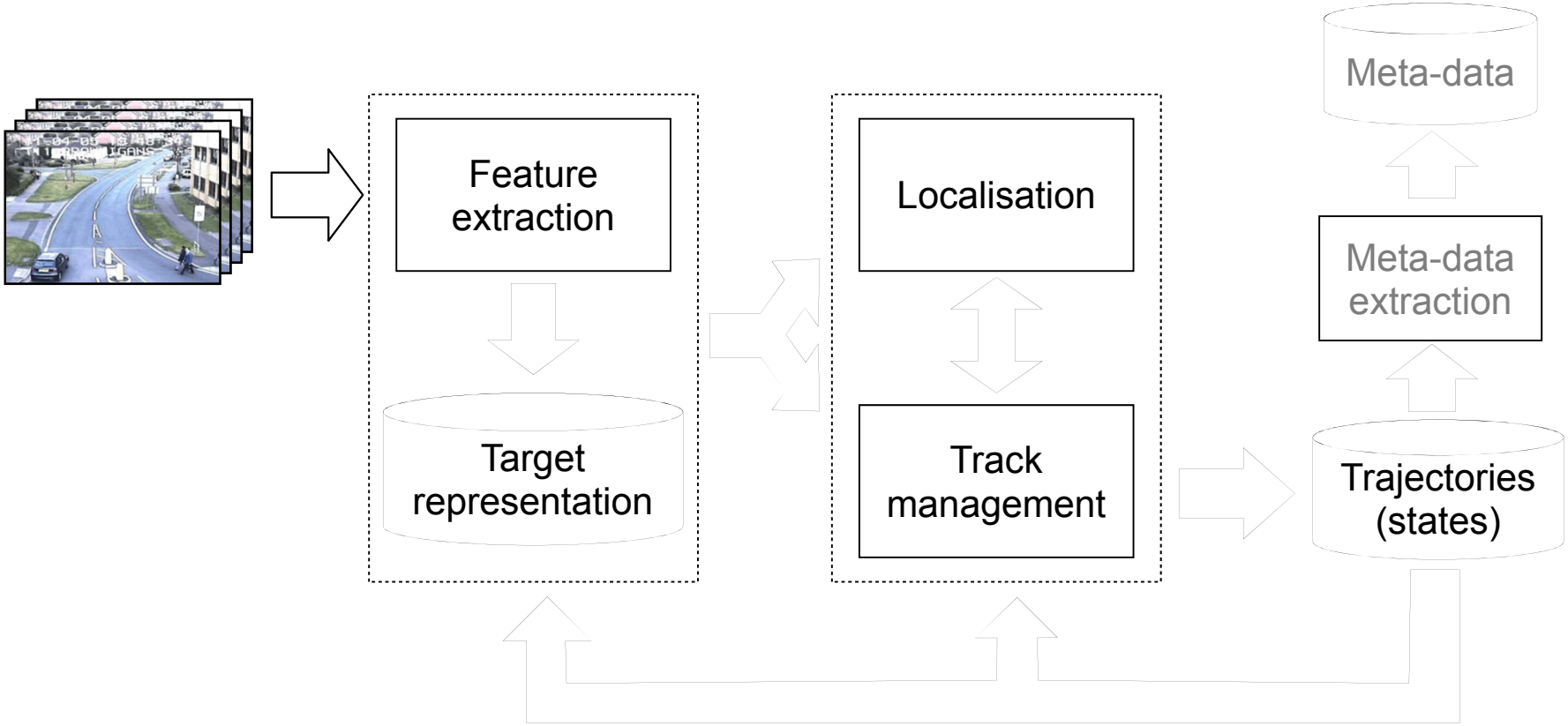
Challenges: clutter

- Background objects with similar appearance





The video tracking pipeline



Particle Filter

- State $\mathbf{x}_k = \mathbf{f}_k(\mathbf{x}_{k-1}, \mathbf{u}_k)$
- Observation $\mathbf{z}_k = \mathbf{h}_k(\mathbf{x}_k, \mathbf{n}_k)$

- Objective
  - to estimate unknown state $\mathbf{x}_k$ based on a sequence of observations $\mathbf{z}_k, k = 0,1,\ldots$
  - to approx. the posterior distribution $p(\mathbf{x}_k \mid \mathbf{z}_{1:k}) \approx \sum_{i=1}^{L} w_k^i \, \delta(\mathbf{x}_k - \mathbf{x}_k^i)$
- Solution (Bayesian)
  - Prediction step
    - Based on state equation
  - Update step
    - Based on likelihood function

Sequential Monte Carlo (Particle Filter)

- Monte Carlo approximation of the Bayes recursion
  - Approximate the posterior *pdf*

$$p(x_k \mid z_{1:k}) \approx \sum_{i=1}^{L} \omega_k^{(i)} \, \delta(x_k - x_k^{(i)})$$

  - Samples propagated using state and observation equations

$$\omega_k^{(i)} \propto \omega_{k-1}^{(i)} \frac{g(z_k \mid x_k^{(i)}) \, \cancel{f(x_k^{(i)} \mid x_{k-1}^{(i)})}}{\cancel{q(x_k^{(i)} \mid x_{k-1}^{(i)}, z_k)}}$$

- Sequential Importance Sampling (SIS)
  - Sample according to dynamics

$$q(x_k \mid x_{k-1}, z_k) = f(x_k \mid x_{k-1})$$

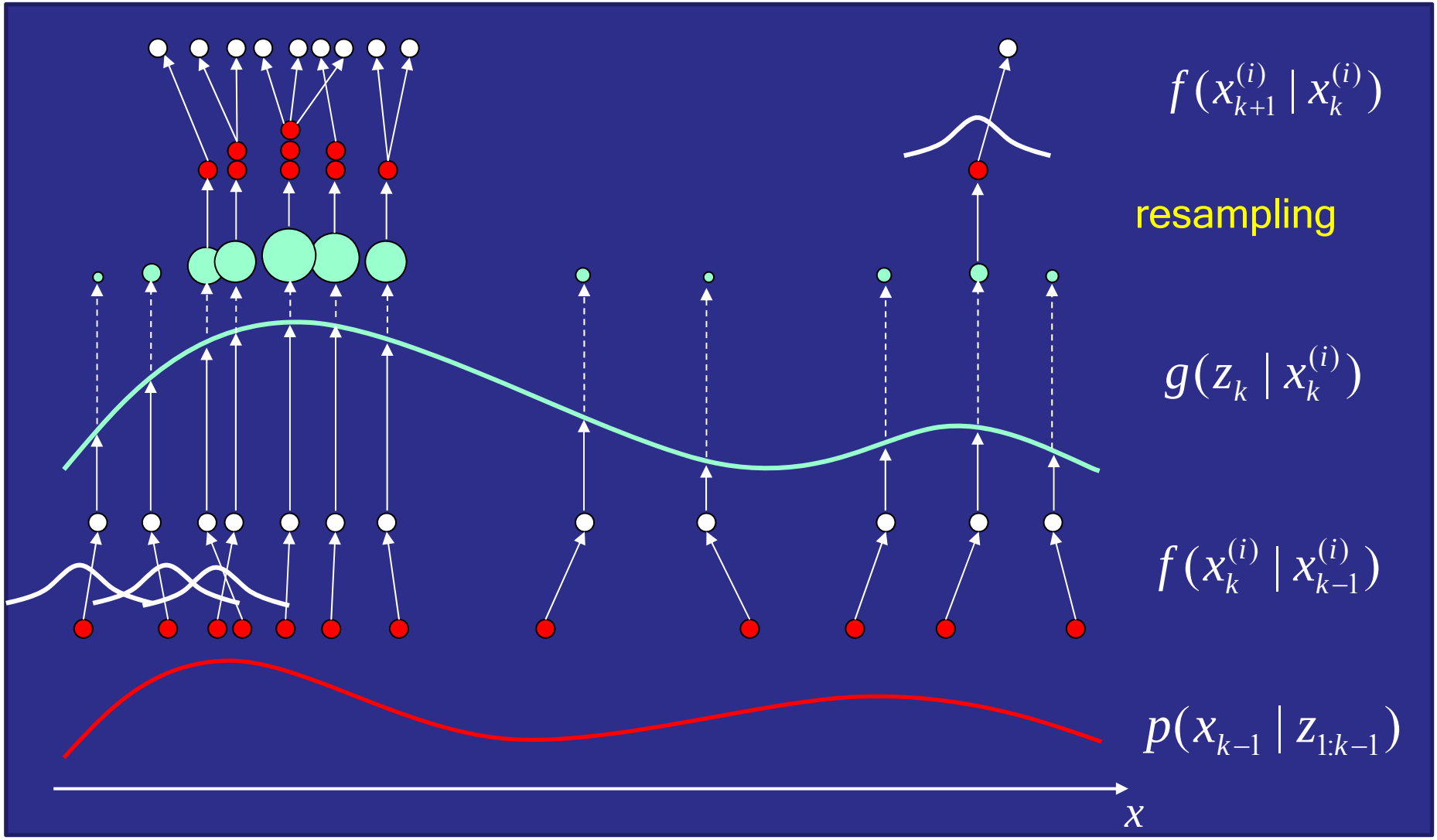
Resampling

- SIS problem
  - The sampling *does not depend on the measurement*
  - After some steps particles have *small weights*

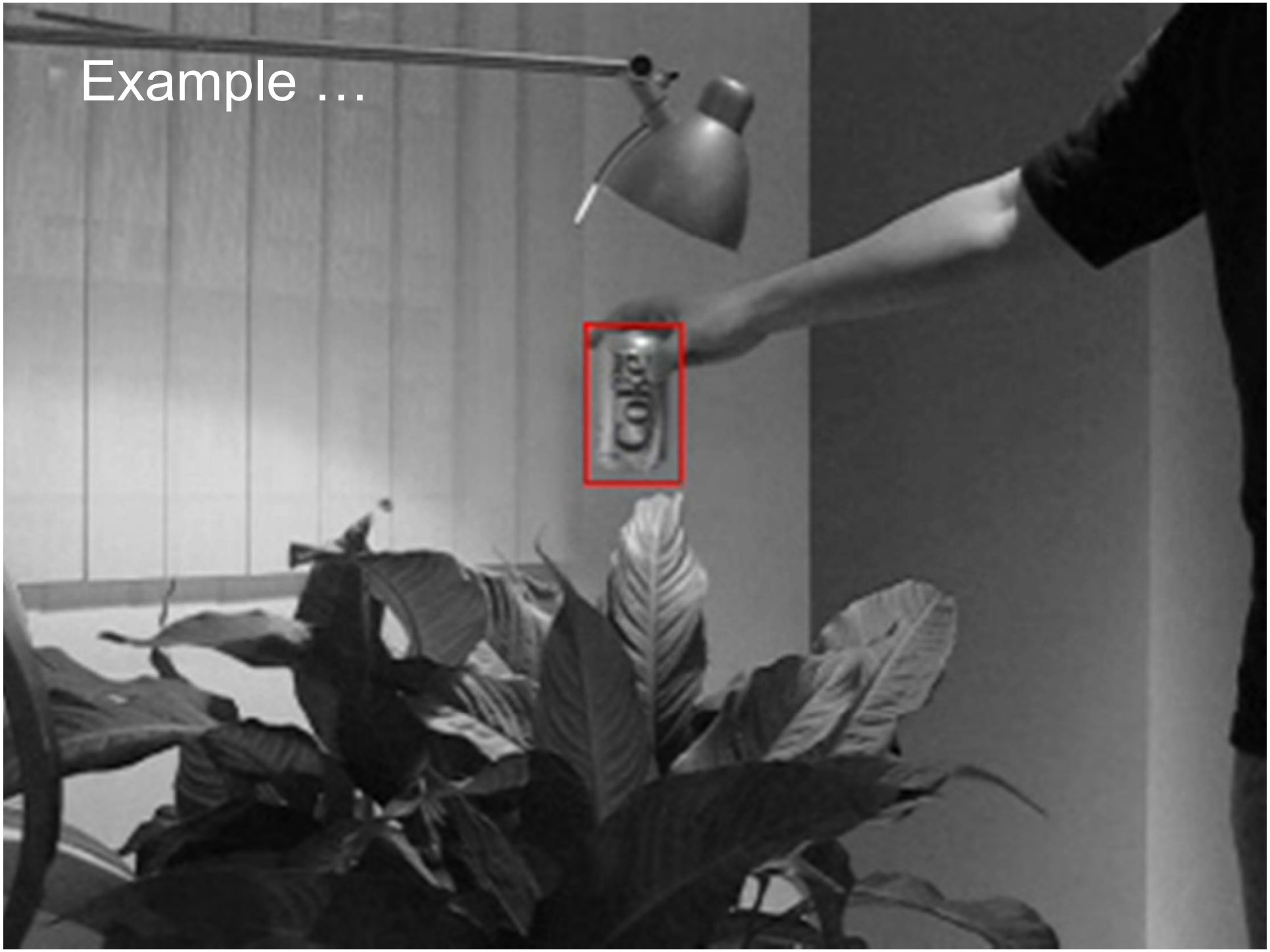
- Solution
  - To eliminate particles with **small weight**
  - To propagate multiple times particles with **large weight**

- Sequential Importance Resampling (SIR)
  - Sample proportionally to $\omega_{k-1}^{(i)}$
  - New weight $\omega_k^{(i)} \propto g(z_k \mid x_k^{(i)})$ likelihood

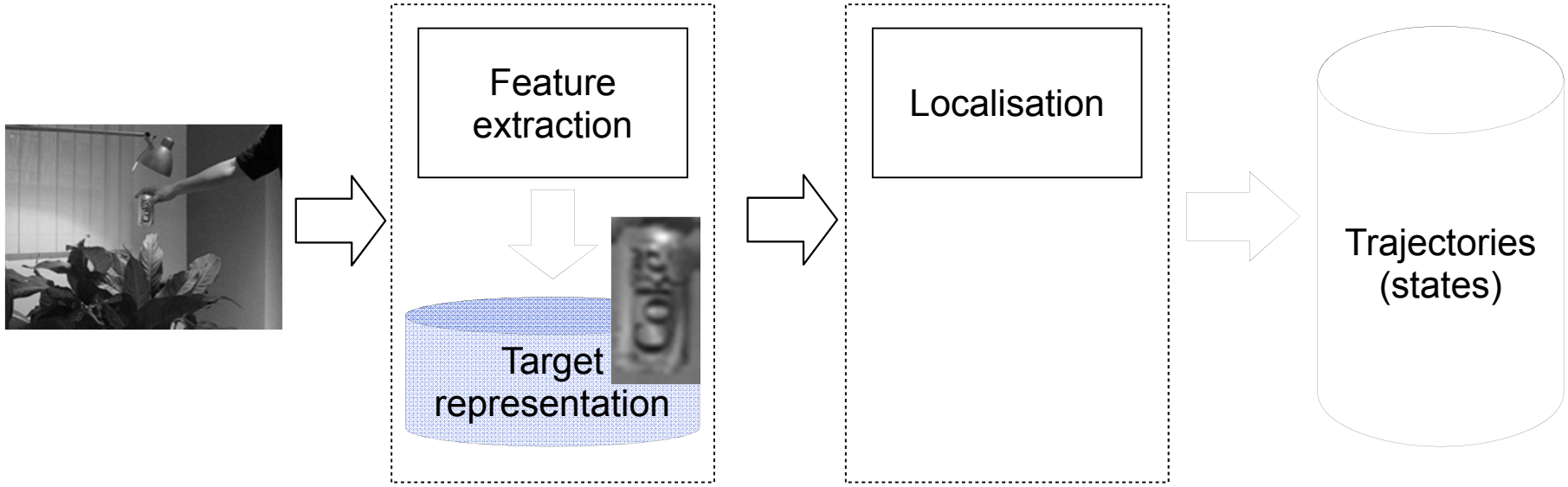
CONDENSATION



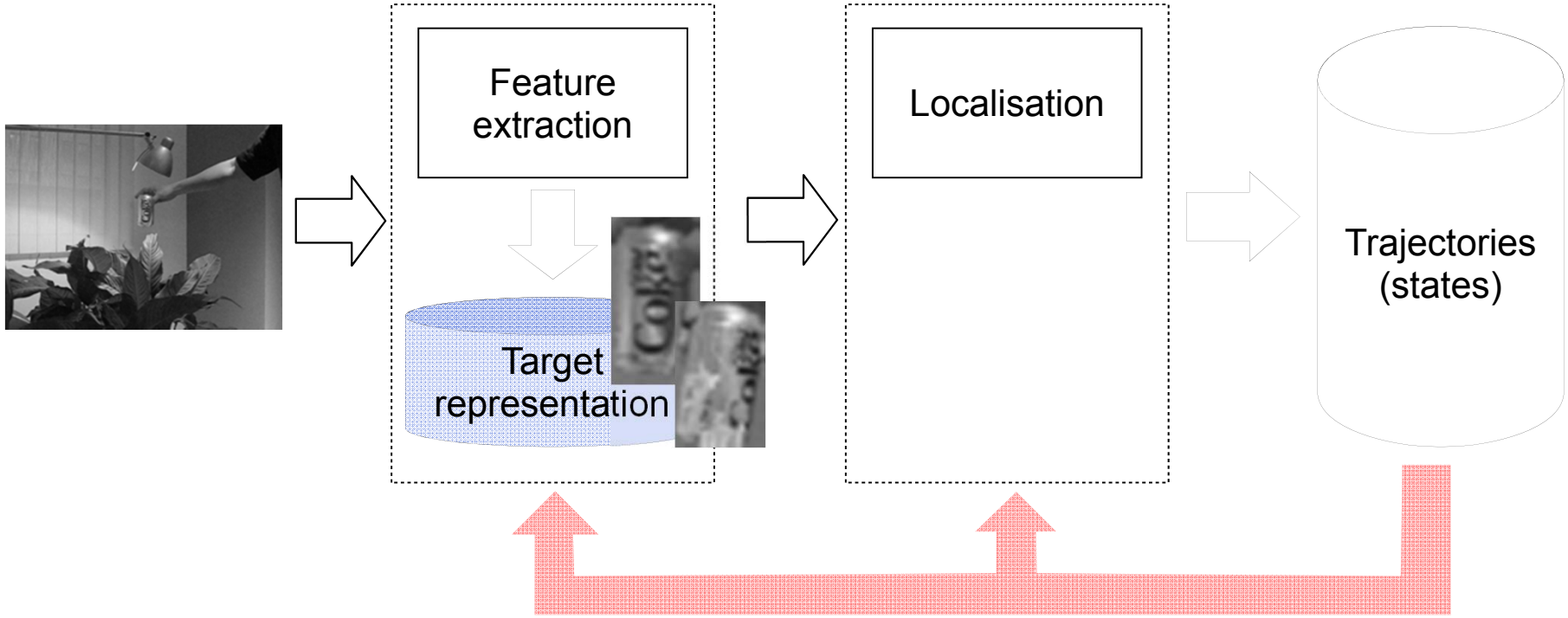
Example …



The (simplified) video tracking pipeline

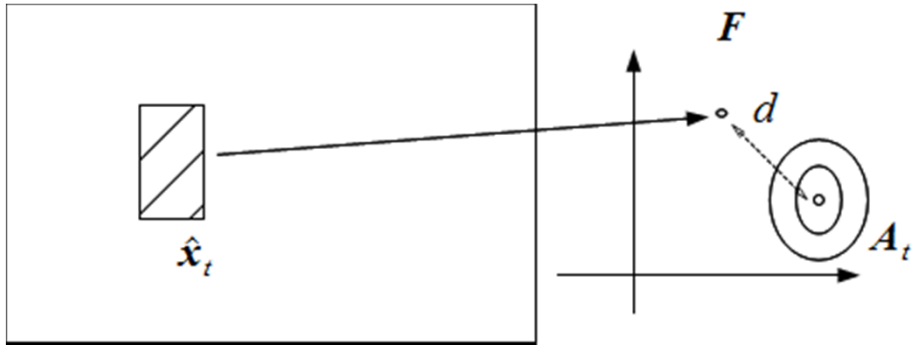


The (simplified) video tracking pipeline

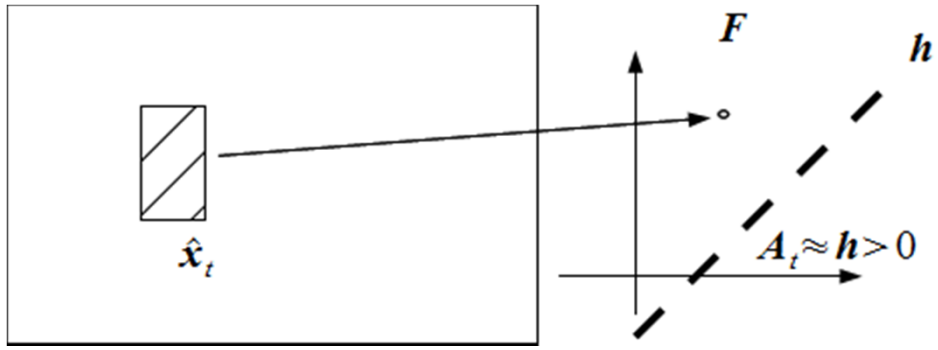


Generative vs discriminative trackers

- Generative tracker
  - (trained and) updated based on object appearance only without considering the appearance of the background or other targets

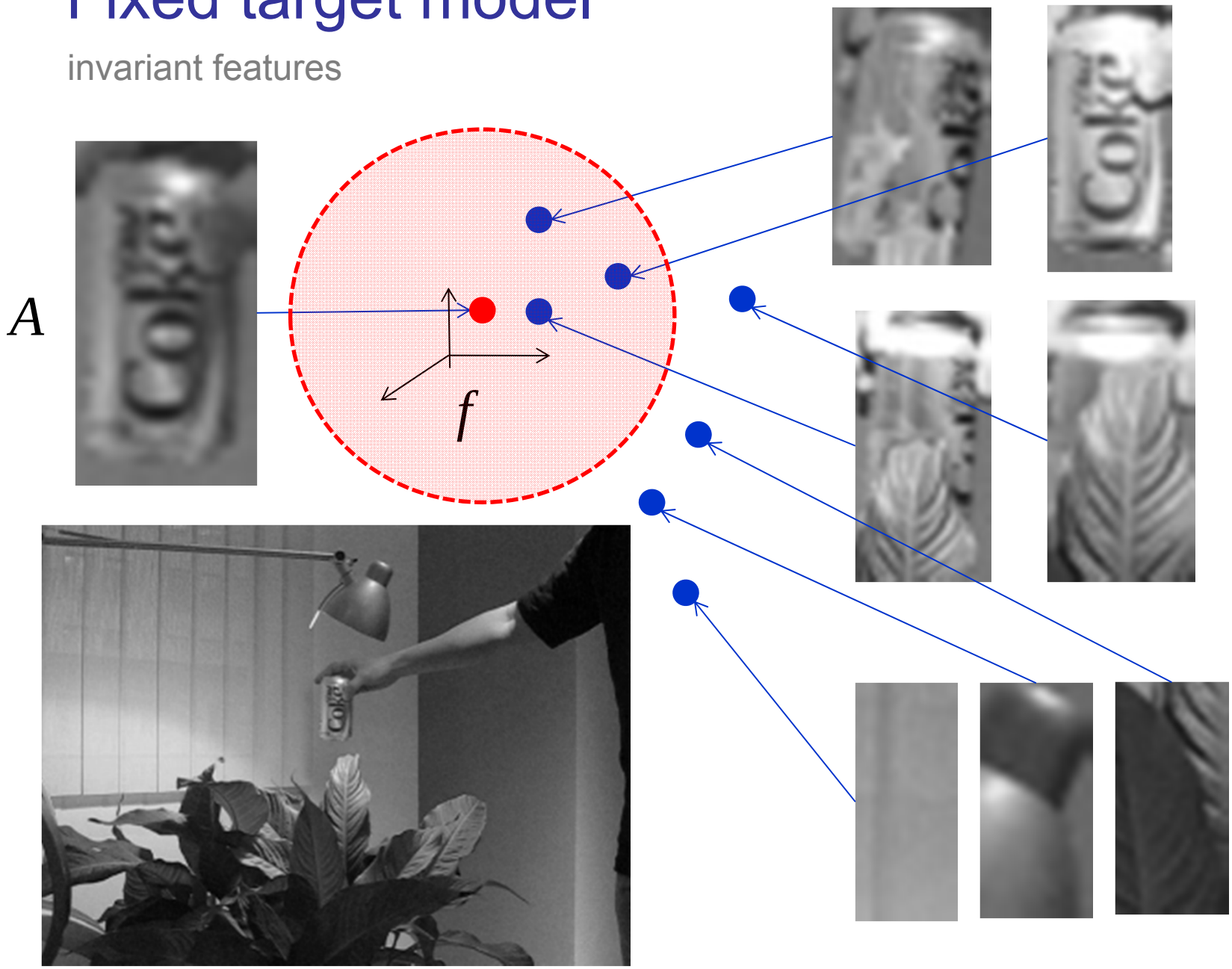


- Discriminative tracker
  - (trained and) updated to learn a decision boundary that can separate target & background and/or one target against other targets



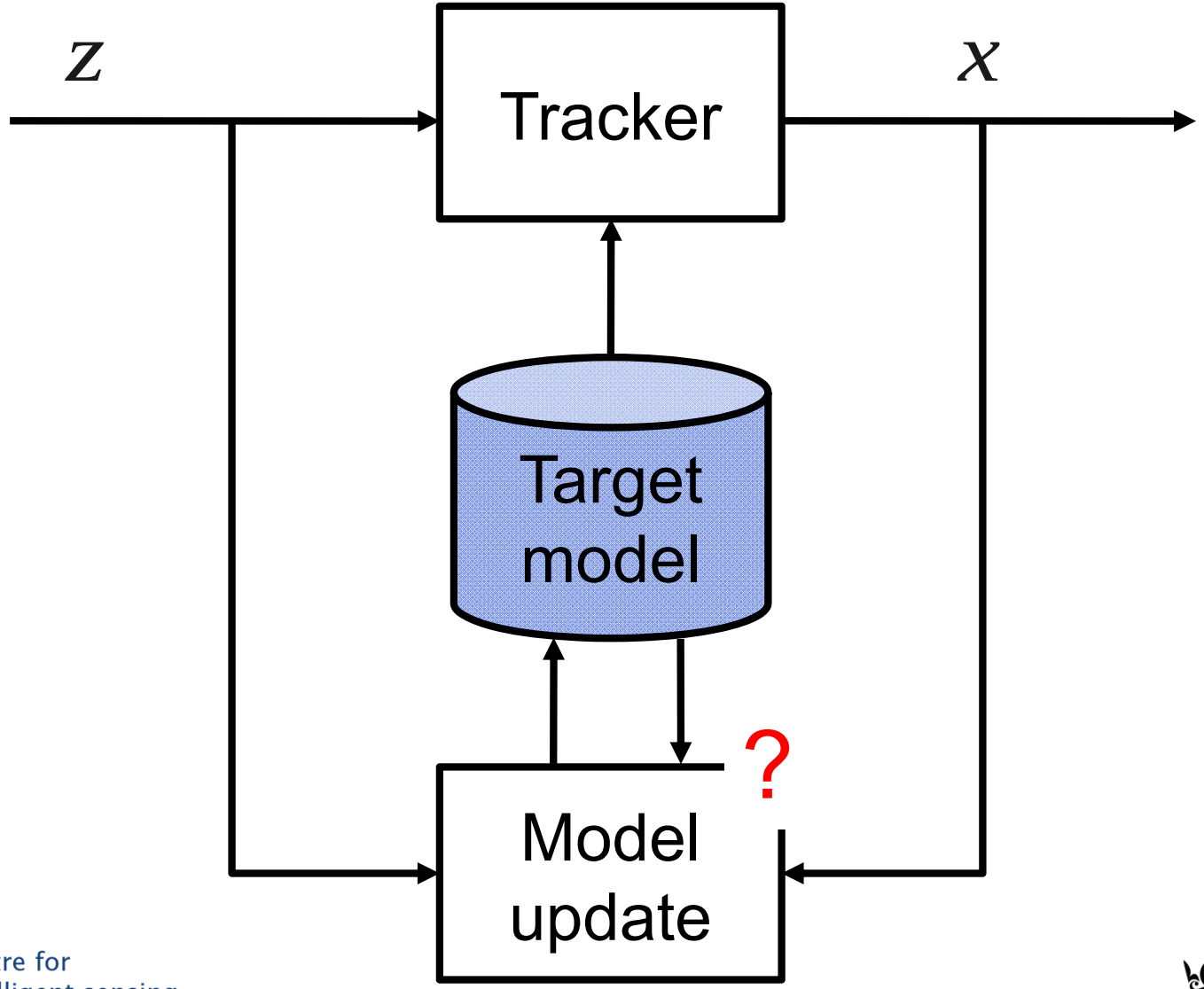
Fixed target model

invariant features

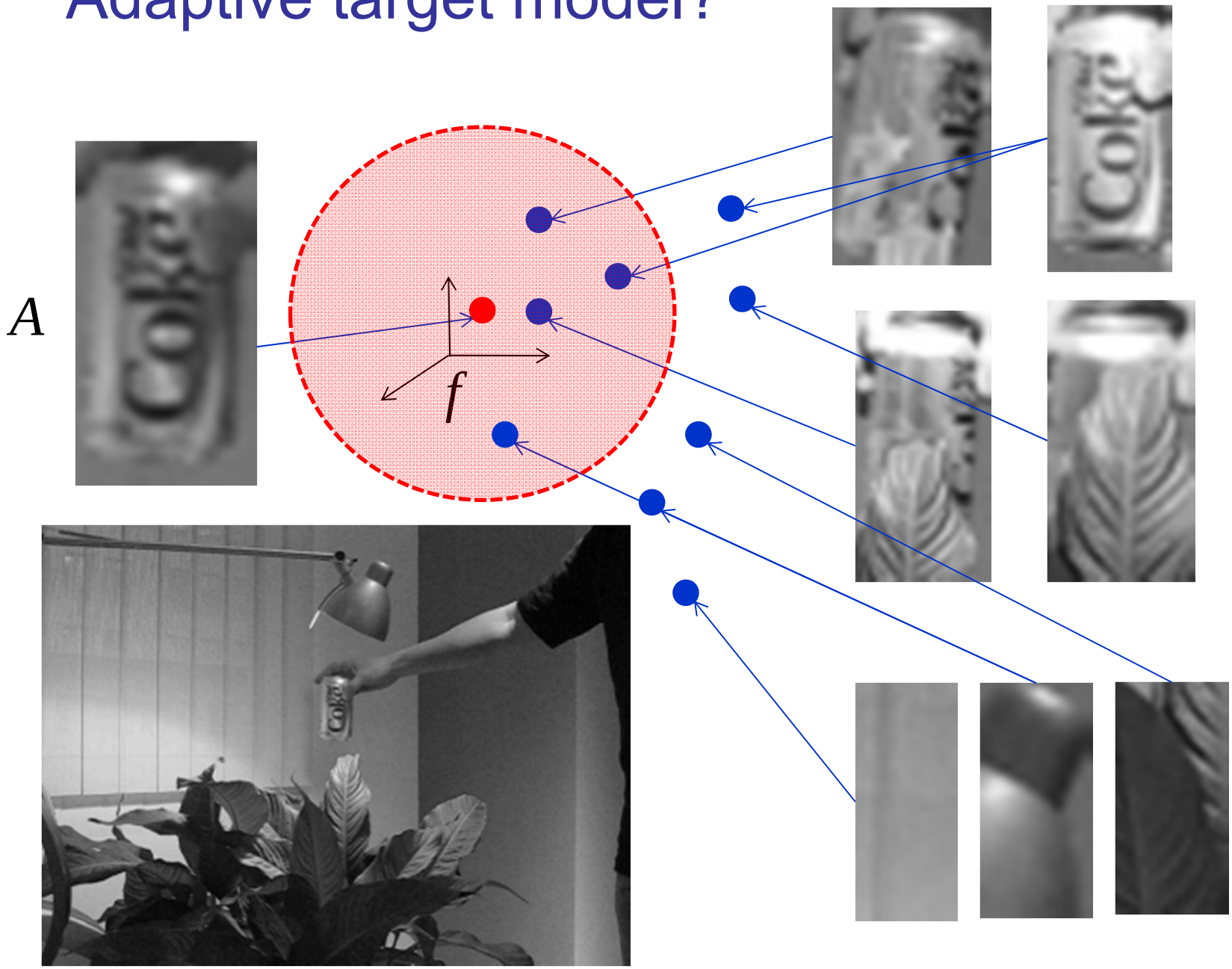


Fixed target model: another example

Coping with appearance changes



Adaptive target model?



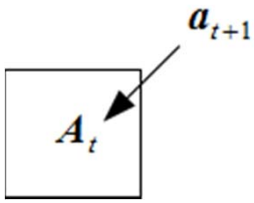
Plasticity *vs* stability dilemma

(model update)

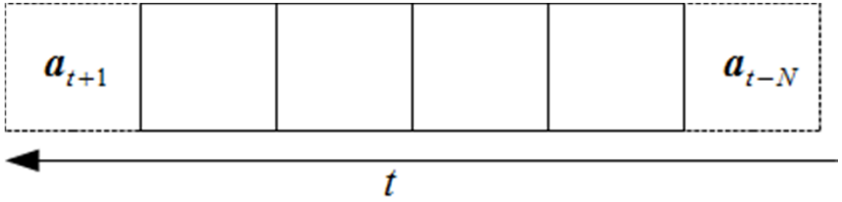
‘extreme’ plasticity

always forget

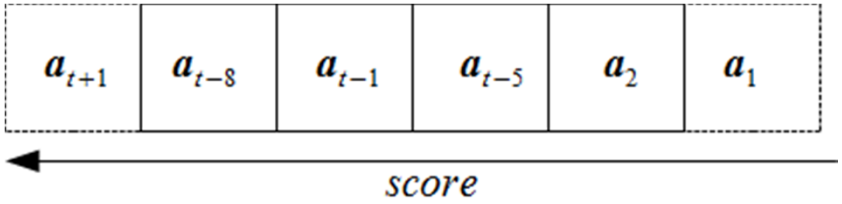
last model only



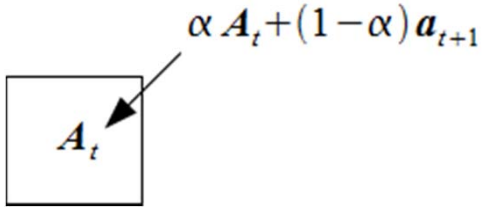
sliding window



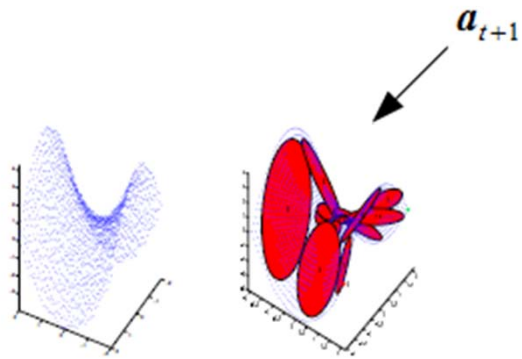
ranking



blending



subspace/manifold



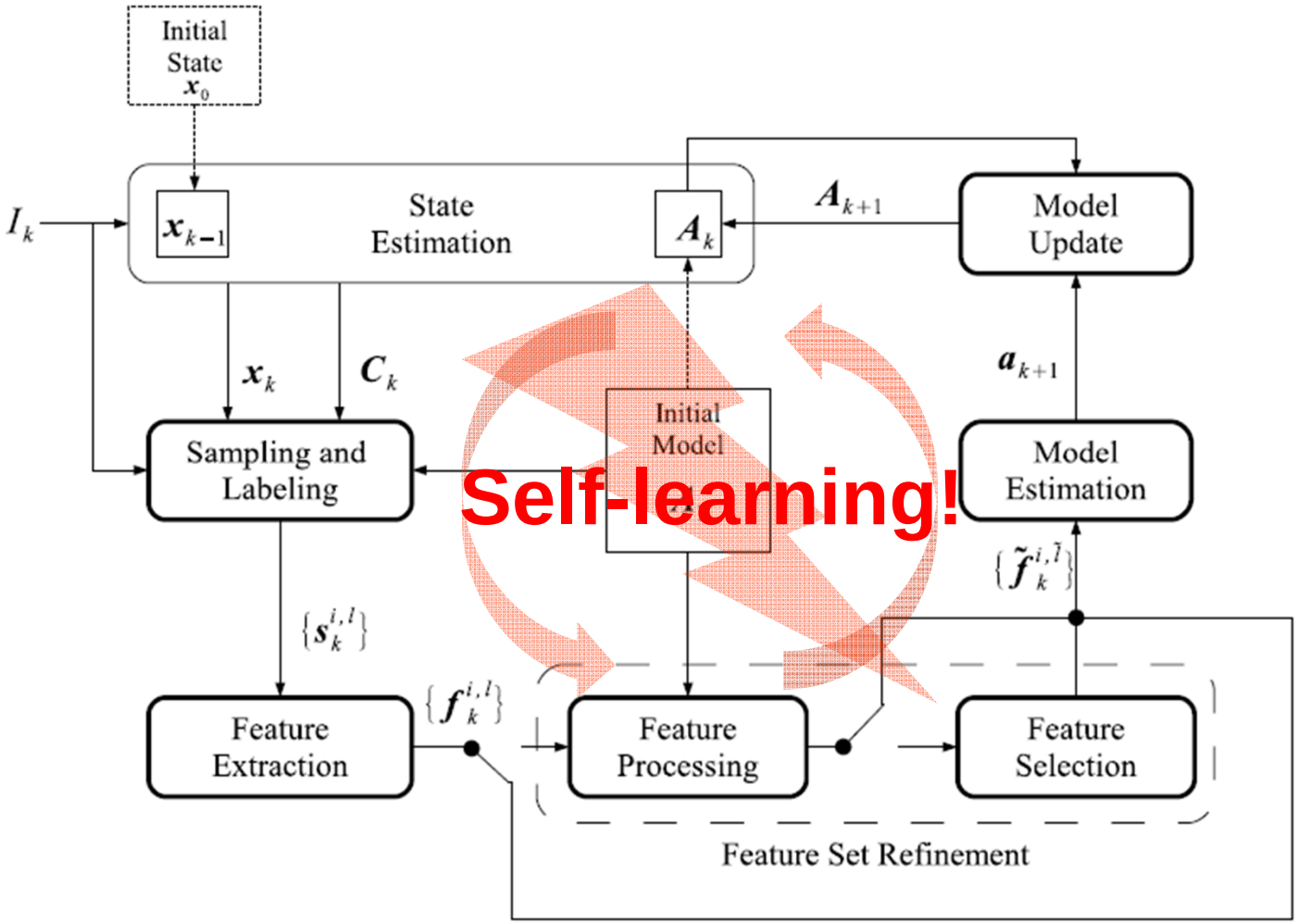
A_t overall target model

a_{t+1} model estimated in I_{t+1}

‘extreme’ stability

full history (potentially)

Adaptive target models



Adaptive appearance modeling for video tracking: survey and evaluation
S. Salti, A. Cavallaro, L. Di Stefano
IEEE Transactions on Image Processing, Vol. 21, Issue 10, October 2012

algorithm	reference	colour code
Boost	Grabner and Bischof *On-line boosting and vision*, CVPR06.	
SemiBoost	Grabner *et al.* *Semi-supervised on-line boosting for robust tracking*, ECCV08	
BeyondSemiBoost	Stalder *et al.* *Beyond semi-supervised tracking*, ICCVW09	
A-BHMC	Kwon & Lee, *Tracking […] via Patch-based Dynamic Appearance Modeling and Adaptive Basin Hopping Monte Carlo Sampling*, CVPR09	
IVT	Ross *et al.* *Incremental learning for robust visual tracking*, IJCV08	
MILBoost	Babenko *et al.* *Robust object tracking with online multiple instance learning*, PAMI11	
TLD	Kalal *et al.* *Online Learning of Robust Object Detectors During Unstable Tracking*, ICCV09	
STRUCK	Hare *et al.* *STRUCK: Structured Output Tracking with Kernels*, ICCV11	
Mean-shift	Comaniciu *et al.* *Kernel-based object tracking*, PAMI03.	
Color-based PF	Pérez *et al.* *Color-based probabilistic tracking*, ECCV02	
FragTrack	Adam et al., *Robust Fragments-based Tracking Using the Integral Histogram*, CVPR06	

Adaptive target modelling: examples

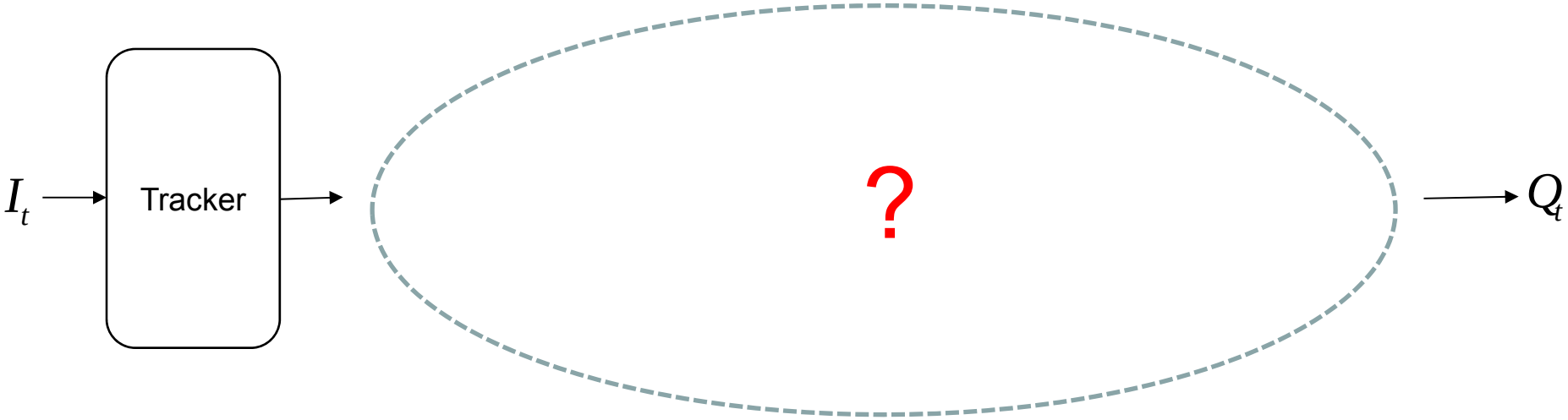
Adaptive appearance modeling for video tracking: survey and evaluation
S. Salti, A. Cavallaro, L. Di Stefano
IEEE Transactions on Image Processing, Vol. 21, Issue 10, October 2012

Adaptive target modelling: examples

Adaptive appearance modeling for video tracking: survey and evaluation
S. Salti, A. Cavallaro, L. Di Stefano
IEEE Transactions on Image Processing, Vol. 21, Issue 10, October 2012

How well am I tracking?

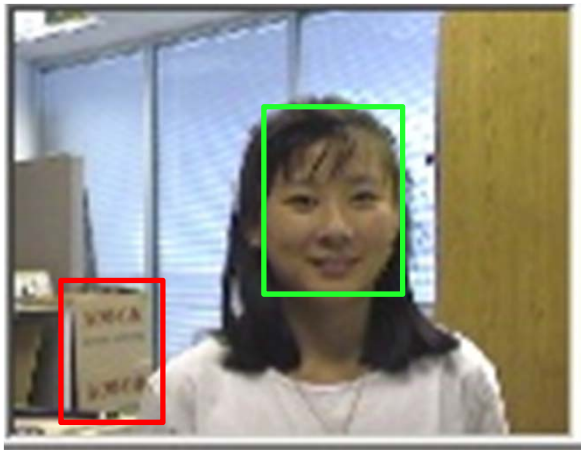
Performance evaluation


I_t
Tracker
?
Q_t

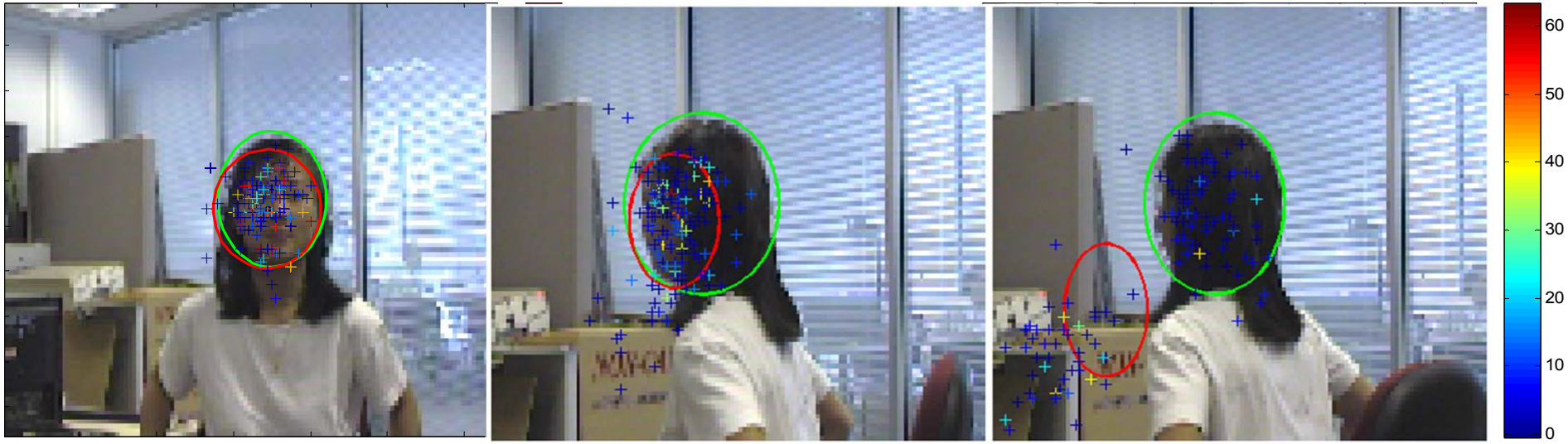
Tracking quality estimation

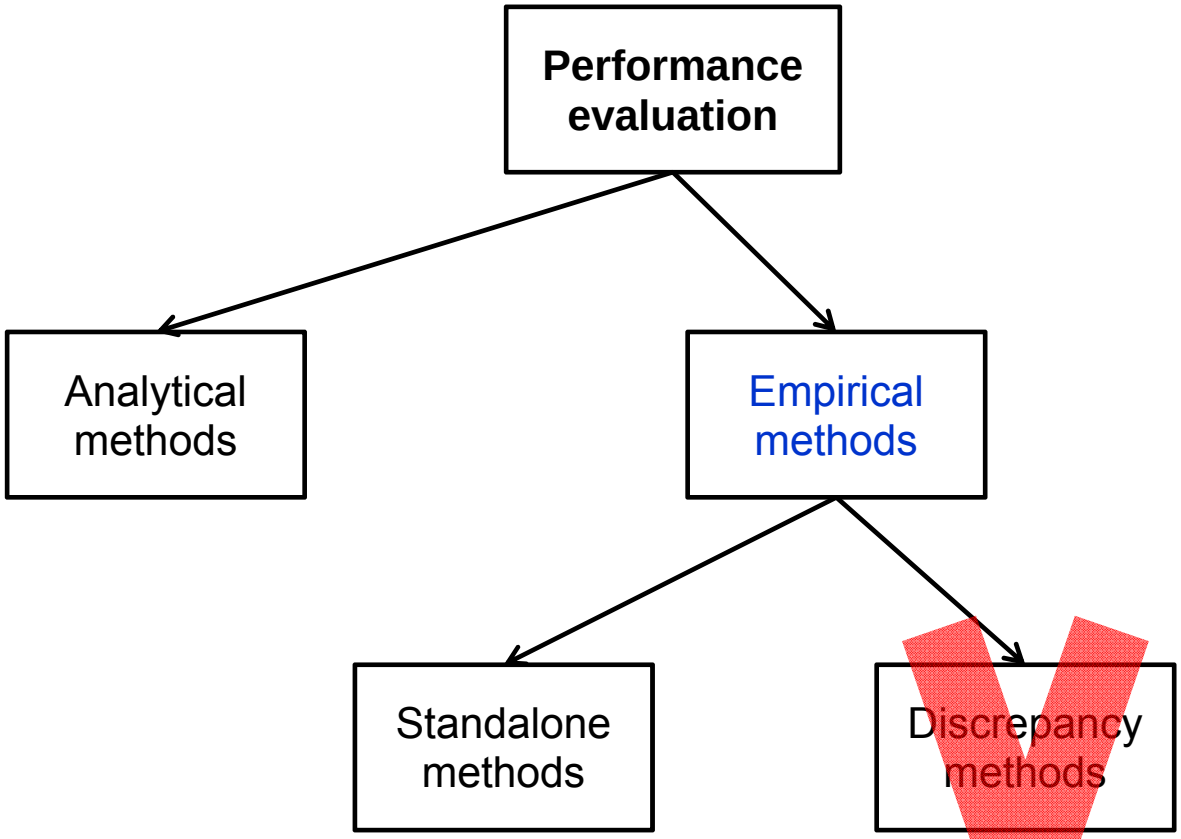
tracking result

ideal result



So, what happened?





- based on quality criteria about the trajectory

- Quantify the deviation of the results from reference results
- Need a ground-truth

What quality criteria?

Observation likelihood

✓Accurate if tracker is correct

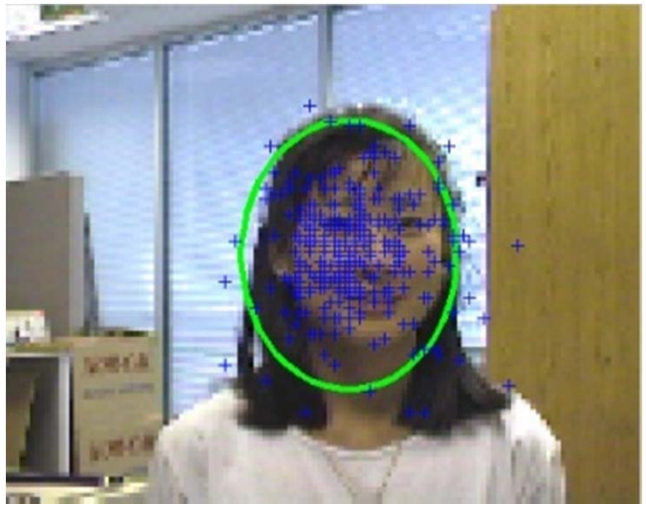
✗ Affected by distractors

$$g(z_k \mid x_k^{(i)})$$

Covariance (uncertainty)

✓Indicates target searching

✗ Affected by distractors

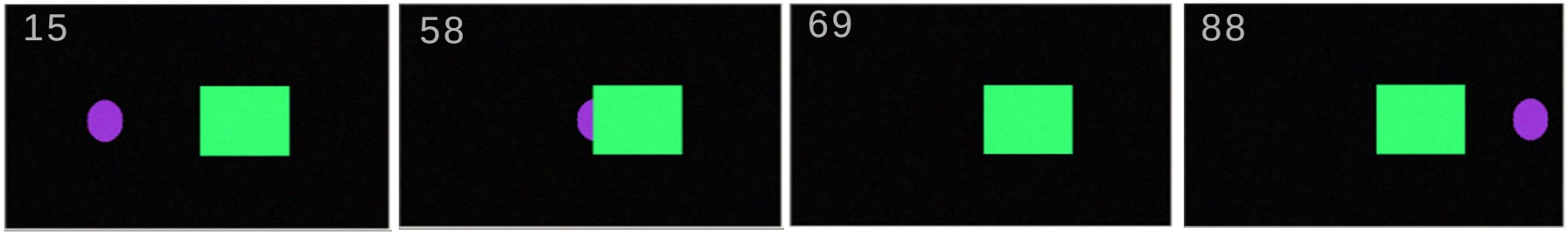


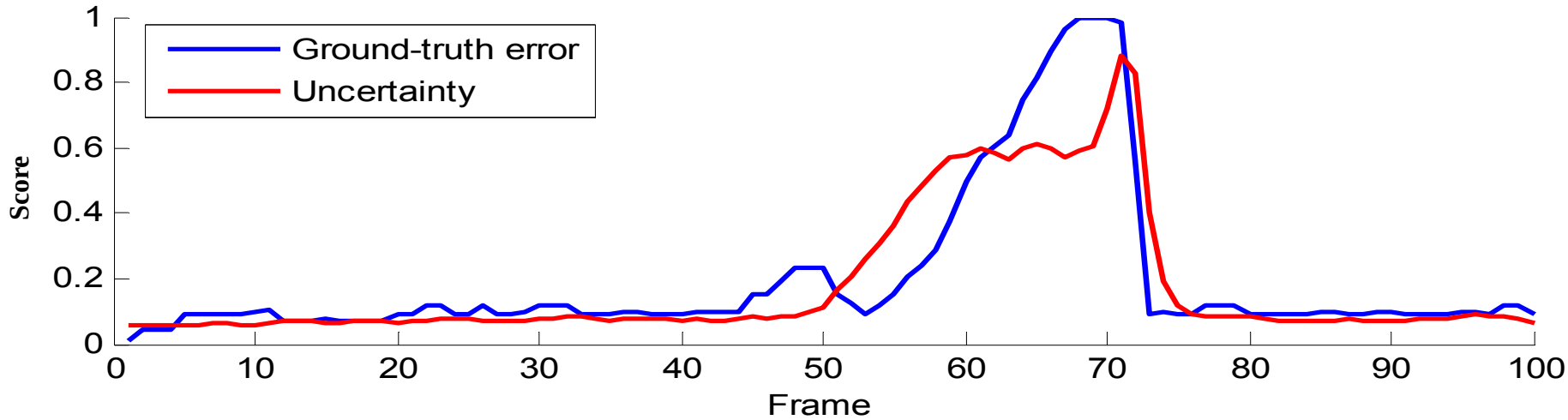
Analysis of the particle distribution

Uncertainty analysis

- Tracker status estimation
  - based on spatial uncertainty
  - multi-hypothesis tracker (e.g. Particle Filter → spread of particles)

Score: based on normalised determinant of the covariance matrix of the particle distribution

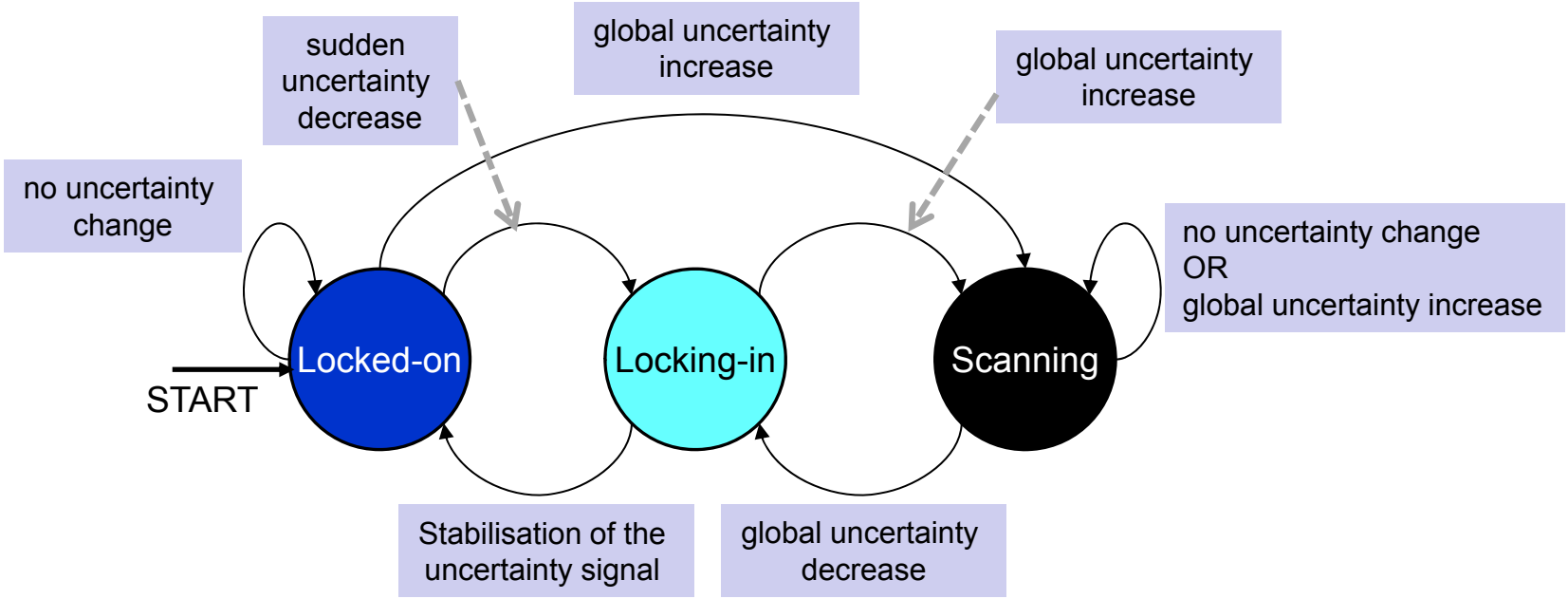




Tracker condition estimation

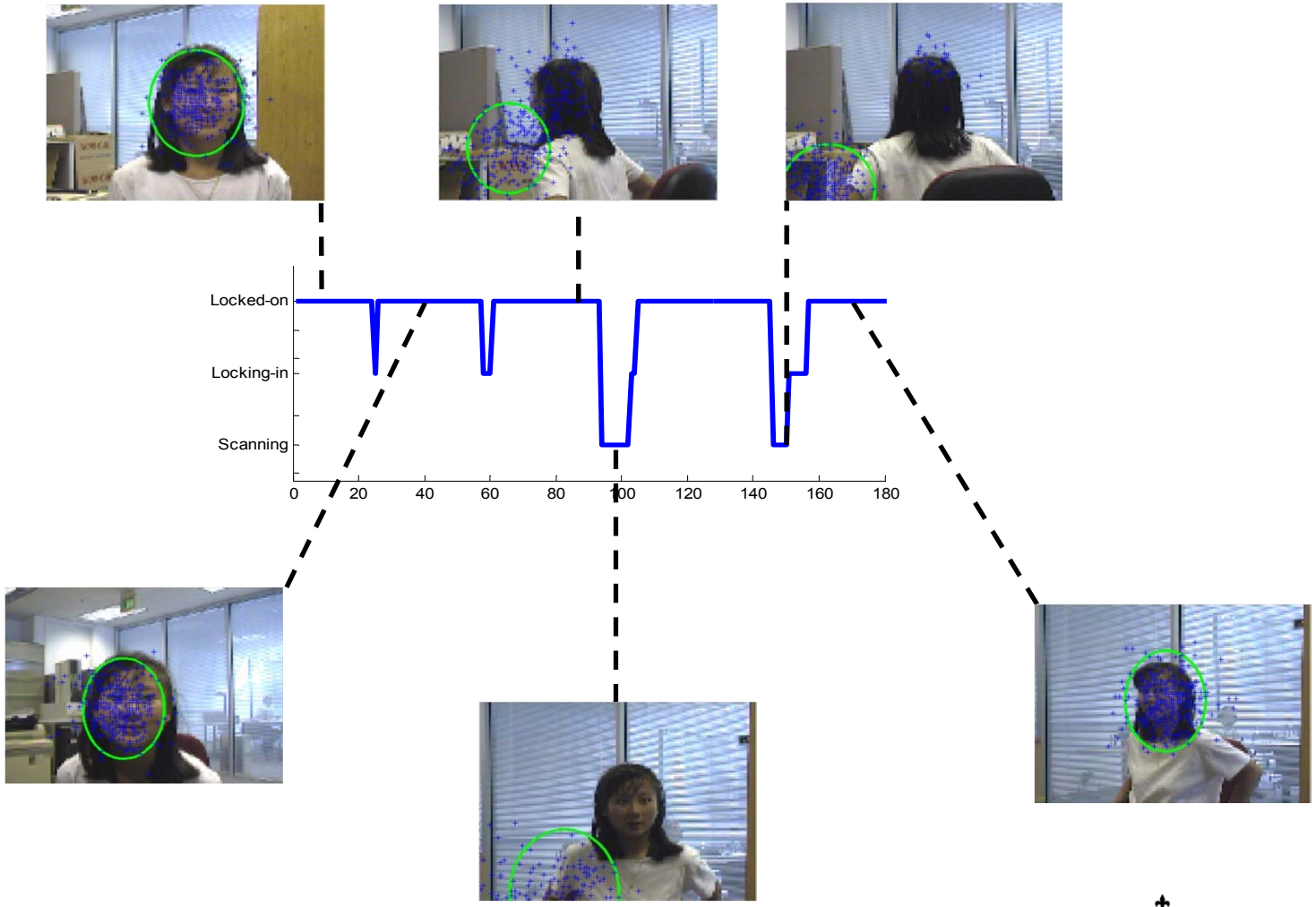
Condition estimation

- Goal: to determine transitions of tracker condition
  - **Locked-on**: tracking the target or a distractor (!)
  - Locking-in: re-focus on the target or a distractor (!) after a failure
  - **Scanning:** searching the target after a failure



Tracker condition estimation

Condition estimation



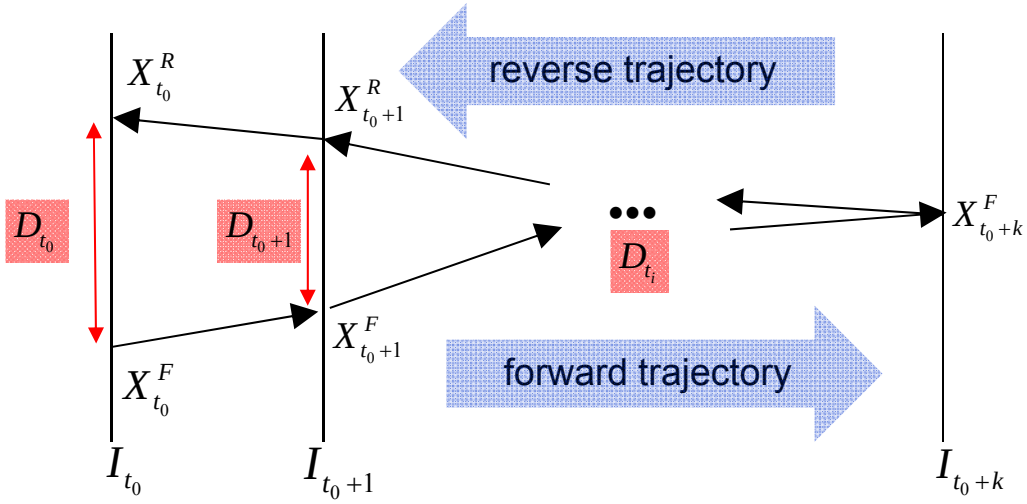
Final criterion: reverse tracking

Recovery detection

Reverse-based evaluation

✓Checks trajectory consistency

✗ Exponential complexity



- Recovery from error
  - Reverse tracking for matching: recovery ↔ reference

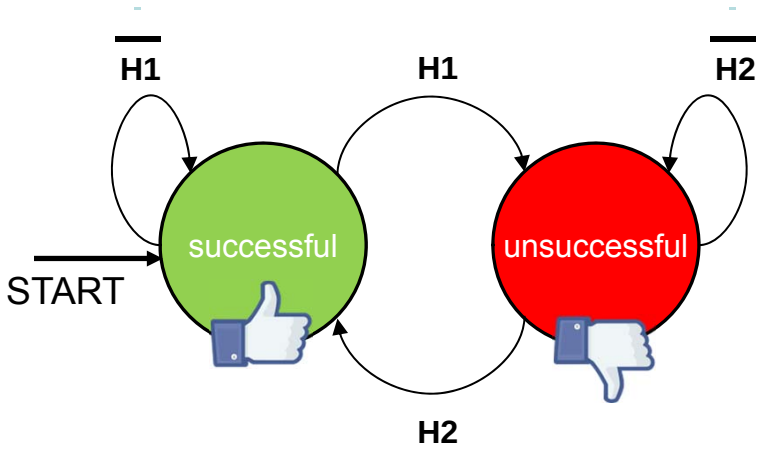
How well am I tracking?

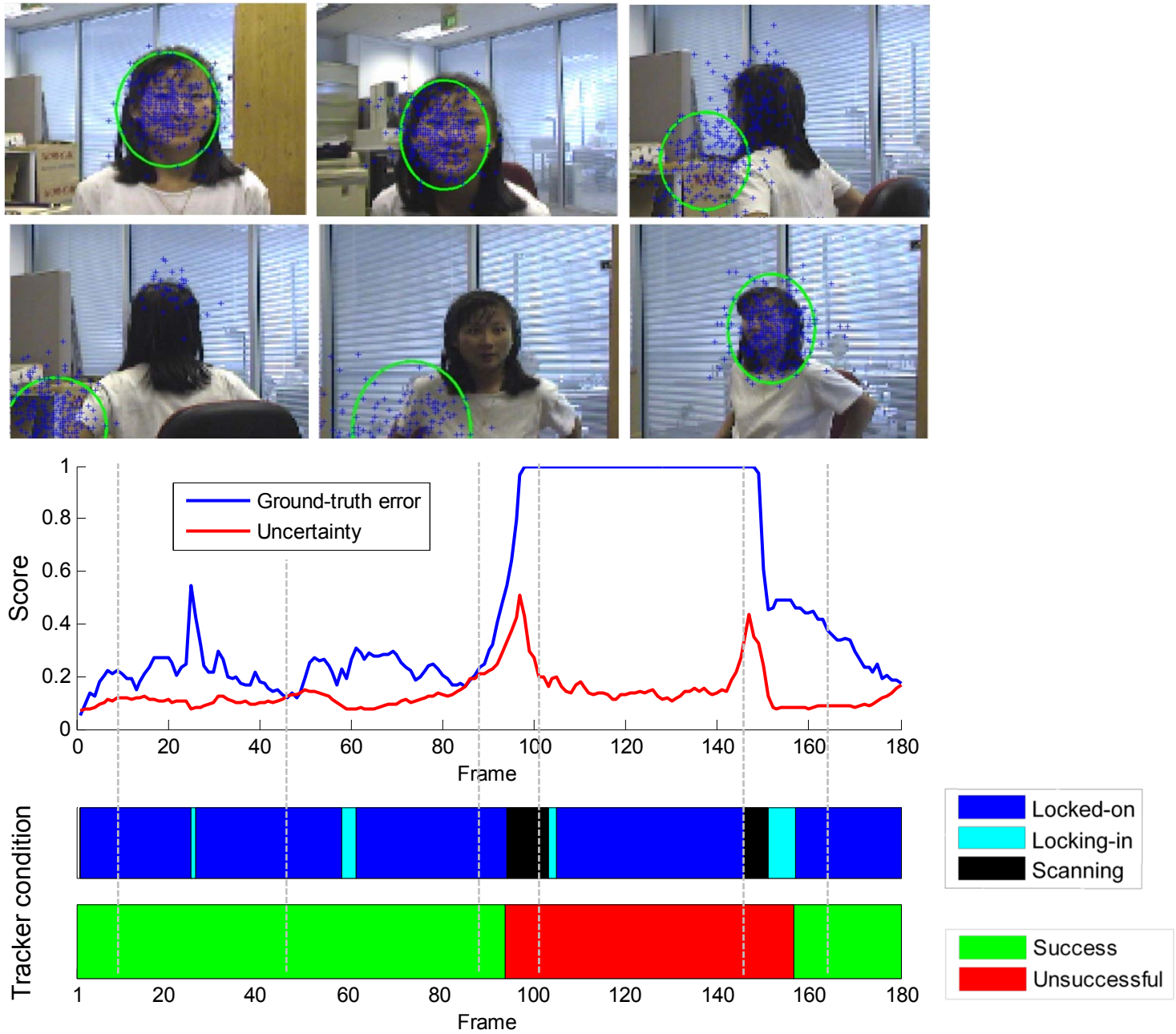
Temporal segmentation

- Temporal segmentation of tracker operation
  - determines correct tracking
    - successful
    - unsuccessful
  - based on
    - tracker condition and
    - correct recovery from error

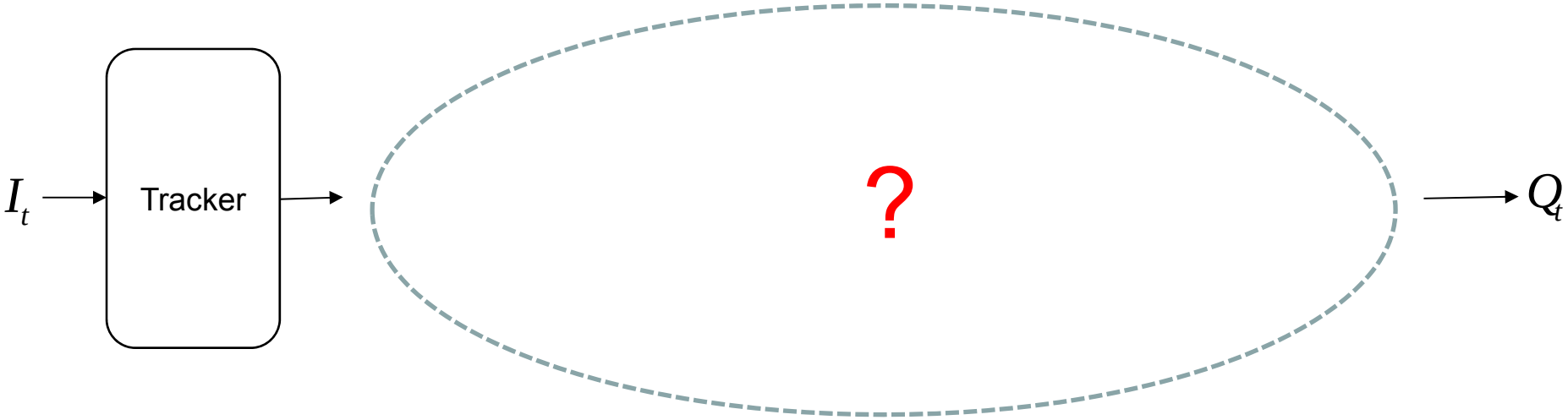
H1:
tracker remains or move to **scanning**

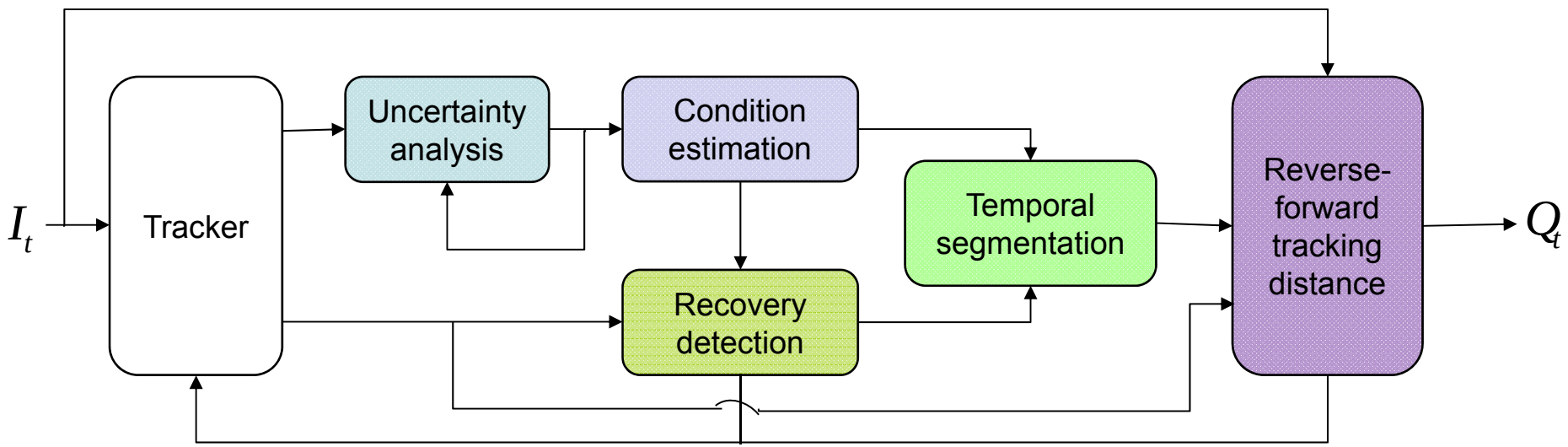
H2:
tracker moves from **locking-in** to **locked-on**
& there is a correct recovery from error




Ground-truth error
Uncertainty
Score
Frame
Tracker condition
Frame
Locked-on
Locking-in
Scanning
Success
Unsuccessful

To summarise …



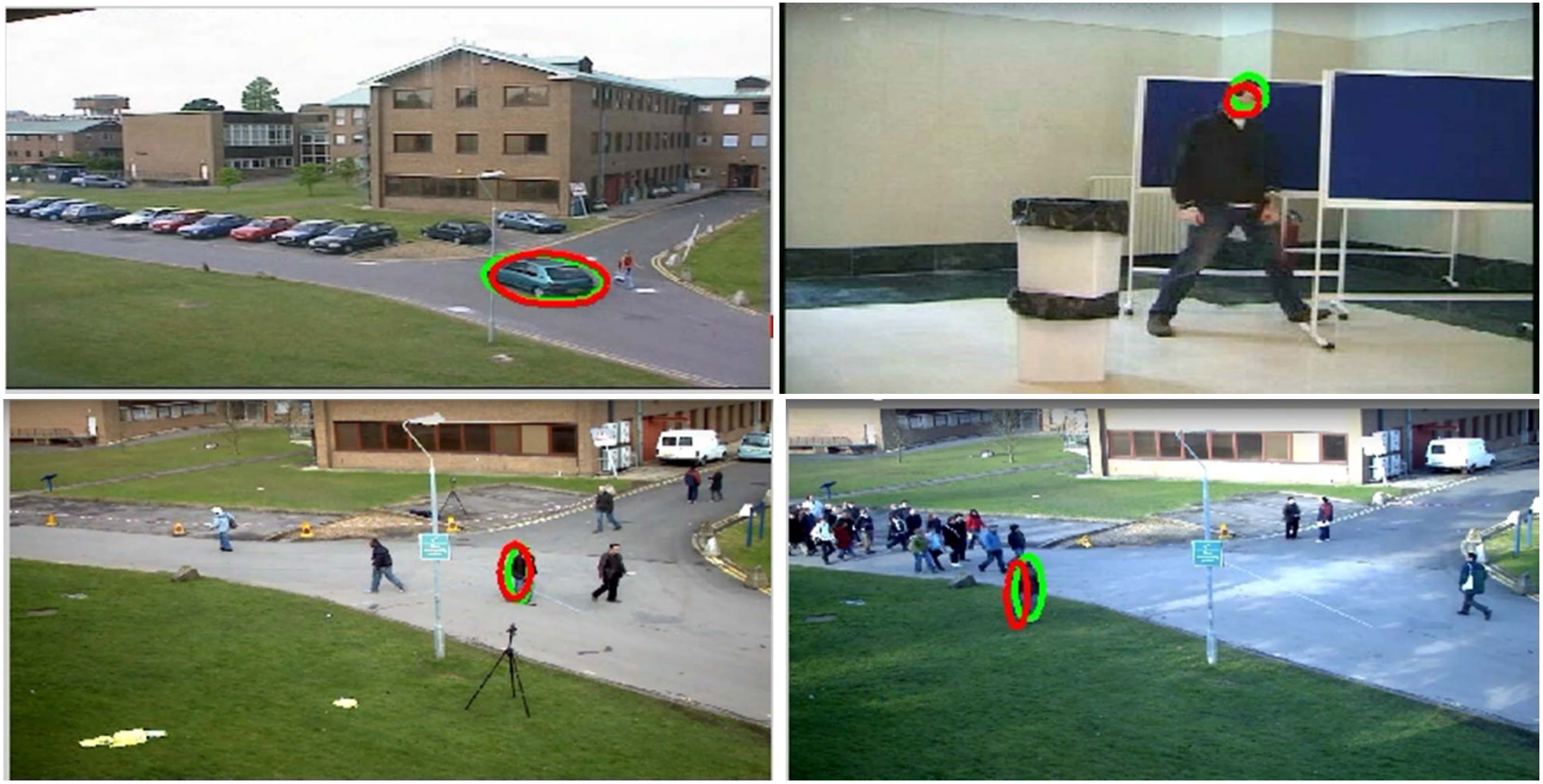


Adaptive on-line performance evaluation of video trackers

J.C. SanMiguel, A. Cavallaro and J.M. Martinez

IEEE Transactions on Image Processing, Vol. 21, Issue 5, May 2012

On-line performance evaluation: results

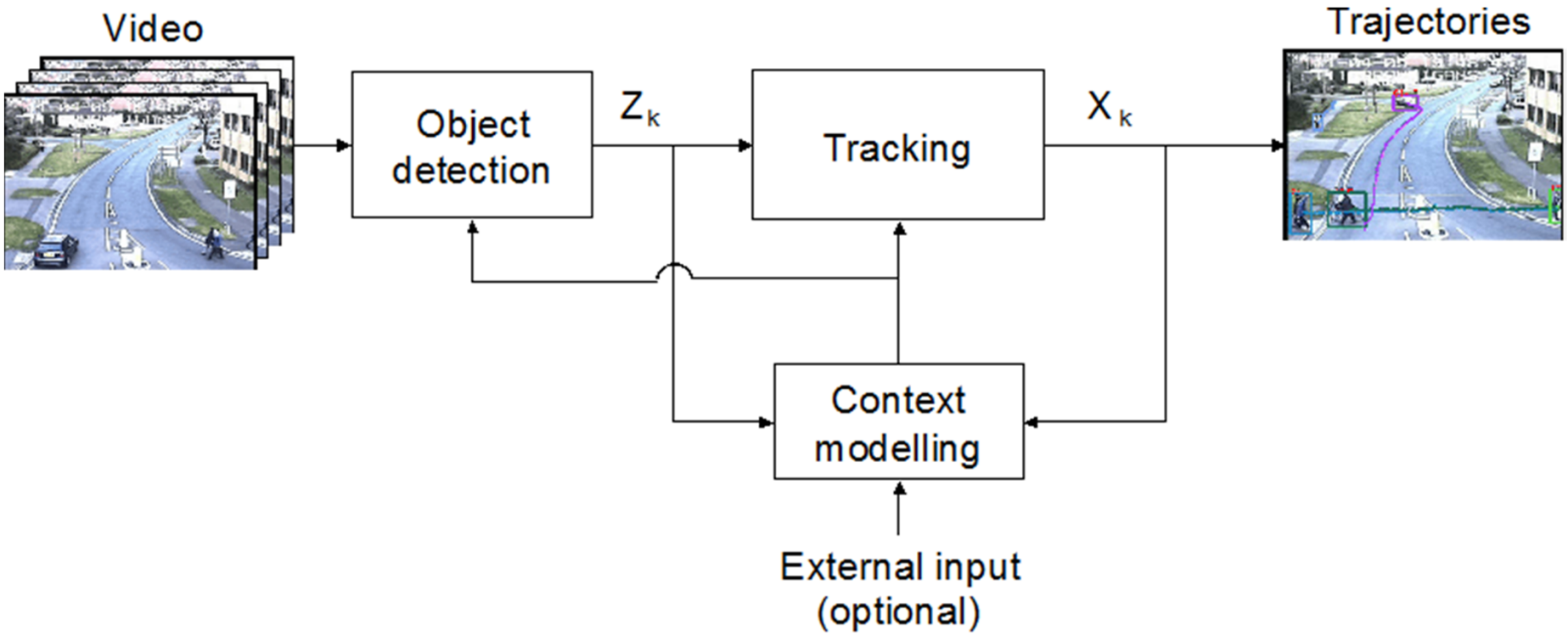


Adaptive on-line performance evaluation of video trackers
J.C. SanMiguel, A. Cavallaro and J.M. Martinez
IEEE Transactions on Image Processing, Vol. 21, Issue 5, May 2012

Can scene modelling help tracking?

(multiple objects)

Tracking with (adaptive) context modelling



Learning scene context for multiple object tracking
E. Maggio and A. Cavallaro
IEEE Transactions on Image Processing, Vol. 18, Issue 8, August 2009, pp. 1873-1884

Multi-target observation and state

- Observations

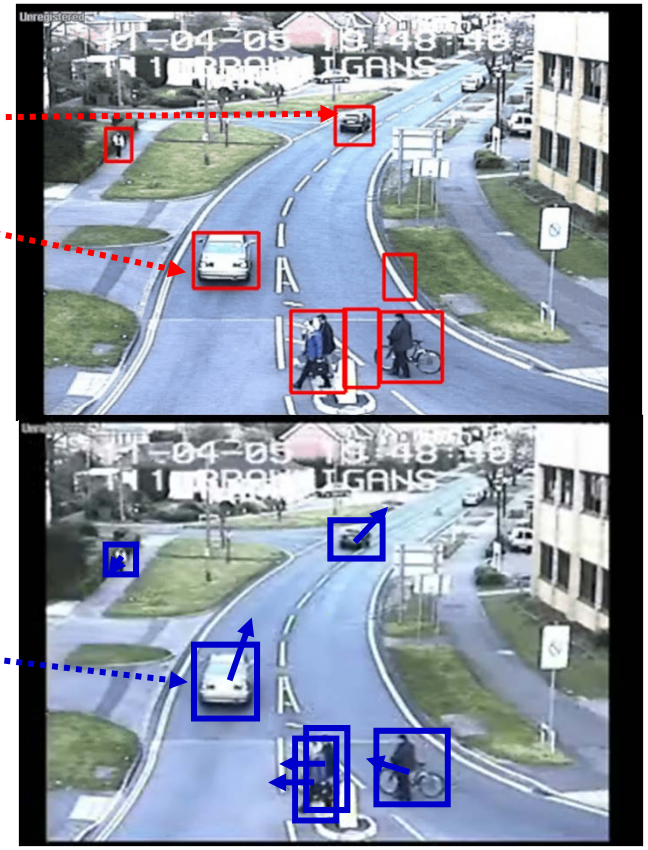
$$Z_k = \{z_{k,1}, \ldots, z_{k,N(k)}\}$$

$$z = [y_1, y_2, w, h]$$

- State

$$X_k = \{x_{k,1}, \ldots, x_{k,M(k)}\}$$

$$x = [y_1, y_2, \dot{y}_1, \dot{y}_2, w, h]$$



Multi-target Bayes filter

Single target

Random vectors

$$\mathbf{x}_k = \mathbf{f}\left(x_{k-1}, \mathbf{u}_k\right)$$
$$\mathbf{z}_k = \mathbf{g}\left(x_k, \mathbf{n}_k\right)$$

Multiple targets

Random Finite Sets (RFS)

$$\Xi_k = S_k\left(X_{k-1}\right) \cup \Gamma_k$$
$$\Omega_k = \Theta_k\left(X_k\right) \cup \mathrm{K}_k$$

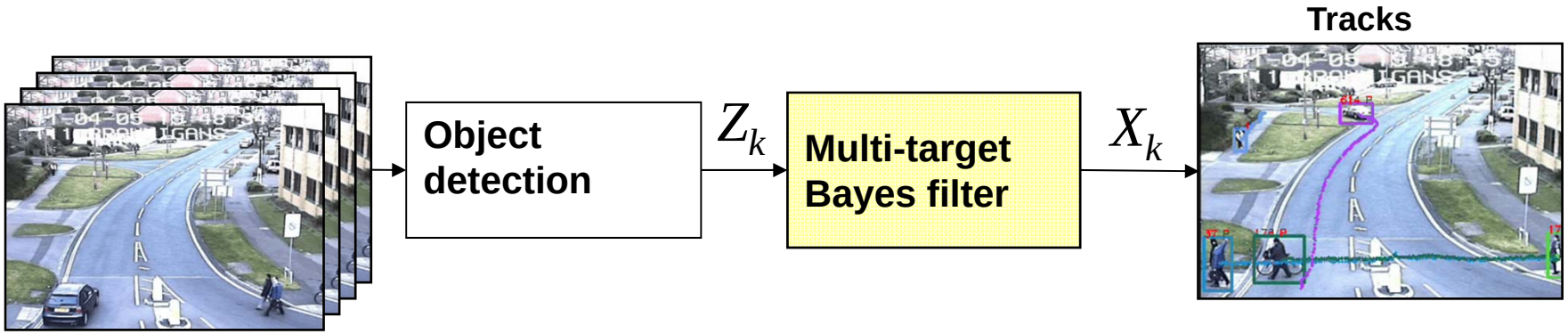
- Recursively propagates the multi-target state
  - Prediction

$$p(X_k \mid Z_{1:k-1}) = \int f(X_k \mid X_{k-1}) p(X_{k-1} \mid Z_{1:k-1}) \mu_s\left(dX_{k-1}\right)$$

  - Update

$$p(X_k \mid Z_{1:k}) = \frac{g(Z_k \mid X_k) p(X \mid Z_{1:k-1})}{\int p(Z_k \mid X_k) p(X_k \mid Z_{1:k-1}) \mu_s\left(dX_k\right)}$$

Multi-target Bayes filter (2)



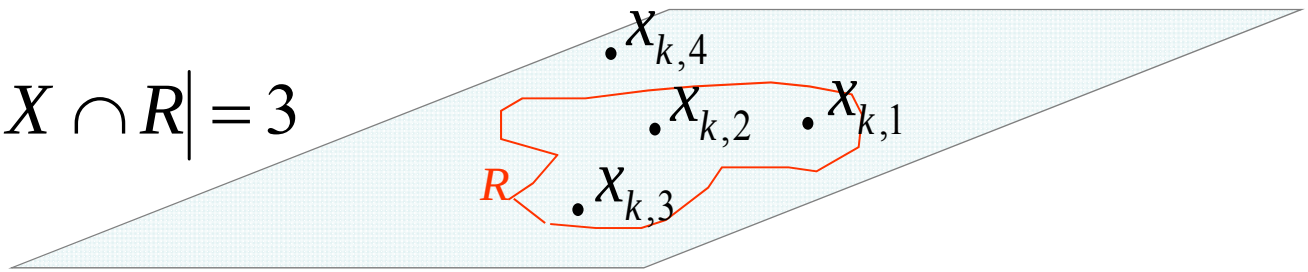
- Model target *birth*, *death*, *clutter* and *missing detections*
- Monte Carlo approximation
  - Dimensionality of the state grows with the number of targets
  - Number of particles is exponential with the number of targets
  - Not feasible in real-world applications! ☹
- Solution
  - Probability Hypothesis Density (PHD) filter

Probability Hypothesis Density (PHD)

- The first order moment of a RFS → PHD
  - The integral on any R gives the expected number of targets in R

$$\int_R D_\Xi(x)dx = E\left[\left|\Xi \cap R\right|\right]$$

$$\left|X \cap R\right| = 3$$



$$\begin{array}{lccccccc}
\textbf{\textit{Multiple-target Bayes filter}} & \longrightarrow & p(X_{k-1}|Z_{1:k-1}) & \longrightarrow & p(X_k|Z_{1:k-1}) & \xrightarrow[Z_k]{} & p(X_k|Z_{1:k}) & \longrightarrow \cdots \\
 & & \downarrow & & \downarrow & & \downarrow & \\
\textbf{\textit{PHD Filter}} & \longrightarrow & D_{k-1|k-1} & \longrightarrow & D_{k|k-1} & \xrightarrow{} & D_{k|k} & \longrightarrow \cdots
\end{array}$$

Recursive PHD filter

- Prediction

$$D_{k|k-1}(x_k) = \gamma(x) + \int \phi(x_k, x_{k-1}) D_{k-1|k-1}(x_{k-1}) dx_{k-1}$$

  – PHD state transition

Birth intensity

$$\phi(x_k, x_{k-1}) = e(x_k) f(x_k \mid x_{k-1})$$

Survival probability

Single target transition

- Update

$$D_{k|k}(x_k) = \left[p_M(x_k) + \sum_{z \in Z_k} \frac{\psi_z(x_k)}{\kappa(z) + \langle \psi_z, D_{k|k-1}(x_k) \rangle} \right] D_{k|k-1}(x_k)$$

  – PHD likelihood

Missing detection

Clutter intensity

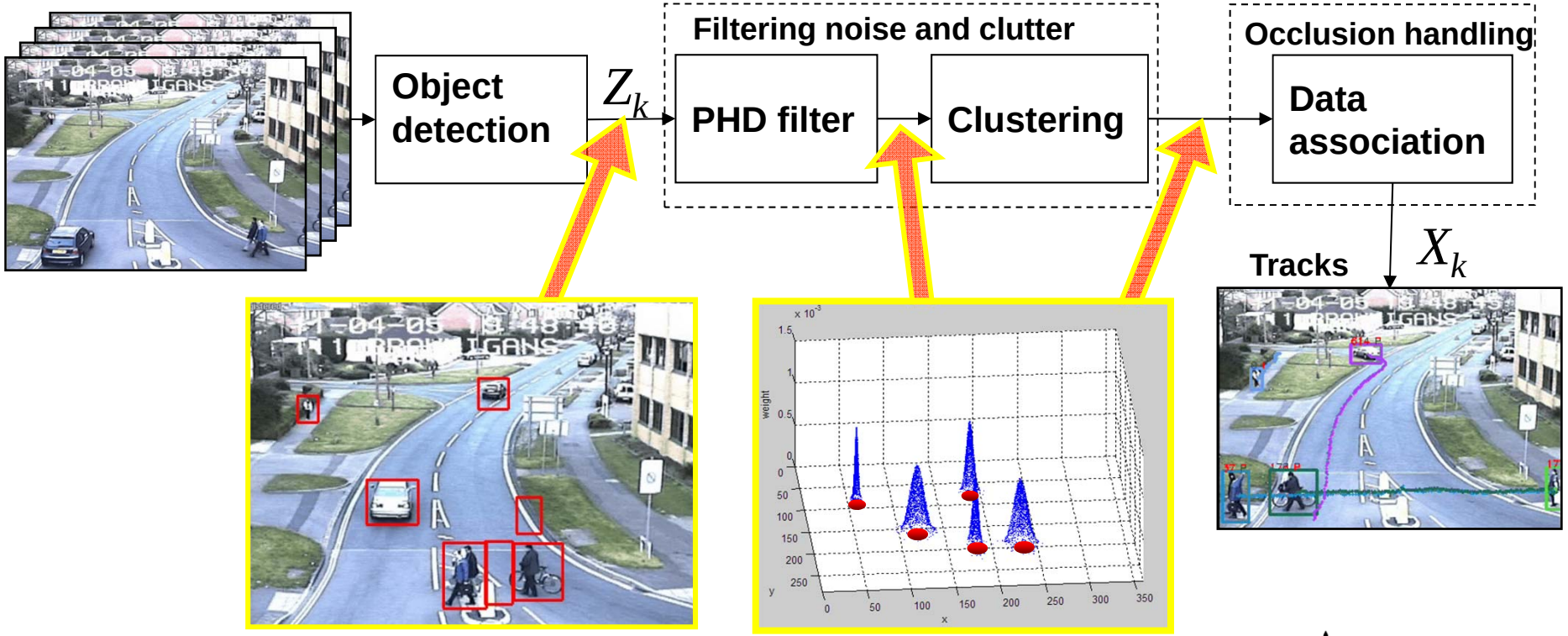
Normalization

$$\psi_z(x) = (1 - p_M(x)) g(z \mid x)$$

Single target likelihood

Tracking with the PHD filter

- **Advantage**: Linear complexity with the target number
- **Drawback**: Data association is not integrated in the Bayesian recursion



Results – CLEAR/i-LIDS dataset

Detections PHD output

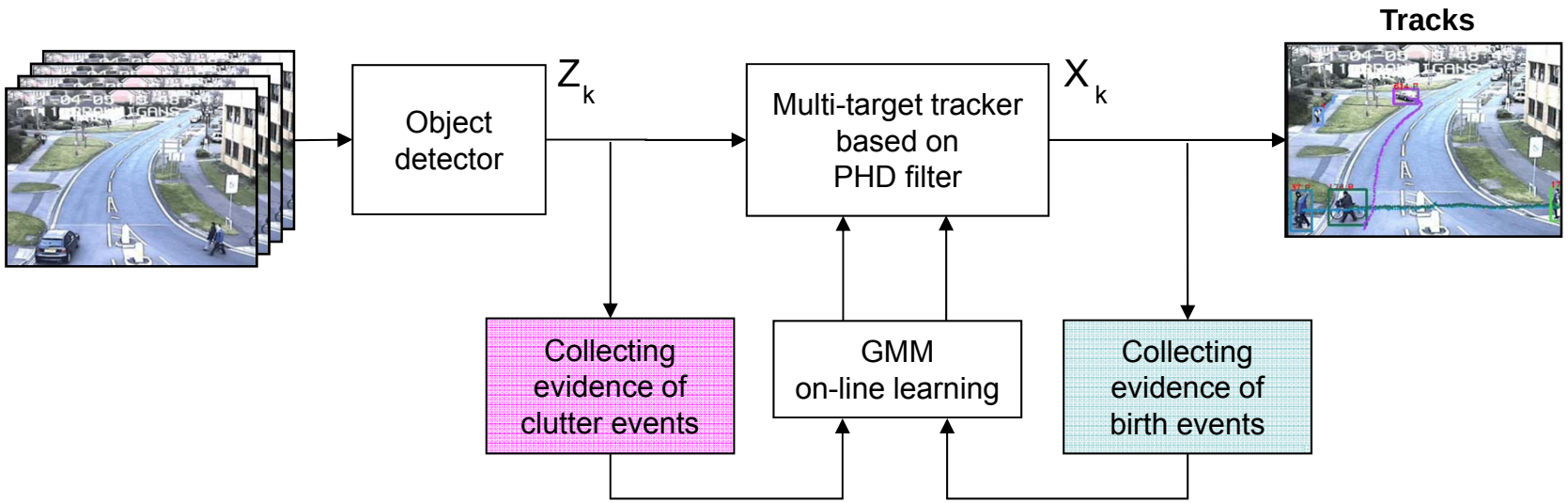
Context information: examples

Clutter area

Entry/exit area

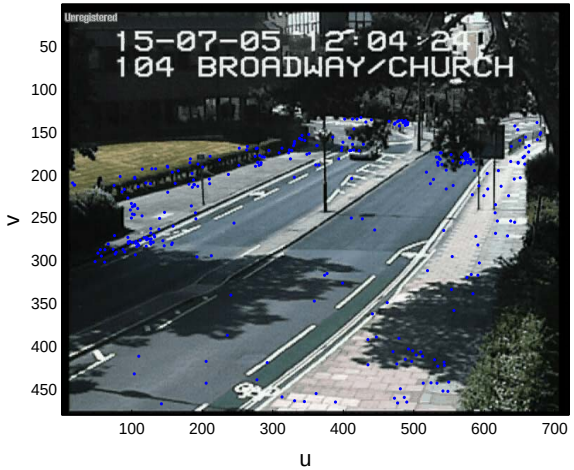
- Sidewalks and roads → likely **entry/exit** points
- Vegetation → likely source of **clutter**

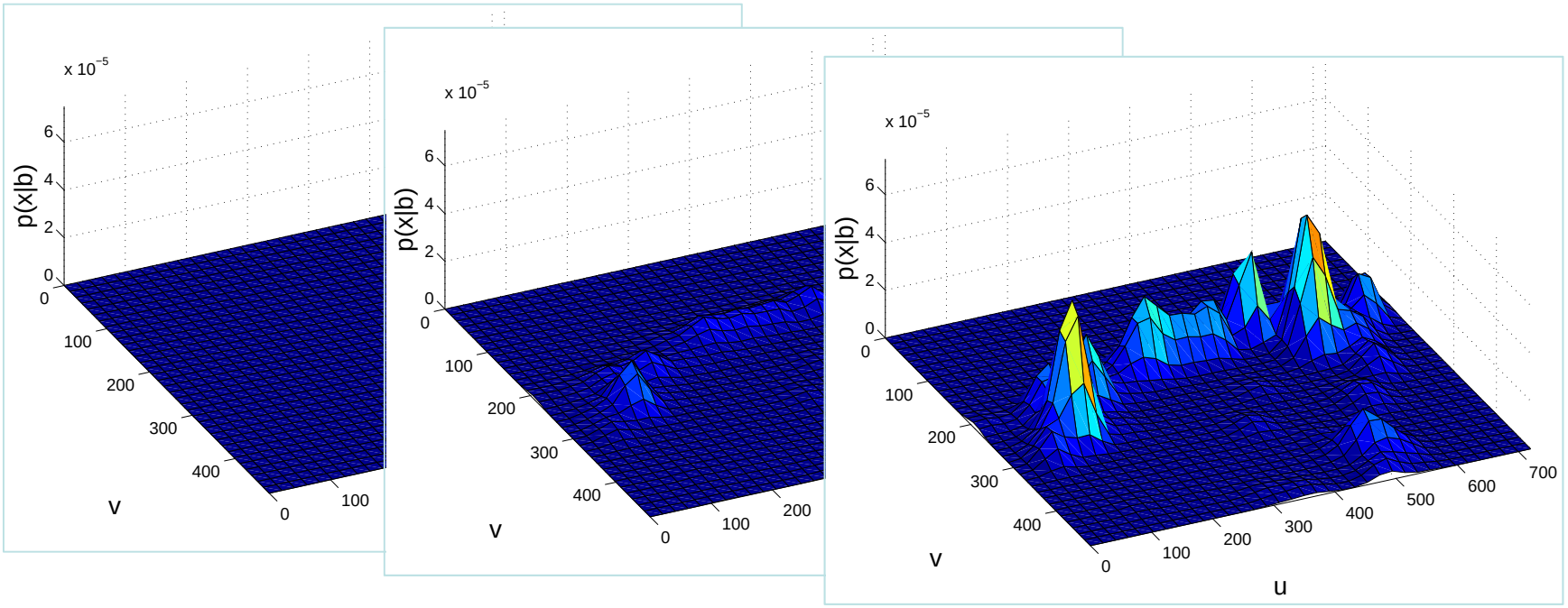
Using scene context in the PHD recursion



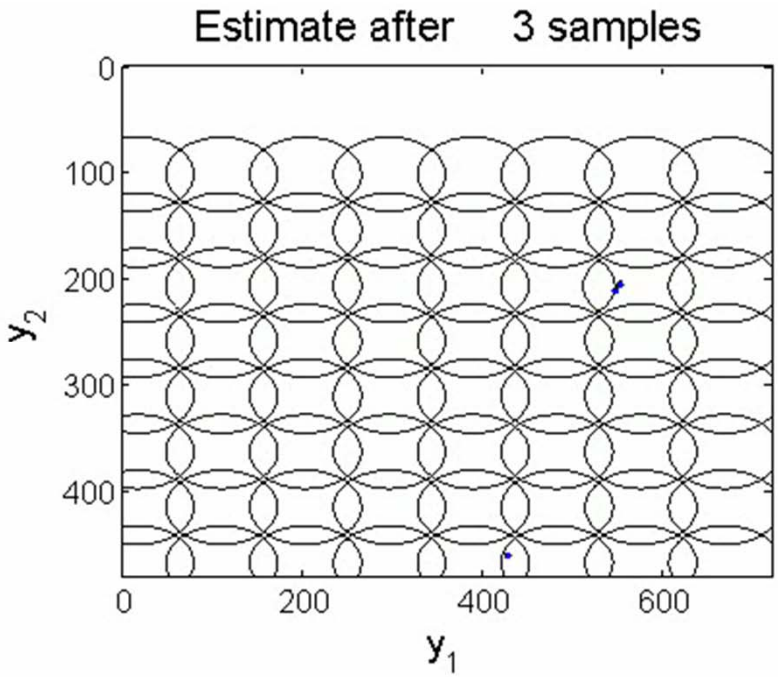
- Goal: context adaptive filtering
  - Birth intensity -- **to model entry areas**
  - Clutter intensity -- **to model detection errors**
- Data collection
  - Birth events from the output of the tracker
  - Clutter events via user feedback
- Density estimation

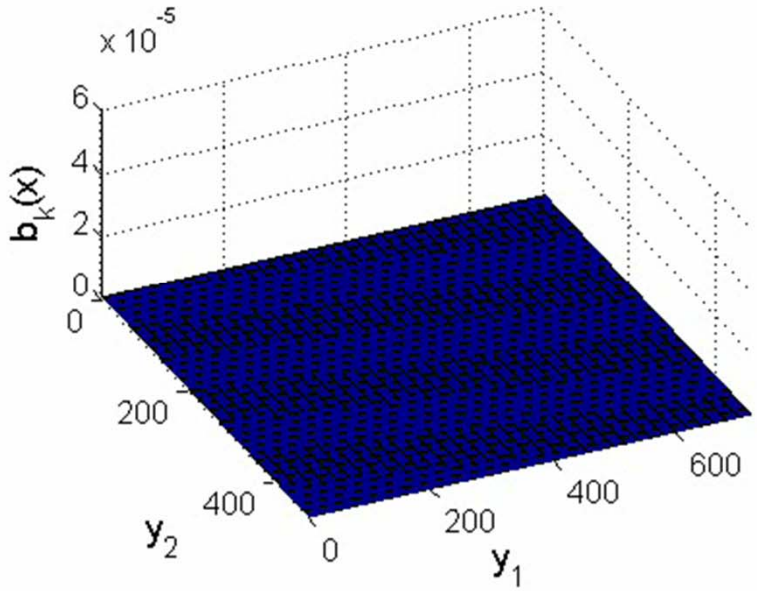
Example: incremental learning birth intensity



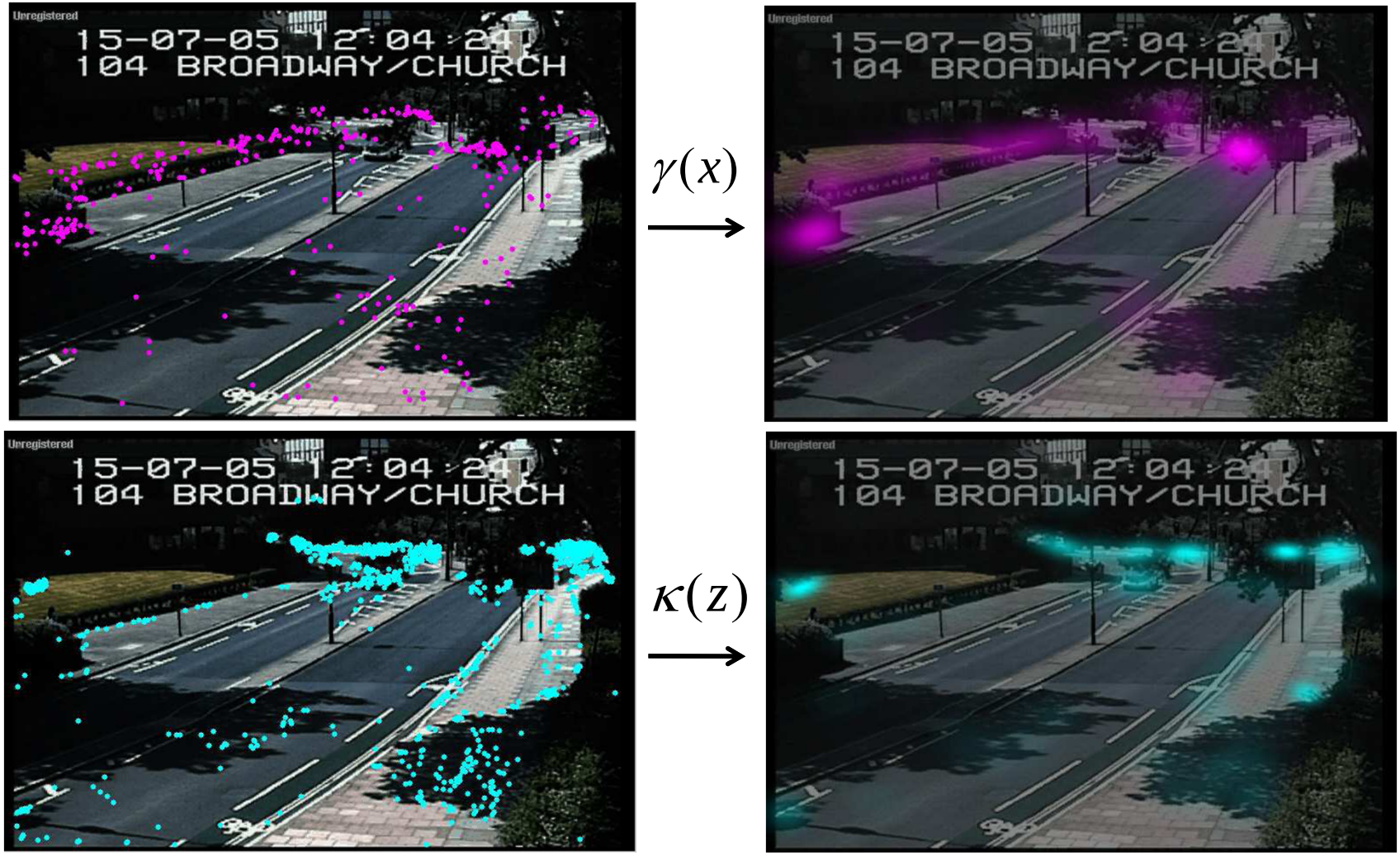


Example: learning birth intensity

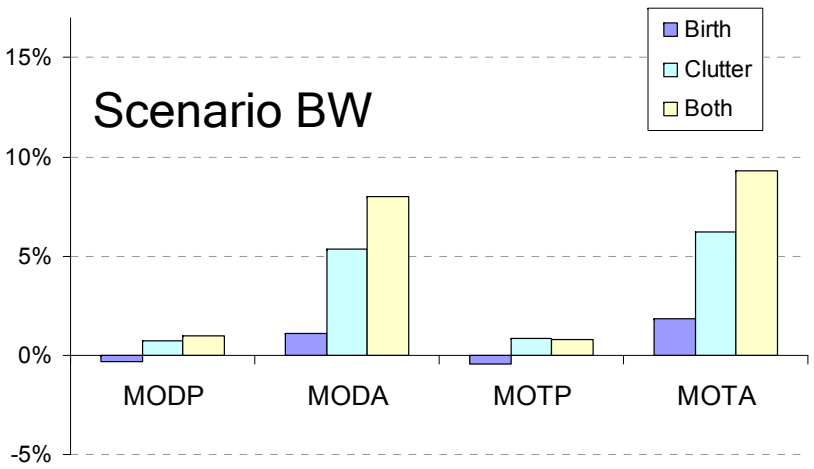


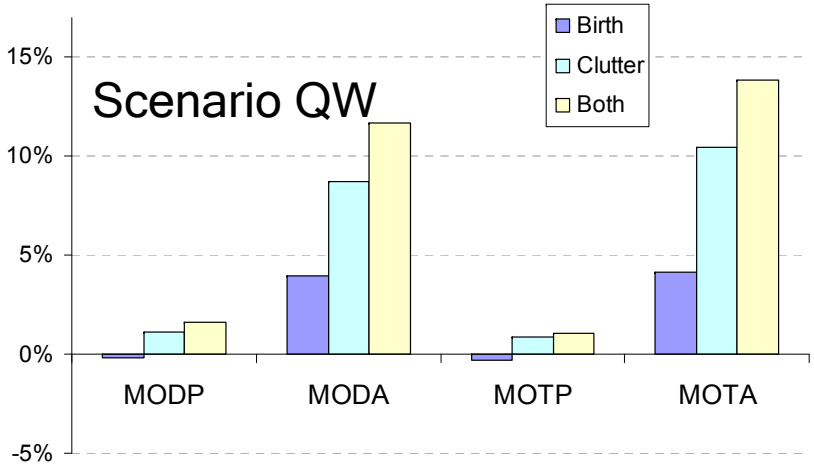


Birth and clutter intensities



Comparison: w. and w/o context learning



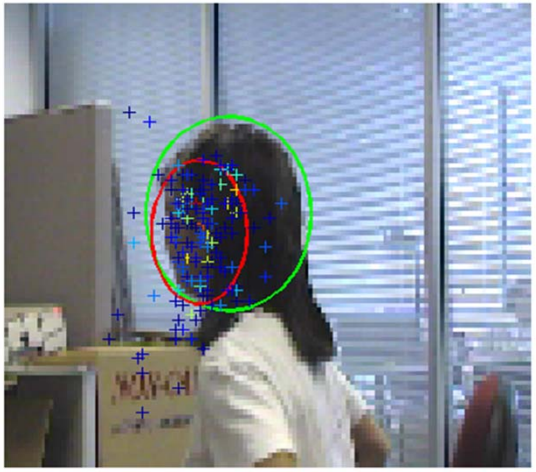


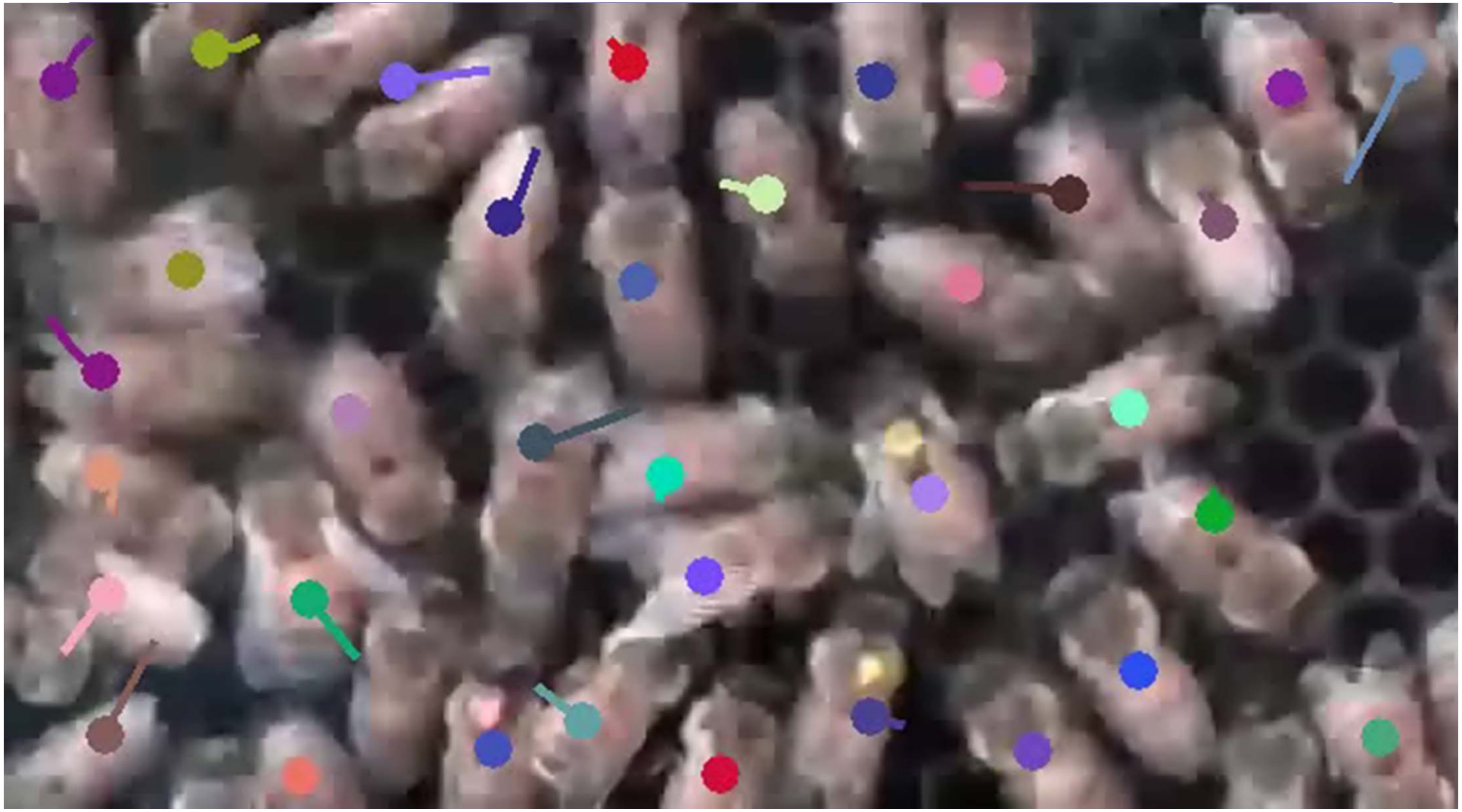
- Performance improvement on CLEAR-2007 dataset
- Reference: PHD tracker with uniform birth and clutter intensities

Learning scene context for multiple object tracking
E. Maggio and A. Cavallaro
IEEE Transactions on Image Processing, Vol. 18, Issue 8, August 2009, pp. 1873-1884

Advances in video tracking: summary

adaptation **self-evaluation** **context learning**





Outlook – projects

EU FP7 project **CENTAUR**
Crowded environments monitoring
for activity understanding and recognition
Marie Curie Industry-Academia Partnerships 2013 - 2016

UK EPSRC project **MAVIP**
Multisource audio-visual production
from user-generated content 2013 - 2015

EU Erasmus Mundus Joint Doctorate **ICE**
Interactive & Cognitive Environments now - 2017

EU Artemis JU project **COPCAMS**
Cognitive & perceptive Cameras
Public-private partnership 2013 – 2016

www.eecs.qmul.ac.uk/~andrea

www.twitter.com/smartcameras

www.youtube.com/smartcameras

Multi-target tracking on confidence maps: an application to people tracking
F. Poiesi, R. Mazzon, A. Cavallaro
Computer Vision and Image Understanding, Vol. 117, Issue 10, pp. 1257-1272, October 2013
http://www.eecs.qmul.ac.uk/~andrea/papers/2013_CVIU_MultiTargetTracking_Poiesi_Mazzon_Cavallaro.pdf

Adaptive appearance modeling for video tracking: survey and evaluation
S. Salti, A. Cavallaro, L. Di Stefano
IEEE Transactions on Image Processing, Vol. 21, Issue 10, October 2012
http://www.eecs.qmul.ac.uk/~andrea/papers/2012_TIP_AdaptiveModeling_Salti_Cavallaro_Distefano.pdf

Adaptive on-line performance evaluation of video trackers
J.C. SanMiguel, A. Cavallaro and J.M. Martinez
IEEE Transactions on Image Processing, Vol. 21, Issue 5, May 2012
http://www.eecs.qmul.ac.uk/~andrea/papers/2012_TIP_AdaptiveOnlineEvaluationTrackers_SanMiguelCavallaroMartinez.pdf

Learning scene context for multiple object tracking
E. Maggio and A. Cavallaro
IEEE Transactions on Image Processing, Vol. 18, Issue 8, August 2009, pp. 1873-1884
http://www.eecs.qmul.ac.uk/~andrea/papers/2009_Maggio_Cavallaro_TIP__learning_context.pdf