A motion-enhanced Hybrid Probability Hypothesis Density filter for real-time Multi-Human Tracking in video surveillance scenarios

Volker Eiselein, Tobias Senst, Ivo Keller, Thomas Sikora

Technische Universität Berlin

Fachgebiet Nachrichtenübertragung

Leitung: Prof. Thomas Sikora

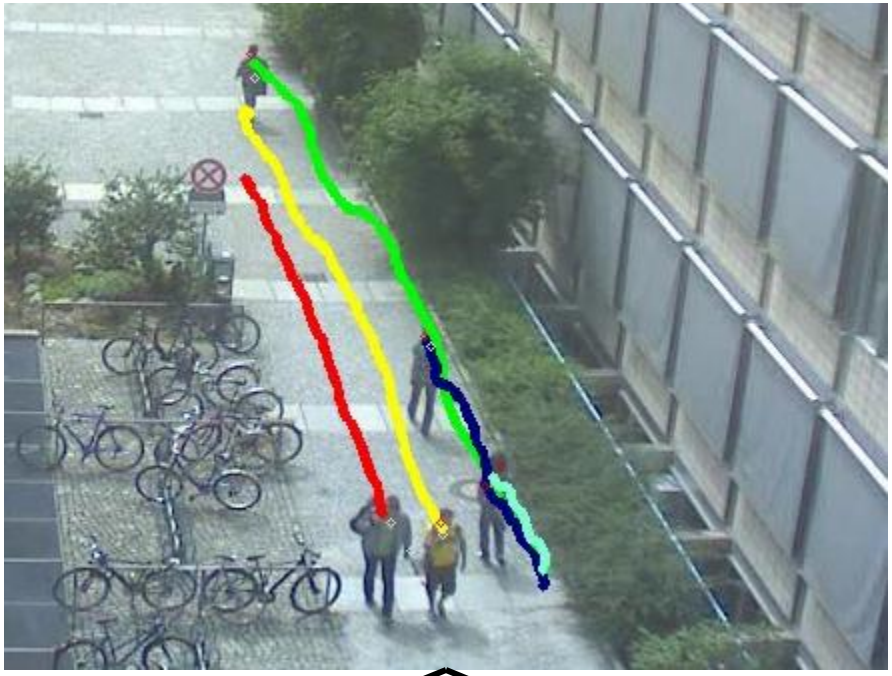
Volker Eiselein

eiselein@nue.tu-berlin.de

VideoSense

2007 - 2013

Multi-Object Tracking



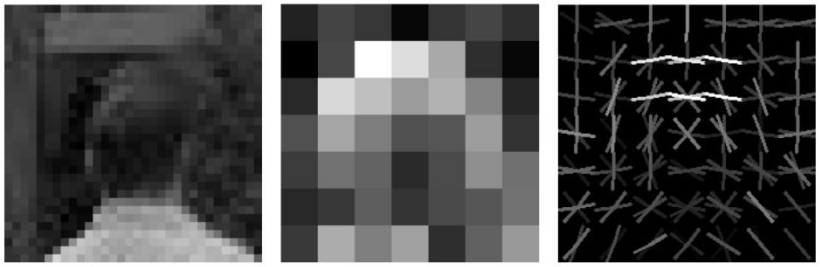
a) Use multiple *single*-object filters

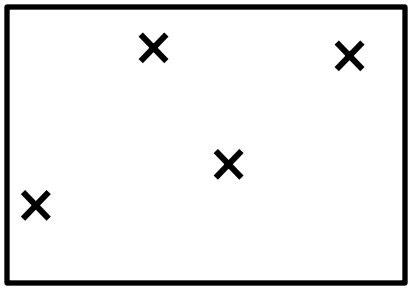
b) Use one *multi*-object filter

Multi-Object Filter

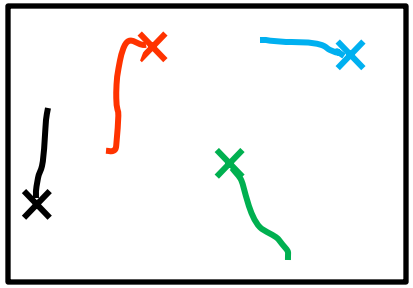
„Tracking-by-Detection“ approach:

1) Get all detections in frame, e.g. by HOG model of shoulder and head:





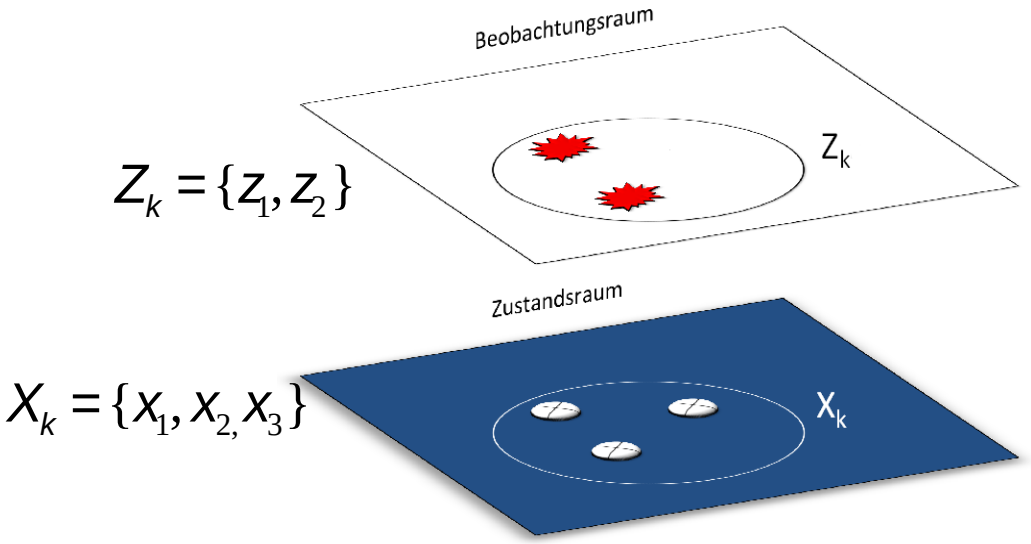
2) Associate these detections to tracks…



Multi-Object Tracking

- *Goal:* Use Bayesian filter as in single-object case

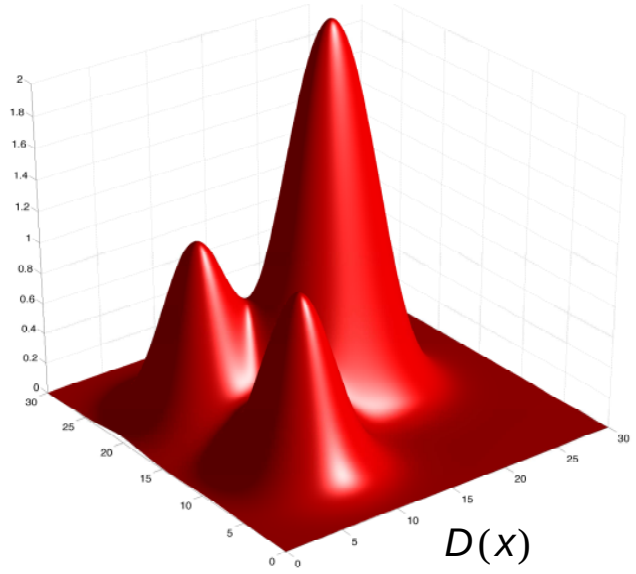
- *Problem:* How to define Bayes‘ theorem for multiple objects?

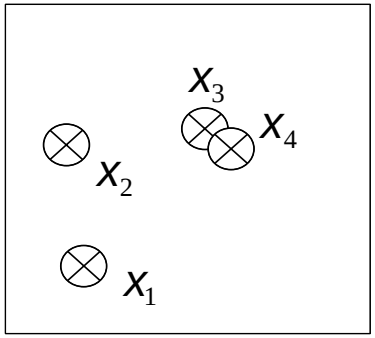


→ Solution: represent objects and measurements as sets

Probability Hypothesis Density (PHD) [Vo et al. 2006]

- For simplification, approximate multi-target statistics by 1st statistical moment (PHD)
- PHD is defined on single state space and gives probability of object existence
- Summed PHD ~ estimate for #objects
- For implementation, Gaussian Mixture Models can be used





Bayes foundation of PHD Filter

Prediction: $$D_{k|k-1}(x_k) = b(x_k) + \sum p_S(x_k \mid x_{k-1}) \cdot f(x_k \mid x_{k-1}) \cdot D_{k-1|k-1}(x_{k-1}) dx_{k-1}$$

Update: $$D_{k|k}(x_k) \approx D_{k|k-1}(x_k) \cdot (1 - p_D(x_k)) + \sum_{z \in Z} \frac{p_D(x_k) \cdot L(z \mid x_k)}{c(z) + \int p_D(x_k) \cdot L(z \mid x_k) \cdot D_{k|k-1}(x_k) dx_k}$$

„No detection“ case

Influence of all received detections

Gaussians of weight > 0.5 are assumed to be objects.

Summary PHD filter

- Associates detections to tracks
- Accounts for „birth“ of new objects
- Accounts for clutter (and missed detections)

Problem formulation

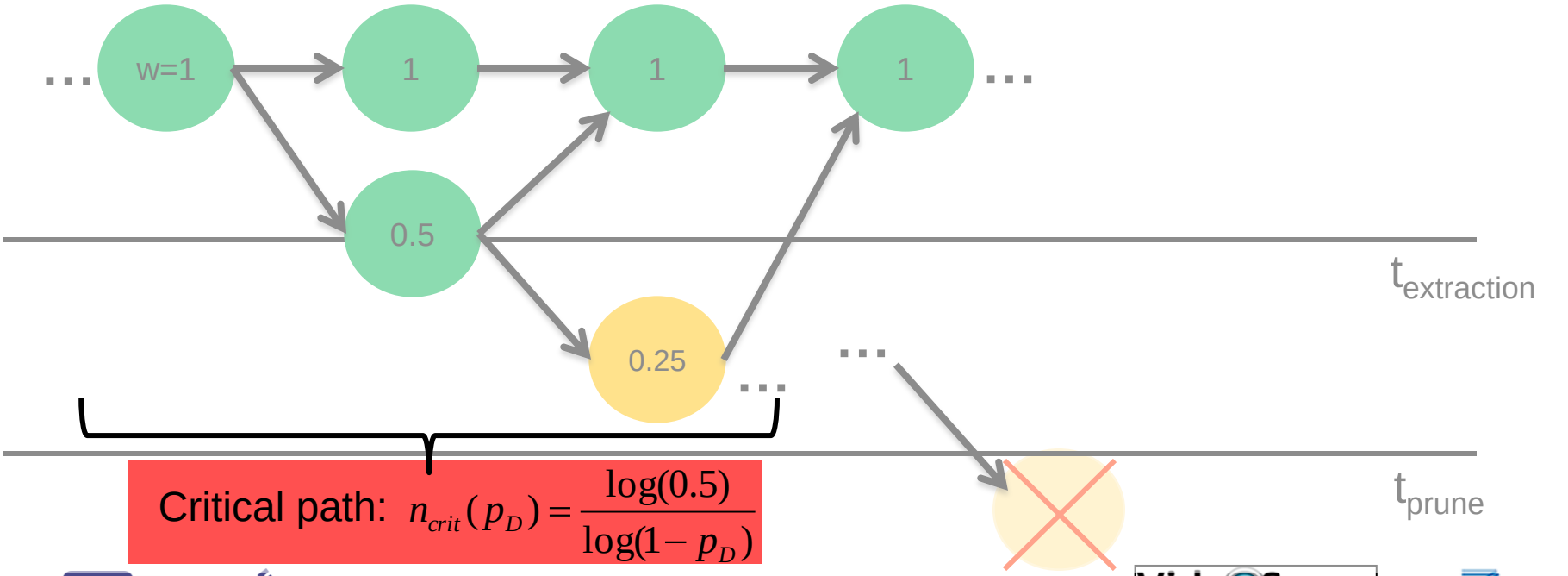
Missed detections can occur due to:

- Low resolution
- Image / Motion Blur
- Occlusions
- Uncommon pose of a person in the scene
- …

Mathematical background

$$D_{k|k}(x_k) \approx D_{k|k-1}(x_k)\cdot(1-p_D(x_k)) + \sum_{z\in Z}\frac{p_D(x_k)\cdot L(z\,|\,x_k)}{c(z)+\int p_D(x_k)\cdot L(z\,|\,x_k)\cdot D_{k|k-1}(x_k)dx_k}$$

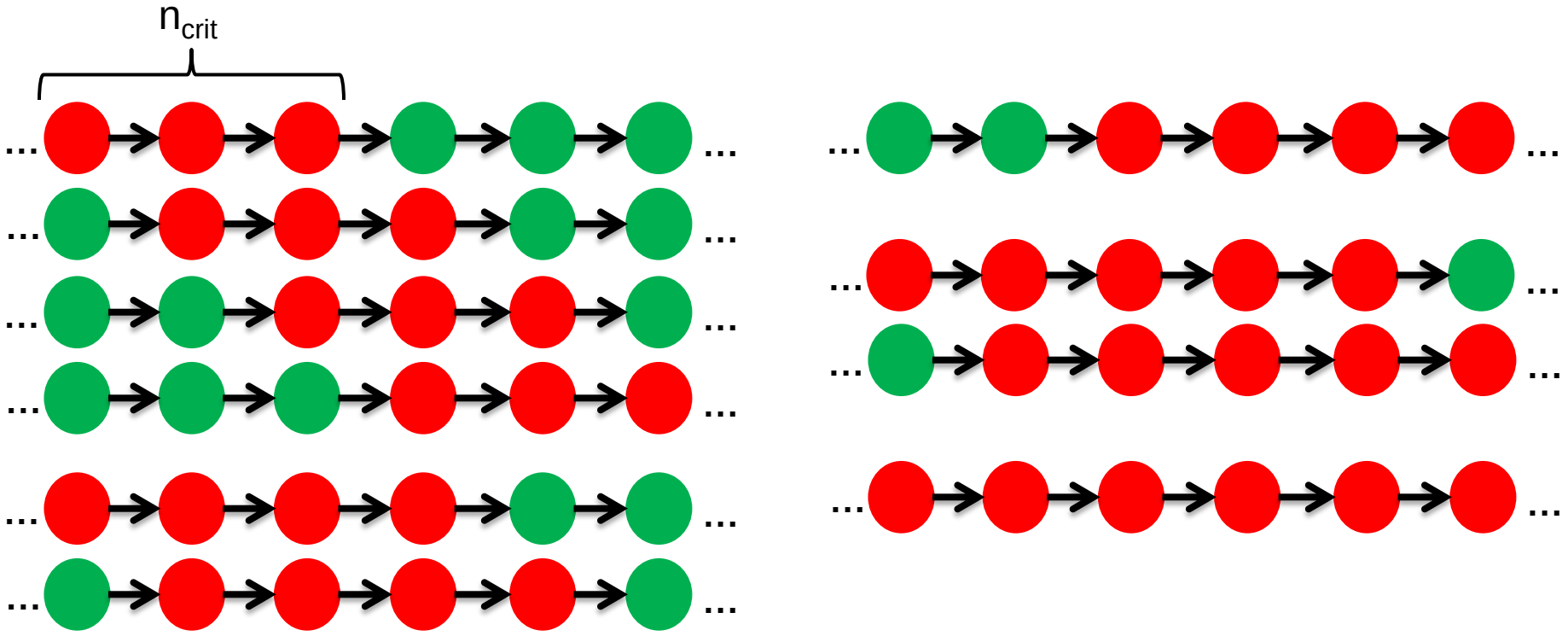
Assume $p_{detection}$=0.5 :



Critical path: $n_{crit}(p_D) = \frac{\log(0.5)}{\log(1-p_D)}$

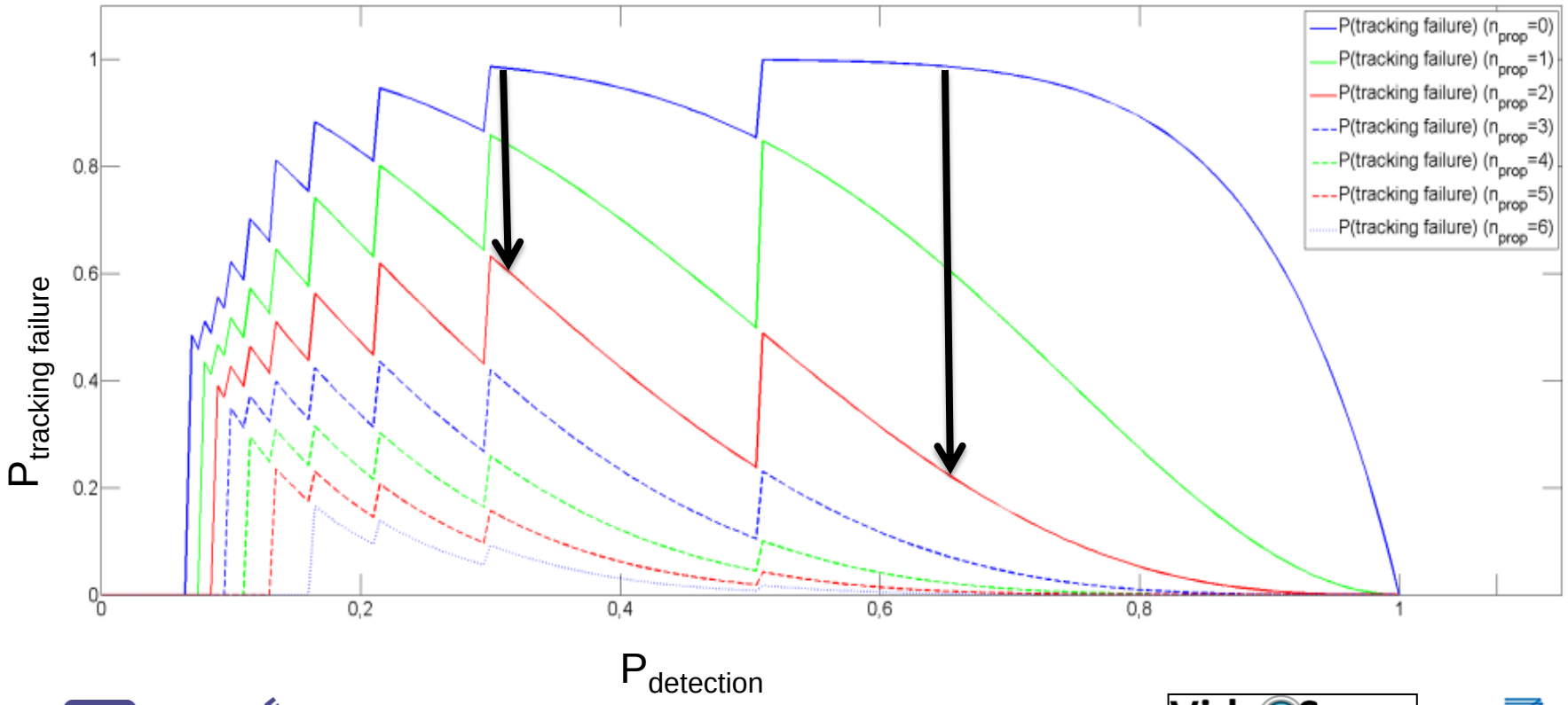
Mathematical background

What is the probability of missing at least n_{crit} detections in a row?



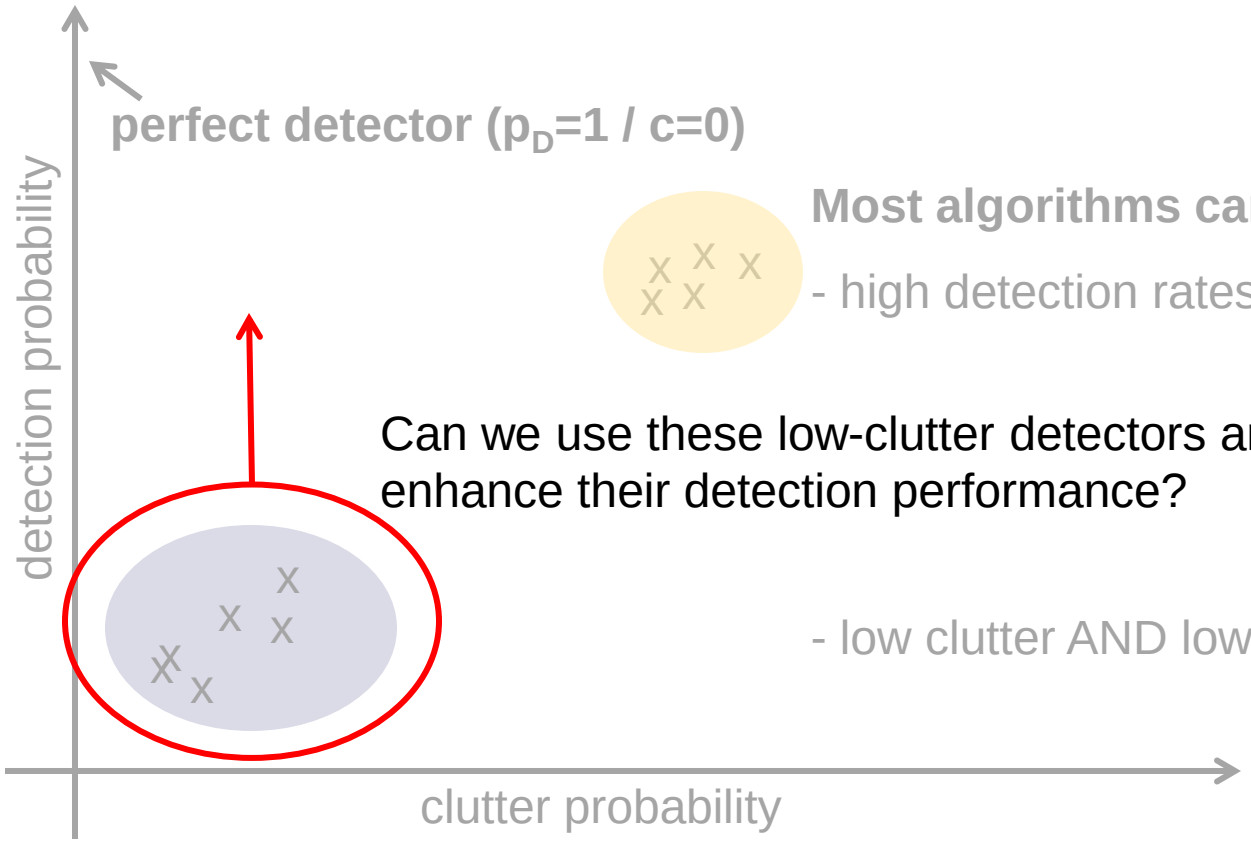
Mathematical background

What is the probability of missing at least n_{crit} detections in a row?



General idea

General perception of human detectors:



Most algorithms can be tuned to either

- high detection rates AND high clutter

Can we use these low-clutter detectors and enhance their detection performance?

- low clutter AND lower detection rates

Approach using motion

Detectors are usually based on:

- gradient information, e.g. HOG

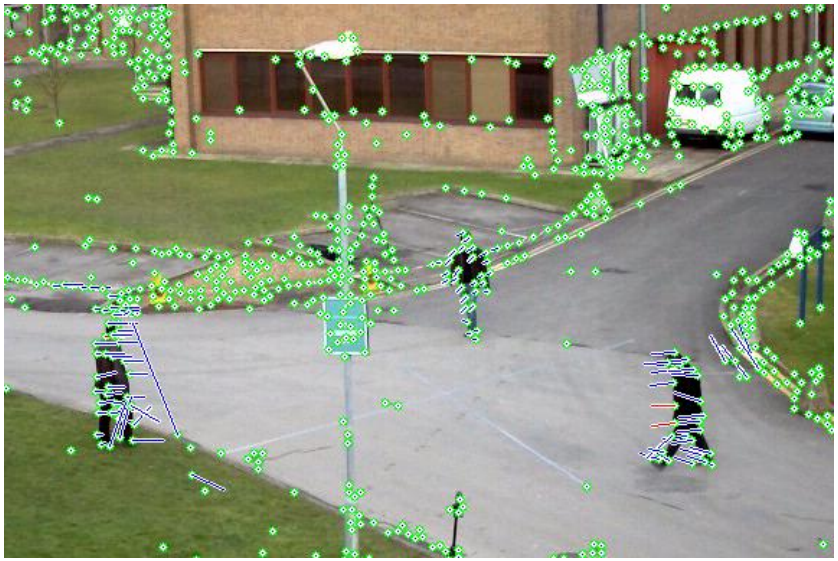
- color information, e.g. target color model / background subtraction

What about motion?

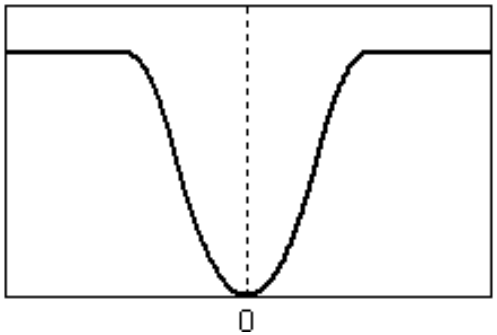
Robust Local Optical Flow

- RLOF – Tracking
  - Robust Local Optical Flow [Senst et. al, Robust Local Optical Flow, TCSVT 2012]
  - Based on modified Hampel Estimator
  - Run time: 9964 motion vectors 36ms (Nvidia GTX 480)

$$E = \sum_{\Omega} \left\| \nabla I \cdot \mathbf{d} + I_t \right\|^2 \Rightarrow \min \quad \Rightarrow \quad E = \sum_{\Omega} \rho\left(\nabla I \cdot \mathbf{d} + I_t, \boldsymbol{\sigma} \right) \Rightarrow \min$$



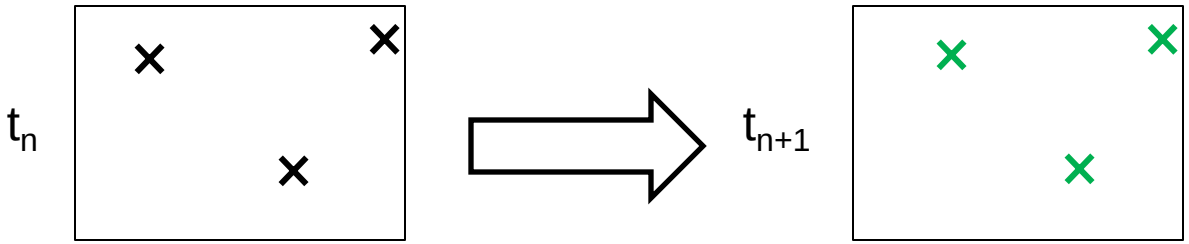
$$\rho(y, \boldsymbol{\sigma}) = \begin{cases} y^2 & , |y| \le \sigma_1 \\ \dfrac{\sigma_1 \left(|y| - \sigma_2 \right)^2}{\sigma_1 - \sigma_2} & , else \\ \sigma_1 \sigma_2 & , |y| \ge \sigma_2 \end{cases}$$



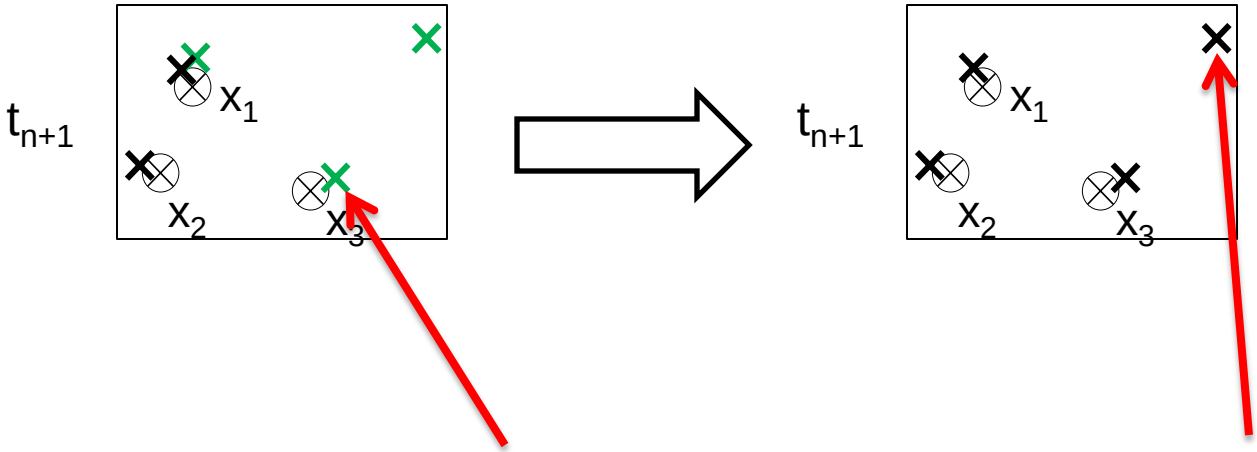
Download at http://www.nue.tu-berlin.de/menue/forschung/projekte/rlof/

Approach using motion

- Use RLOF to propagate detections between several frames:



- Merge into new detection set by gating distance:



Process helps in case of missed detections but also propagates clutter (Trade-off!)

Detailed analysis using OSPA-T metric [Ristic et. al 2011]

Set-based metric between two object sets X,Y:

$$D_{p,c}(X,Y) = \left[\frac{1}{n}\left(\min_{\pi\in\Pi_n}\sum_{i=1}^{m}\left(d_c(x_i, y_{\pi(i)})\right)^p + \left((n-m)\cdot c^p\right)\right)\right]^{\frac{1}{p}}$$

$d(x,y)$ - Base distance between track points and ground truth

$d_c(x,y) = \min(c, d(x,y))$ - Cut-off distance using base distance and maximum penalty c

Π_n - Set of permutations between ground truth and estimated tracks

$1 \le p < \infty$ - metric order

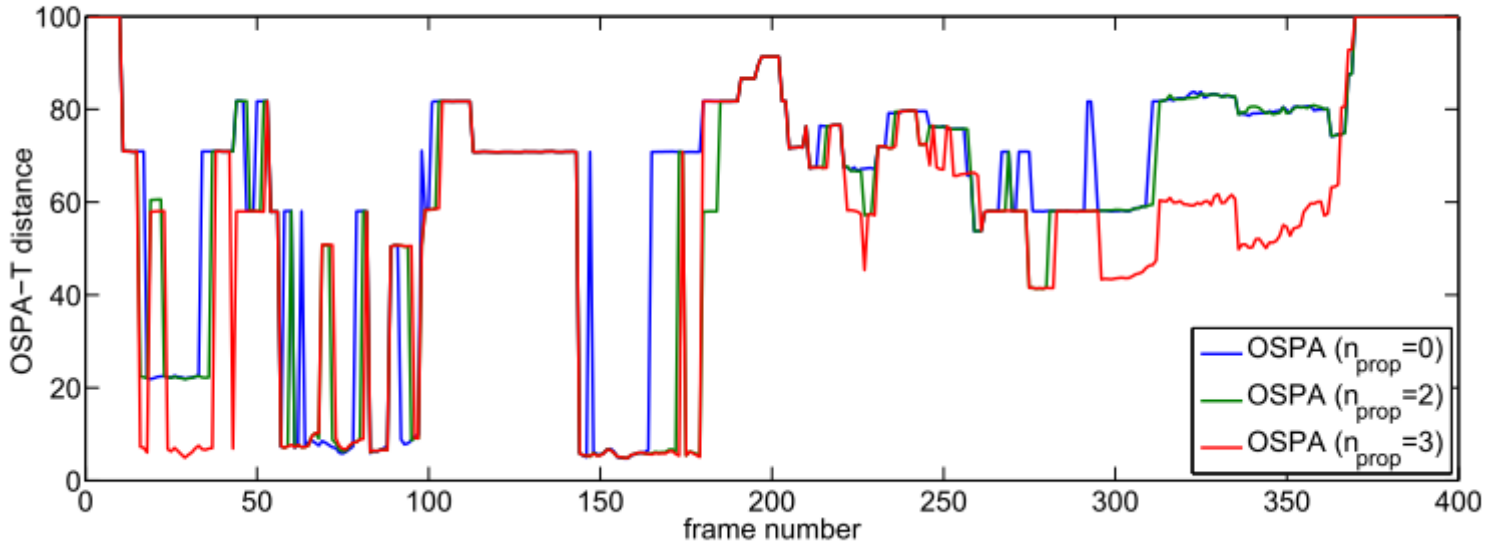
Possible noise introduced by Optical Flow

Potential enhancement by Optical Flow

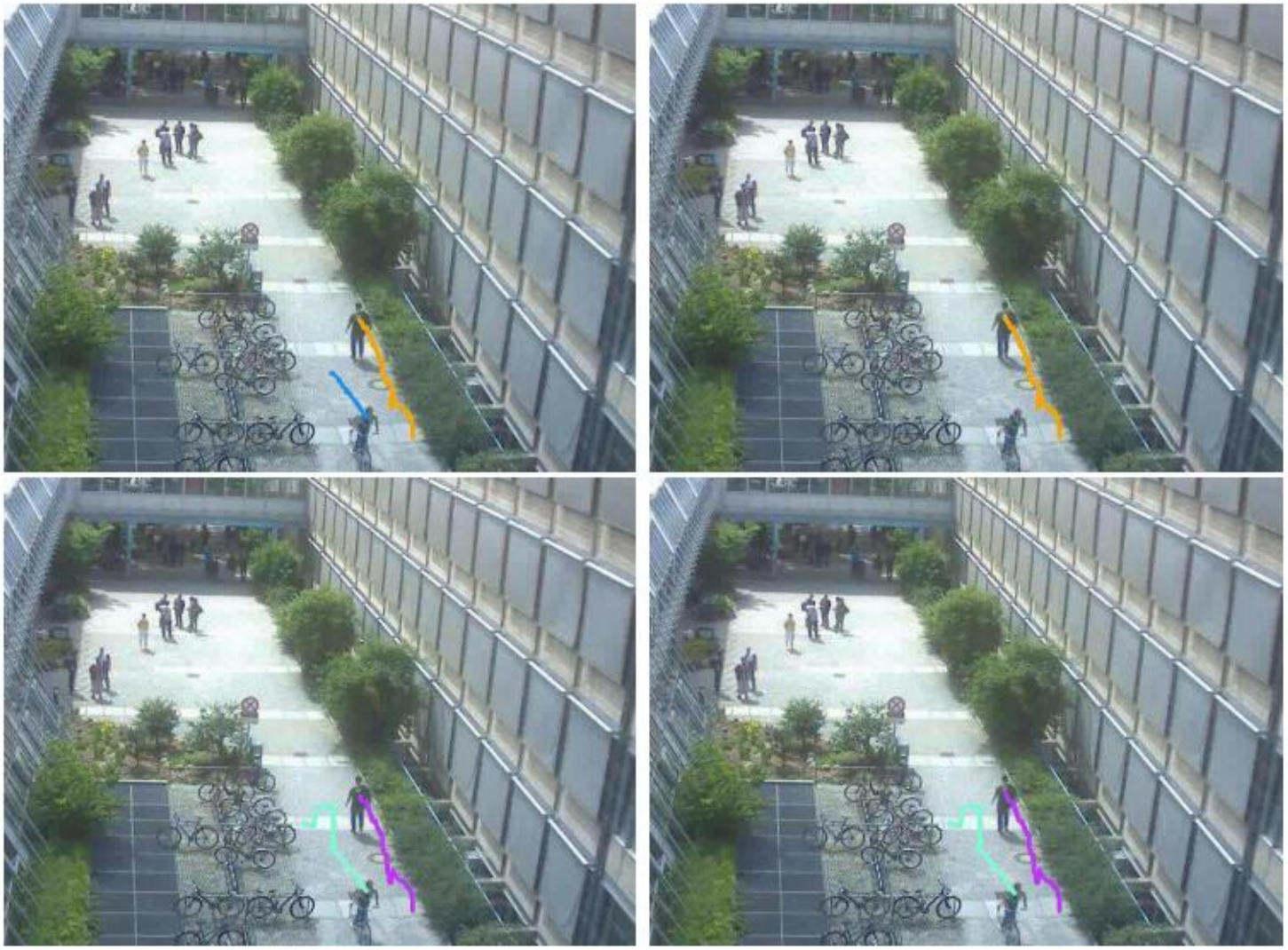
→ Solve for trade-off and optimal number of propagated frames

Results

- PHD filter can be enhanced by Optical Flow information
- Best number of propagated frames depends on p_D and clutter (our detector: n_{prop}~2-3)
- Frame rate can be kept at ~15 fps (10 objects, single threaded C++, not optimized)



Visual results



Exemplary result videos

Thank you - Questions?