Technische Universität Berlin

Volker Eiselein, Daniel Arp, Michael Pätzold and Thomas Sikora

Prof. Dr.-Ing. Sikora

Real-time Multi-Human Tracking using a Probability Hypothesis Density Filter and multiple detectors

Introduction

In multi-object tracking, generally not only the objects' states but also their number is unknown and changes over time. Furthermore, missed detections and clutter cause additional problems.

The Probability Hypothesis Density (PHD) filter proposed by *Mahler, 2006* is a multi-object Bayes filter which has recently attracted a lot of interest in the tracking community mainly for its linear complexity and its ability to deal with high clutter especially in radar/sonar scenarios.

We propose two extensions to this algorithm in order to deal with notably different constraints in computer vision applications. By using a new tree-based path extraction algorithm for a Gaussian Mixture PHD filter and fusing information of two human detectors into a novel likelihood model, we are able to deal better with missed detections and enhance the tracking results significantly.

The Gaussian Mixture PHD filter

Proposed by *Vo et al., 2006* the GM-PHD filter models the PHD function by a mixture of Gaussian distributions and performs tracking in two steps:

- **Prediction Step**: 1) Predict states according to motion model and survival probability, 2) Add birth distributions:

$$D_{k|k-1}(\mathbf{x}) = b(\mathbf{x}) + \int p_S(\mathbf{x'}) \cdot f(\mathbf{x}|\mathbf{x'}) \cdot D_{k-1|k-1}(\mathbf{x'}) d\mathbf{x'}$$

- **Update Step**: Generate new Gaussians for every <detection, state> pair. Compute weights according to:

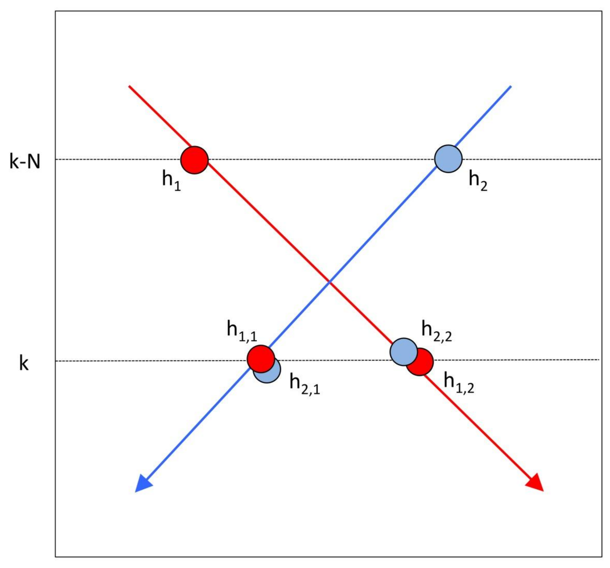
$$w_k^{[j]}(\mathbf{z}) = \begin{cases} (1-p_D)\cdot w_{k|k-1}^{[j]}, & (\text{"no detection"}) \\ \frac{p_D(\mathbf{x})\cdot L_z(\mathbf{x})\cdot w_{k|k-1}^{[j]}}{\mathcal{C}+\int p_D(\mathbf{x})\cdot L_z(\mathbf{x})\cdot w_{k|k-1}^{[j]} d\mathbf{x}}, & \text{if } \mathbf{z}\in Z_k \end{cases}$$

- Labels for every Gaussian allow to identify and track different objects by maintaining a tree structure for all distributions

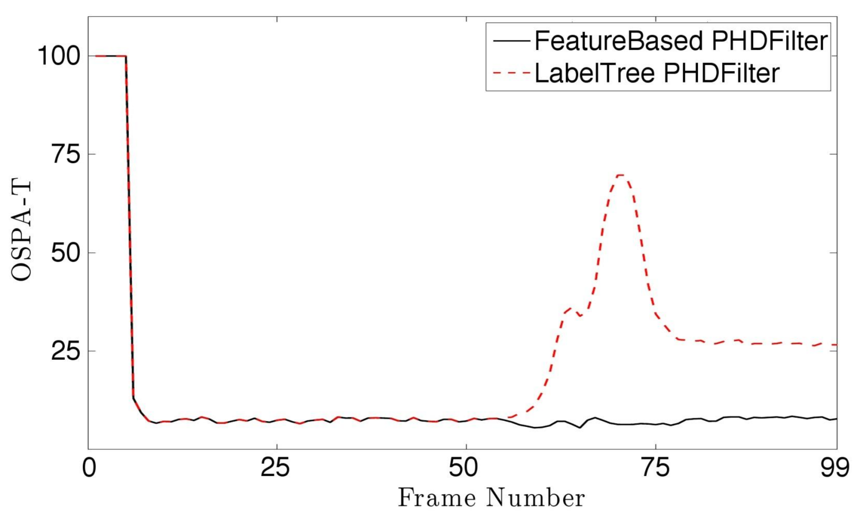
Introducing image information into feature-based label trees

- Filter uses only human detections and cannot distinguish between different persons!
- When two persons meet, many detectors do not maintain both detections, e.g. due to occlusions.
- **Solution**: Use image information to resolve ambiguities.

1. Assign every ID a characteristic image cue (e.g. colour histogram).
2. Use image cues for state extraction of near objects.
3. When objects leave the vicinity area, assign labels and remove wrong label branches.



Label assignment problem in standard GM-PHD filter: After the crossing situation (time step k), two state hypotheses exist for both IDs. For the system, it is not clear which one to choose.



Using the proposed feature-based label tree, the tracking results can be enhanced considerably (OSPA-T distance over 1000 averaged Monte Carlo runs)

Fusing information of two detectors

- In order to reduce the filter's sensitivity towards missed detections, we propose to use two detectors complementing each other:

1. A human detector based on a a-priori learned model of a human's head and shoulders ("head detector").
2. A detector based on background subtraction and connected components ("blob detector").

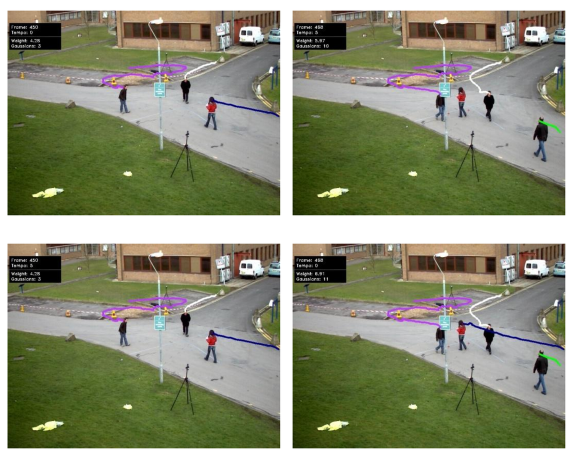
- For multiple detectors, an iterated corrector scheme was proposed in *Mahler, 2006*. This works well with high detection probabilities, but these are not guaranteed in Computer Vision applications.
- The following corrector step is proposed in order to increase the likelihood for single detections in cases of low detection probability:

$$D_{k|k}(\mathbf{x}) = \widehat{L}\cdot D_{k|k-1}(\mathbf{x}) = \Big(\frac{(1-p_D^{[1]})+(1-p_D^{[2]})}{2} + \frac{L_{Z_k^{(1)}}(\mathbf{x})+L_{Z_k^{(2)}}(\mathbf{x})}{2}\Big)\cdot D_{k|k-1}(\mathbf{x})$$

with $$L_{Z_k^{(i)}}(\mathbf{x}) = \sum_{z_j^{(i)}\in Z_k^{(i)}} \frac{p_D^{[i]}\cdot L_{z_j^{(i)}}(\mathbf{x})}{\mathcal{C}+\int p_D^{[i]}(\mathbf{x})\cdot L_{z_j^{(i)}}(\mathbf{x})\cdot D_{k|k-1}(\mathbf{x})d\mathbf{x}}$$

Tracking Results

- Tests on PETS 2009 and on "walk" sequence

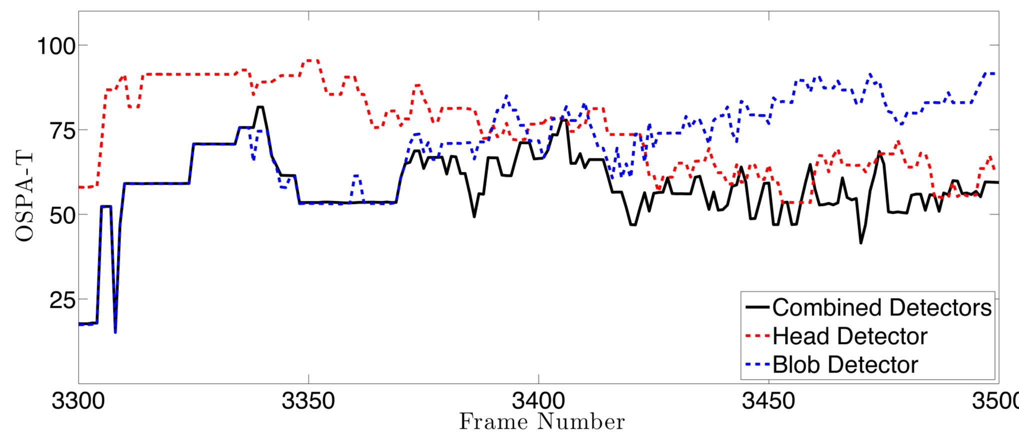


The proposed feature-based label tree extension (below) tracks the lady in the red parka through an occlusion situation while her track is lost otherwise.

OSPA-T distance (PETS)	full video
Head detector	76.63
Blob detector	50.17
Iterated (Blob first)	47.51
Iterated (Head first)	50.75
Proposed method	**38.22**

OSPA-T („walk")	(excerpt)	full video
Head detector	74.6	56,6
Blob detector	70.7	38.5
Iterated (Blob first)	59.9	39.5
Iterated (Head first)	69.9	38.4
Proposed method	**58.2**	**36.5**

Exemplary frame of "walk" sequence. Left: Iterated-corrector approach (head detector first). Right: Proposed method



The proposed combination of detectors enhances the tracking results (OSPA-T measure on excerpt of "walk" sequence)

Conclusion

- Proposed feature-based label trees enhance tracking results of GM-PHD filter in Computer Vision applications.
- A combination of multiple detectors can also enhance tracking results considerably while the proposed novel likelihood model allows its use in scenarios of lower detection probabilities.

Homepage: www.nue.tu-berlin.de
Email: {eiselein,sikora}@nue.tu-berlin.de

VideoSense
2007 - 2013