Intelligent Vehicles that (Fore)See

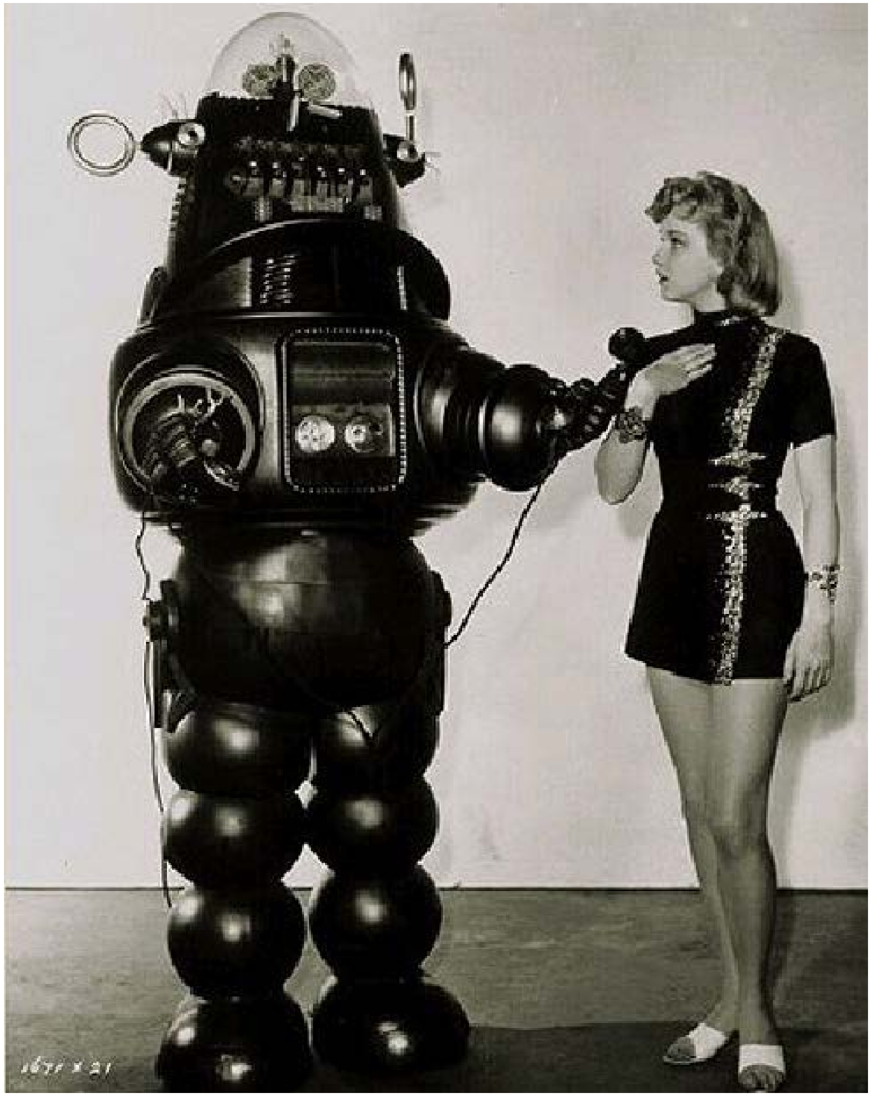
Dariu M. Gavrila

DAIMLER

Environment Perception
Daimler R & D, Ulm, Germany

Intelligent Systems Laboratory
Univ. of Amsterdam, The Netherlands

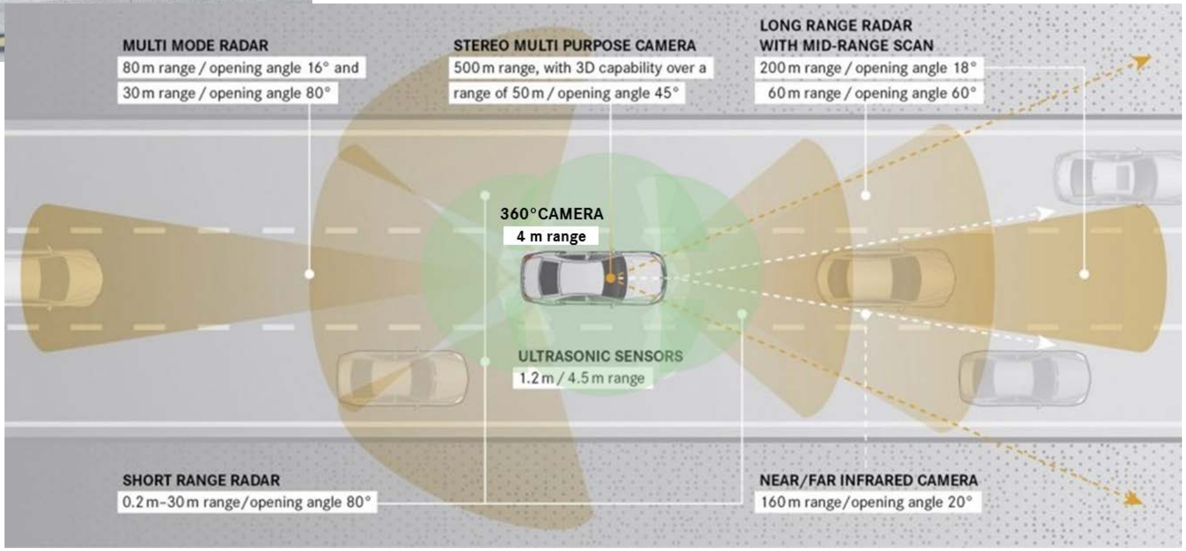
We originally thought Machine Intelligence would look like



1956 "Forbidden Planet"
Robby the Robot (Flickr)

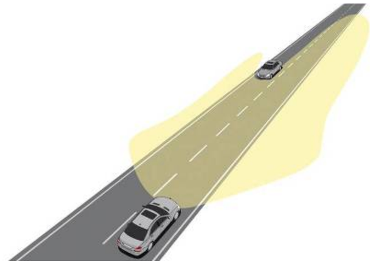
when in fact, Machine Intelligence is already with us, and has a familiar embodiement ...

Mercedes-Benz S Class (2013)

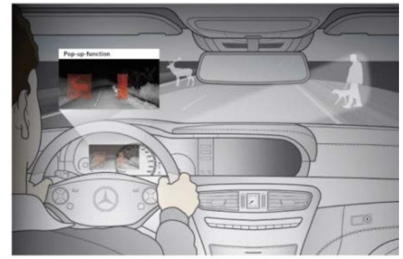


Driver Assistance Functions (MB S-, E- and C-Class, 2013-2014)

Traffic Signs

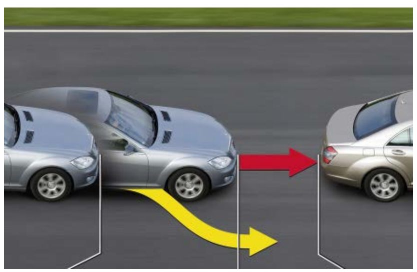


Adaptive High Beam



Nightview

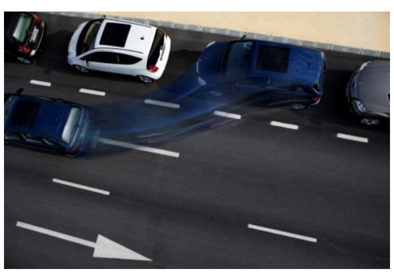
Attention



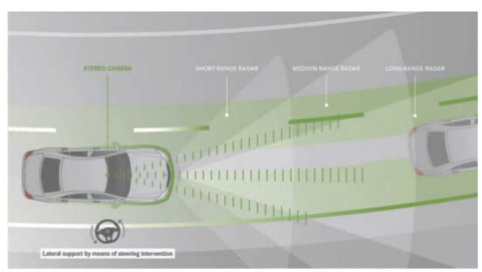
Pre-Crash Braking (longitudinal & lateral traffic) with Pedestrian Recognition

S-Class only

(Active) Body Control



Parking



Adaptive Cruise Control with Steering Assist

(Active)
Lane Keeping

My Research Focus So Far

Perception and modeling of humans and their activities, for systems to interact intelligently with a human-inhabited environment

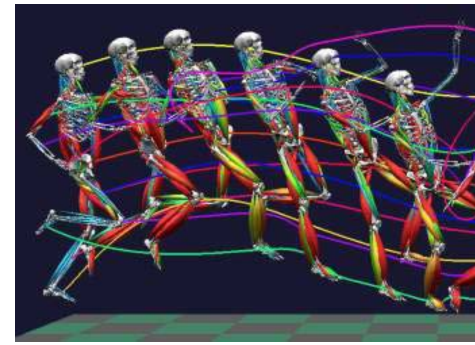
Learn visual appearance, dynamics and behaviors of humans

- High dimensional data, complex manifold structure
- Non-linearity in the observation and dynamical model
- Noisy and/or incomplete sensor data
- Integration of multiple information sources
- Efficient (approximate) inference

Application Domains

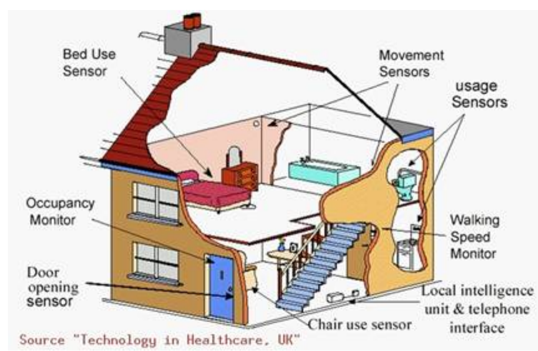
Intelligent Vehicles

Surveillance



Biomechanics
(Rehabilitation, Sports, Ergonomy)

Entertainment
(Animation, Interactive Games)



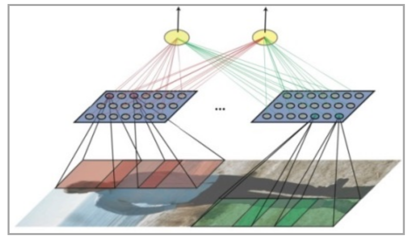
Smart Homes

Social Robotics

Elderly Care

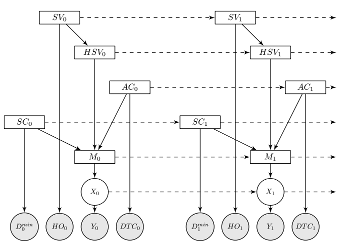
Outline

Regions of Interest (Stereo, Motion, Geometry)



Object Classification

Pose Estimation & Tracking



Behavior Modeling & Path Prediction

Collision Risk Assessment

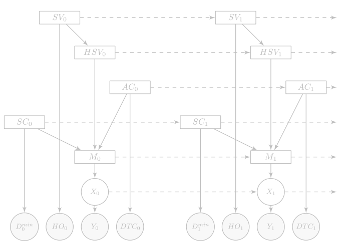
Driver Warning / Vehicle Control

Outline

Regions of Interest
(Stereo, Motion, Geometry)

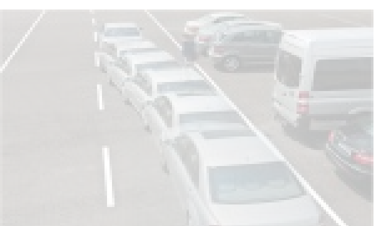
Object Classification

Pose Estimation & Tracking



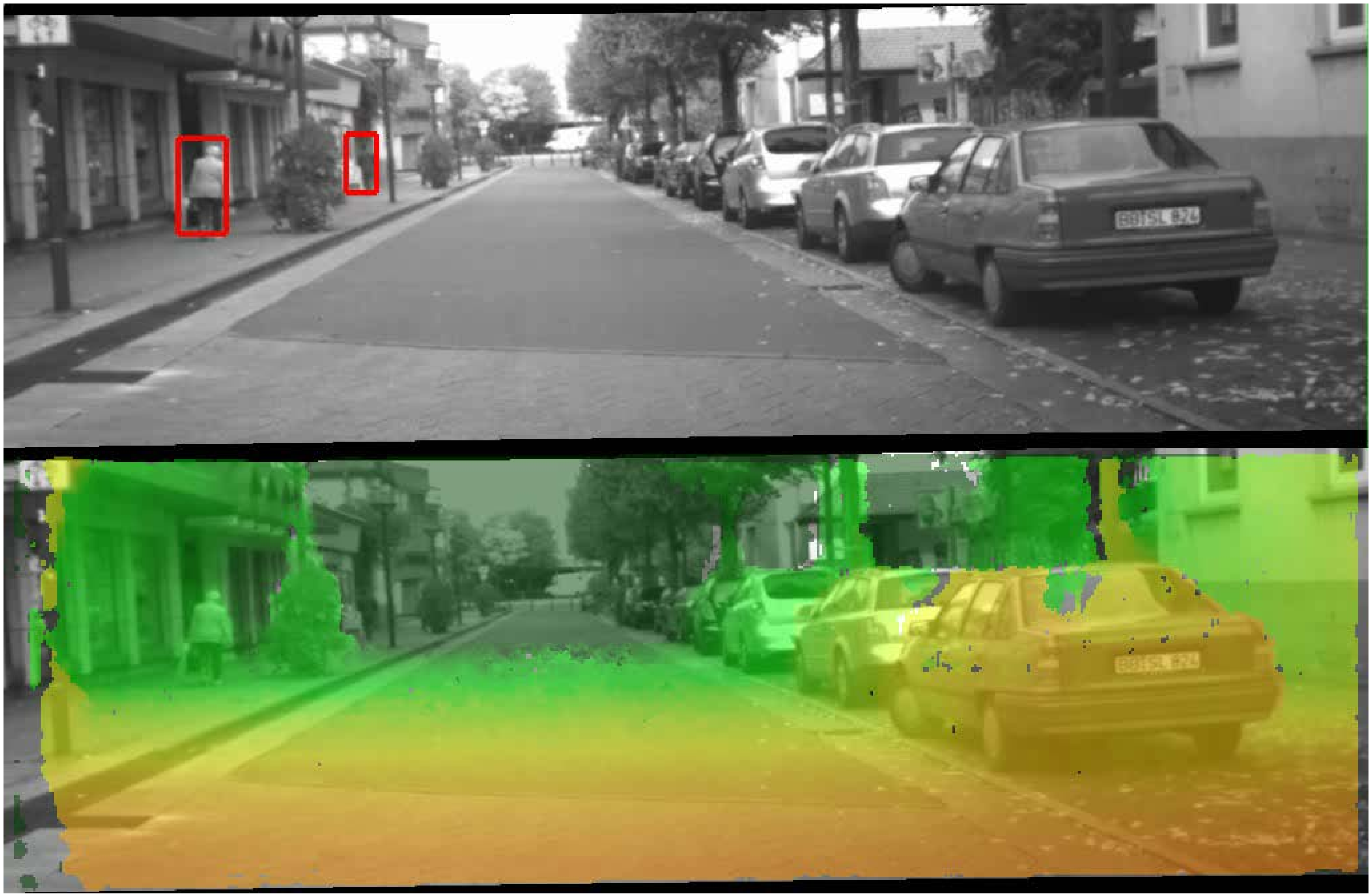
Behavior Modeling & Path Prediction

Collision Risk Assessment



Driver Warning / Vehicle Control

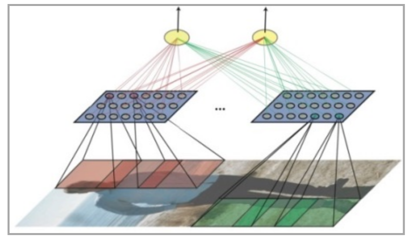
Dense Stereo: Better ROIs, Classification and Localization



C. Keller, M. Enzweiler, M. Rohrbach, D.-F. Llorca, C. Schnörr, and D.M. Gavrila. „The Benefits of Dense Stereo for Pedestrian Detection." *IEEE Trans. on Intelligent Transportation Systems*, 2011.

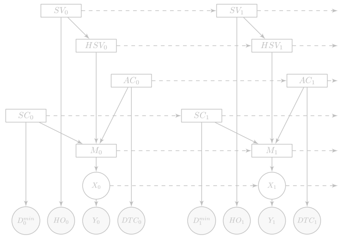
Outline

Regions of Interest
(Stereo, Motion, Geometry)



Object Classification

Pose Estimation & Tracking

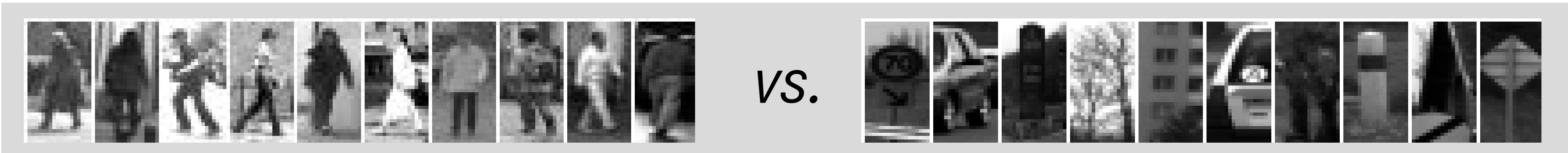


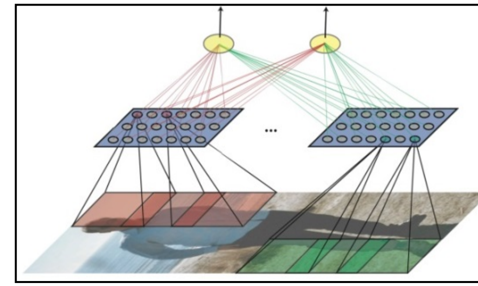
Behavior Modeling & Path Prediction

Collision Risk Assessment

Driver Warning / Vehicle Control

Pedestrian Classification: Appearance-Based

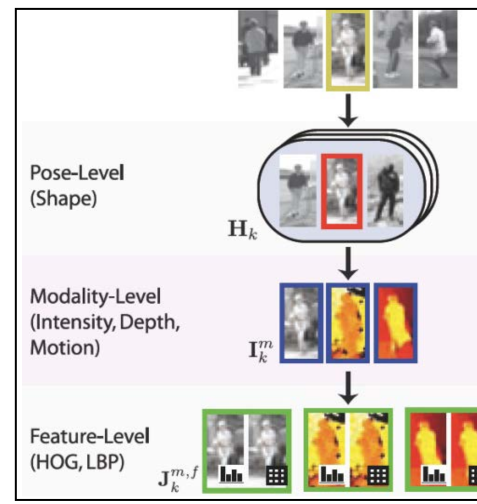




What features? What pattern classifier?

M. Enzweiler and D. M. Gavrila. Monocular Pedestrian Detection: Survey and Experiments. *IEEE Trans. on Pattern Analysis and Machine Intelligence,* 2009.

S. Munder and D. M. Gavrila. An Experimental Study on Pedestrian Classification. *IEEE Trans. on Pattern Analysis and Machine Intelligence,* 2006.

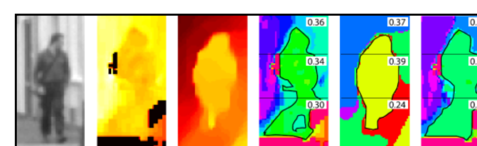


How to combine features and pattern classifiers?

M. Enzweiler and D. M. Gavrila. A Multi-Level Mixture-of-Experts Framework for Pedestrian Classification. *IEEE Trans. on Image Processing,* 2011.

How to deal with occlusion?

M. Enzweiler, A. Eigenstetter, B. Schiele and D.M. Gavrila. Multi-Cue Pedestrian Classification with Partial Occlusion Handling. *Proc. IEEE Conf. on Comp. Vision and Pattern Recognition,* 2010.

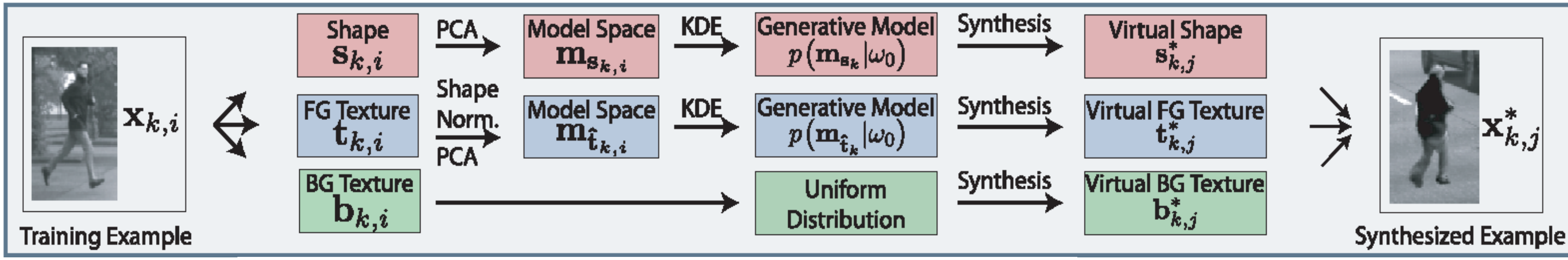


Mixture of experts:

$$P(\omega_0|\mathbf{x}_i) \approx \sum_{k=1}^{K} w_k(\mathbf{x}_i)\mathbf{F}_k(\mathbf{x}_i)$$

pose, modality, body parts, features

Generative Model for Pedestrian Appearance: allows sampling „virtual“ pedestrians

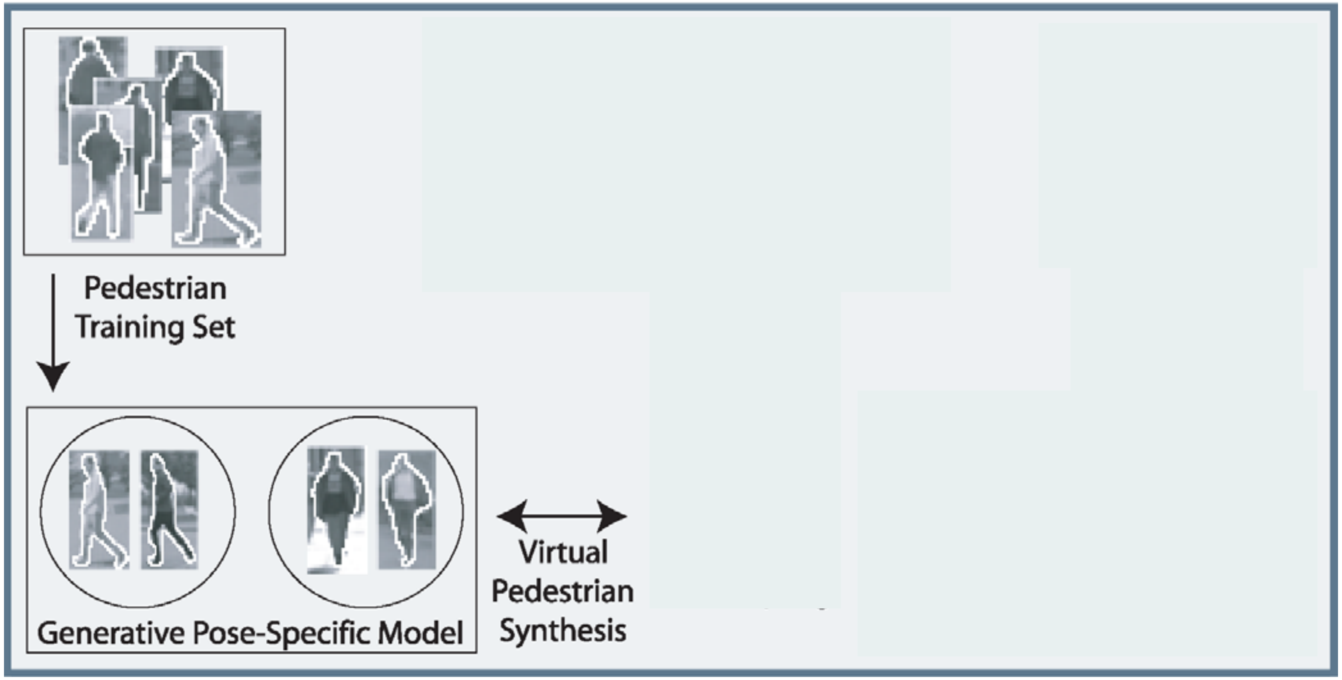


Shape variation

Texture variation

M. Enzweiler and D. M. Gavrila. A Mixed Generative-Discriminative Framework for Pedestrian Classification. *Proc. CVPR* 2008.

Mixed Generative-Discriminative Classification Framework

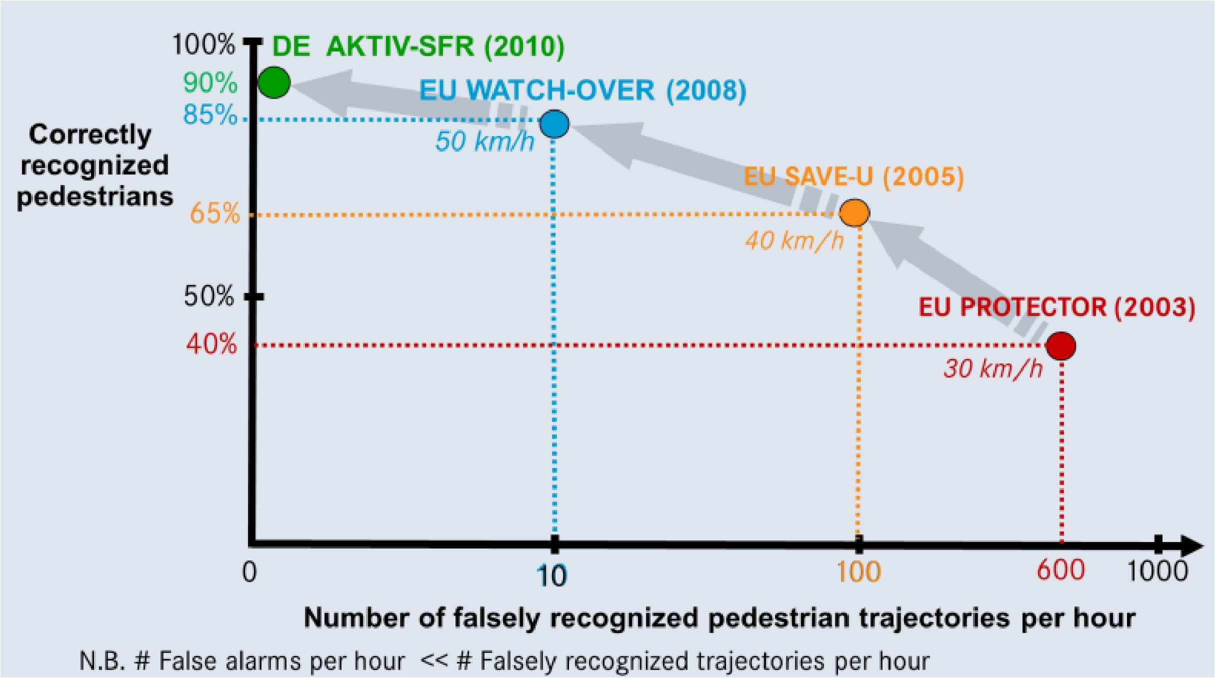


In experiments, 30% less false positives (outperforming case where an equal number of manual labels were added)

Not all virtual samples are informative:
Rejection sampling by active learning

M. Enzweiler and D. M. Gavrila. A Mixed Generative-Discriminative Framework for Pedestrian Classification. *Proc. CVPR* 2008.

Pedestrian Recognition: from Research to Product



After a decade of research …

S Class (2013)

E Class (2013)

C Class (2014)

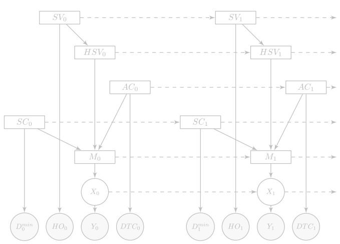
2013-2014: Market introduction PRE-SAFE® brake with stereo vision-based pedestrian recognition in Mercedes-Benz models

Outline

Regions of Interest
(Stereo, Motion, Geometry)

Object
Classification

**Pose Estimation
& Tracking**



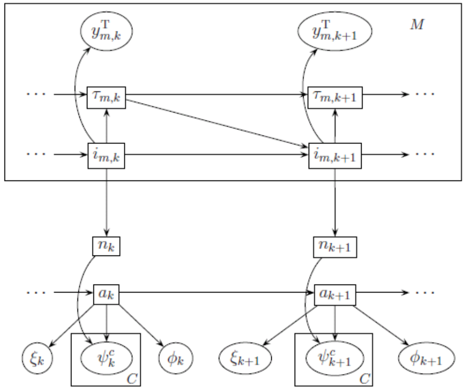
Behavior Modeling
& Path Prediction

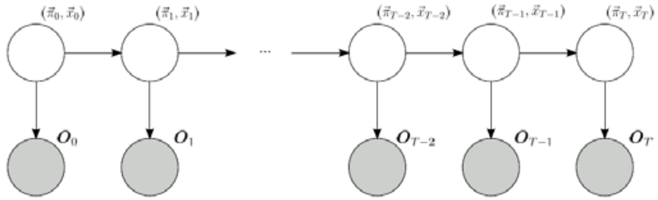
Collision Risk
Assessment

Driver Warning /
Vehicle Control

Probabilistic Temporal Models – Directed Graphs

Conditional dependencies between variables denoted by links



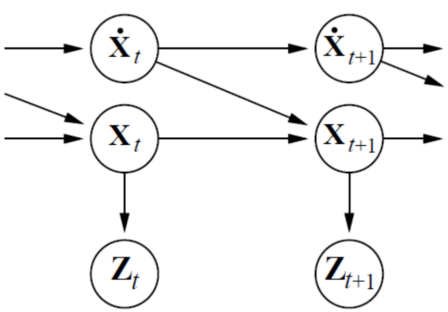


Dynamic Bayesian Networks (DBN)

contains arbitrary many latent and observed variables (discrete or continuous), which are replicated at each time slice

Hidden Markov Model (HMM)

DBN with a single, discrete latent variable and a single observed variable



Linear Dynamical Systems (LDS)

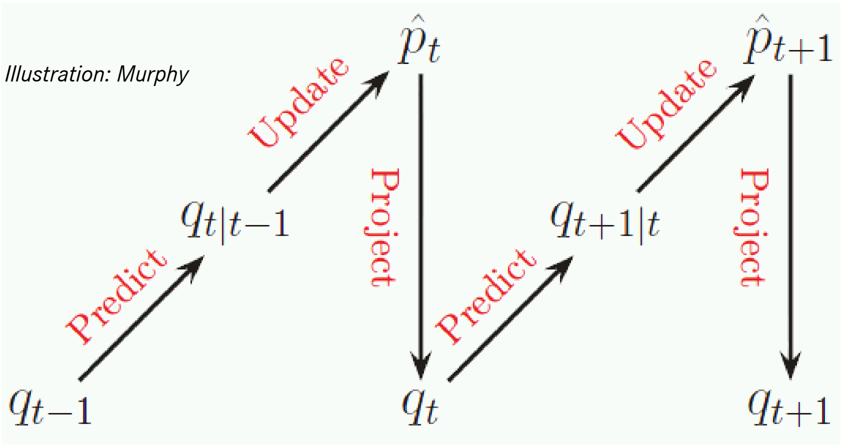
DBN with continuous variables,
linear Gaussian dynamical and observation model

Inference – Bayesian Filtering

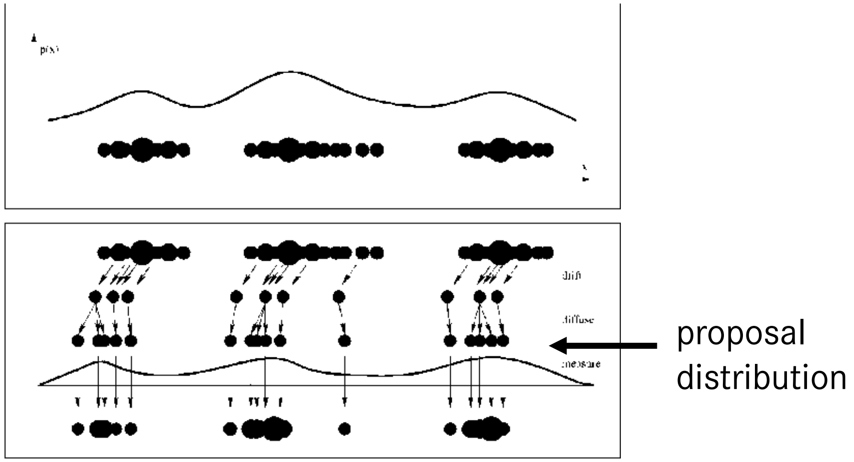
Few interesting problems allow exact inference (e.g. Kalman Filter for LDS)

Approximate inference techniques

1. Exact inference on approximate model → parametric approach: Assumed Density Filter (e.g. Boyen Koller, GPB-1, GPB-2, IMM)
2. Approximate inference on exact model → non-parametric approach: Particle Filter



Assumed Density Filter

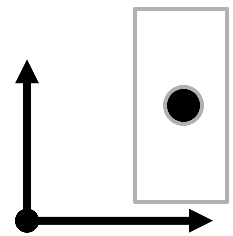


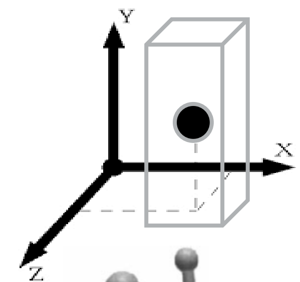
Particle Filter

Pose/Tracking - Related State Variables

2D **3D**

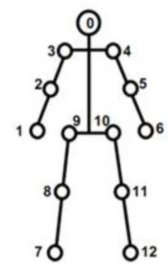
- Location $\boldsymbol{L}_t$

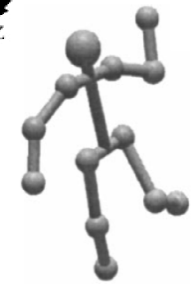




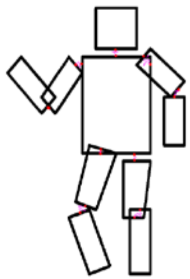
Trade-off between modeling accuracy and ability to estimate parameters robustly

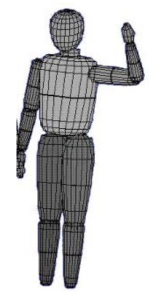
- Pose $\boldsymbol{P}_t$





- Shape $\boldsymbol{S}_t$





- Texture $\boldsymbol{T}_t$

Pose/Tracking - Related State Variables

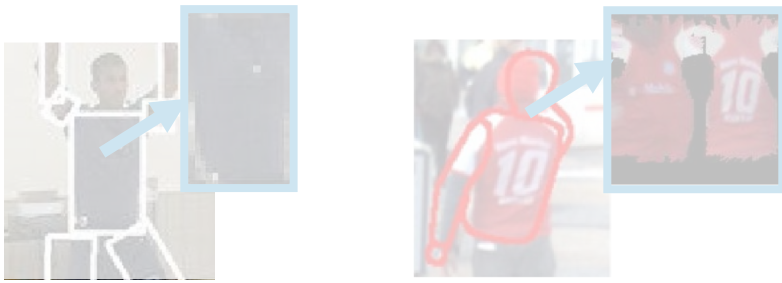
- Location L_t

Particle filtering

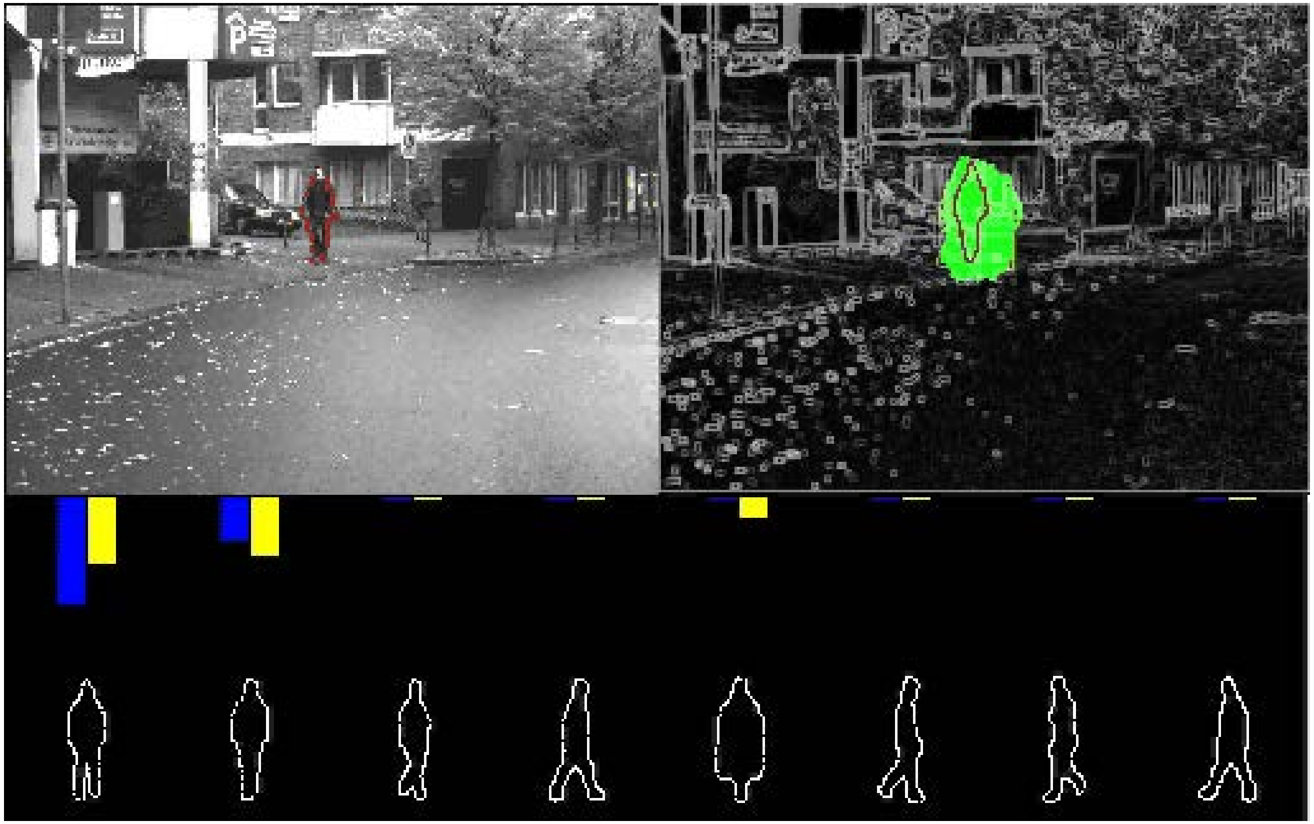
- *Partitioned sampling*
- *Mixture of predictive and detector distribution*
- *Particle optimization*
- *„Coarse to fine“ likelihood function*

Trade-off be
modeling ac
and ab
es
parameters
robustly

- Texture T_t



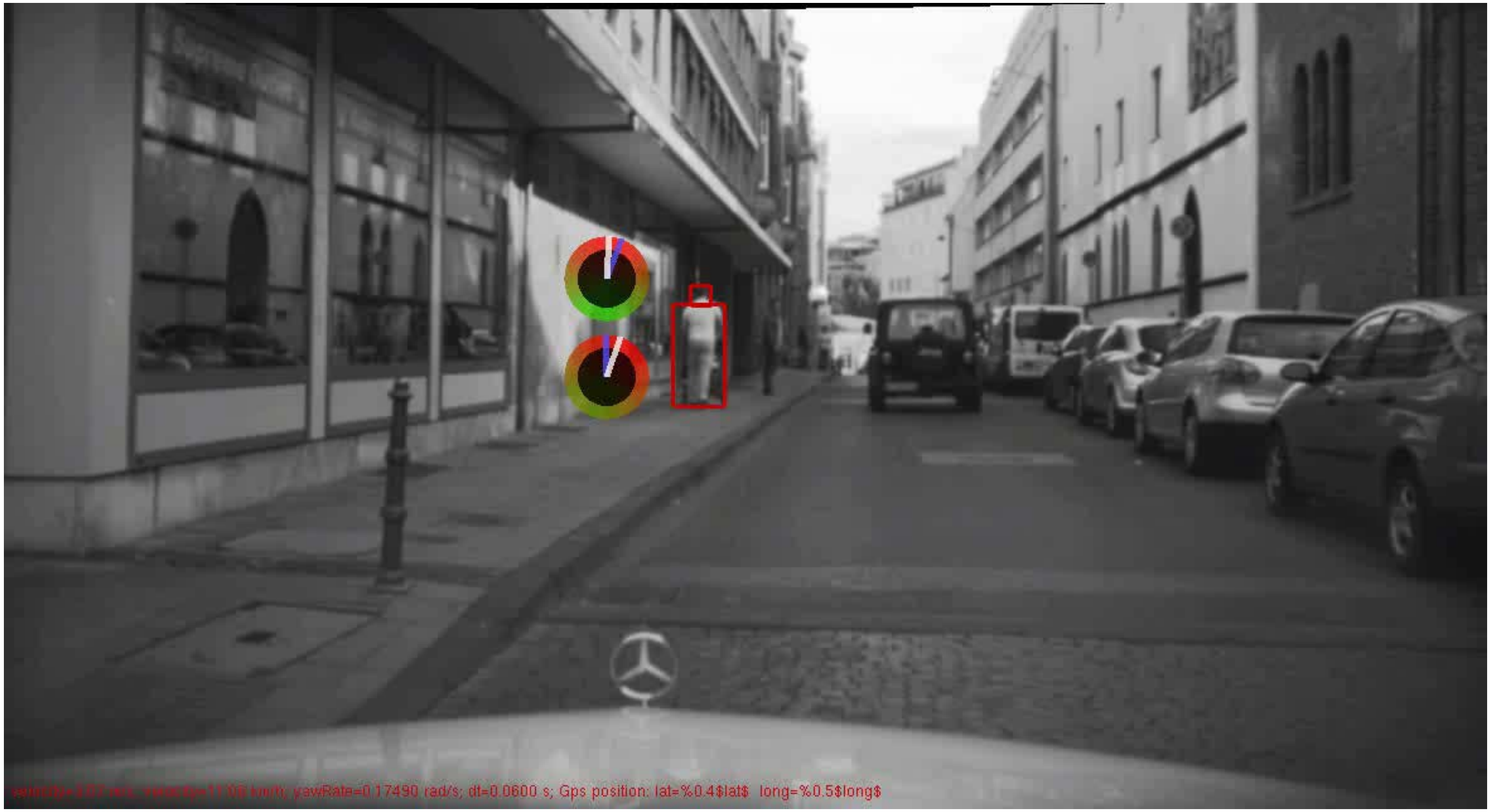
Estimation of (Deformable) 2D Pedestrian Appearance



S. Munder, C. Schnörr and D. M. Gavrila, Pedestrian Detection and Tracking Using a Mixture of View-Based Shape-Texture Models. *IEEE Trans. on Intelligent Transportation Systems*, 2008.

- Partitioned sampling $\boldsymbol{L}_t \rightarrow \boldsymbol{S}_t \rightarrow \boldsymbol{T}_t$
  ($\boldsymbol{S}_t$ has mixed discrete/continuous representation, 20+ dim, implictly encodes $\boldsymbol{P}_t$)
- Proposal distribution: mixture of predictive distribution and detector distrib.
- Particle optimization (Active Shape Models [Cootes *et al*., 1995])

Estimation of 3D Head- and Body- Orientation



(results are only shown for one pedestrian at a time)

- $\boldsymbol{L}_t$ → $\boldsymbol{P}_t$. Low-dimensional state space $\boldsymbol{P}_t\,(\varphi_{\mathrm{Head}}, \varphi_{\mathrm{Body}})$ → no partitioned sampling, no particle optimization necessary
- Proposal distribution equals predictive distribution

F. Flohr, M. Dumitru-Guzu, J. F. P. Kooij and D. M. Gavrila. A probabilistic framework for joint pedestrian head and body orientation estimation. *IEEE Trans. on Intelligent Transportation Systems,* 2014

Estimation of 3D Articulated Human Pose

Cluttered, dynamic background, arbitrary (single) human motion, normal clothing, few overlapping cameras

M. Hofmann and D. M. Gavrila. Multi-view 3D Human Pose Estimation in Complex Environment. *Int. Journal of Computer Vision,* 2012.

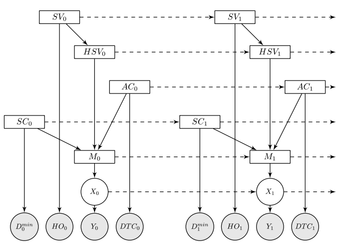
- $\boldsymbol{L}_t \rightarrow \boldsymbol{P}_t \rightarrow \boldsymbol{T}$ ($\boldsymbol{P}_t$ has 13 DOF, $\boldsymbol{S}_t$ estimated by off-line process)
- Coarse-to-fine likelihood function
- Proposal distribution: mixture of extrapolations of K-most likely solutions within time interval (batch mode) and detector distribution
- Particle optimization (inverse kinematics)

Outline

Regions of Interest
(Stereo, Motion, Geometry)

Object
Classification

Pose Estimation
& Tracking



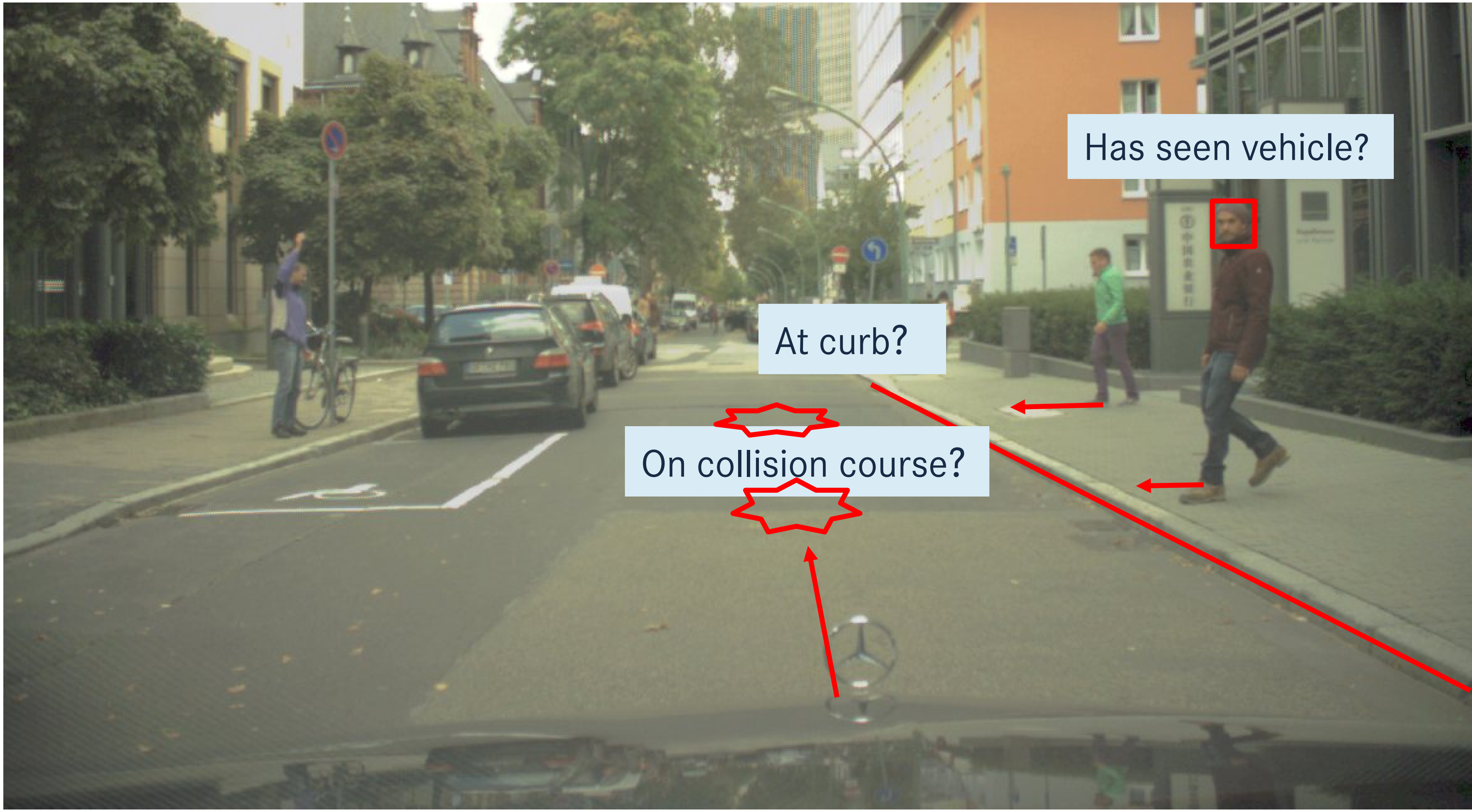
Behavior Modeling
& Path Prediction

Collision Risk
Assessment

Driver Warning /
Vehicle Control

Pro-active pedestrian safety: will the pedestrian cross?

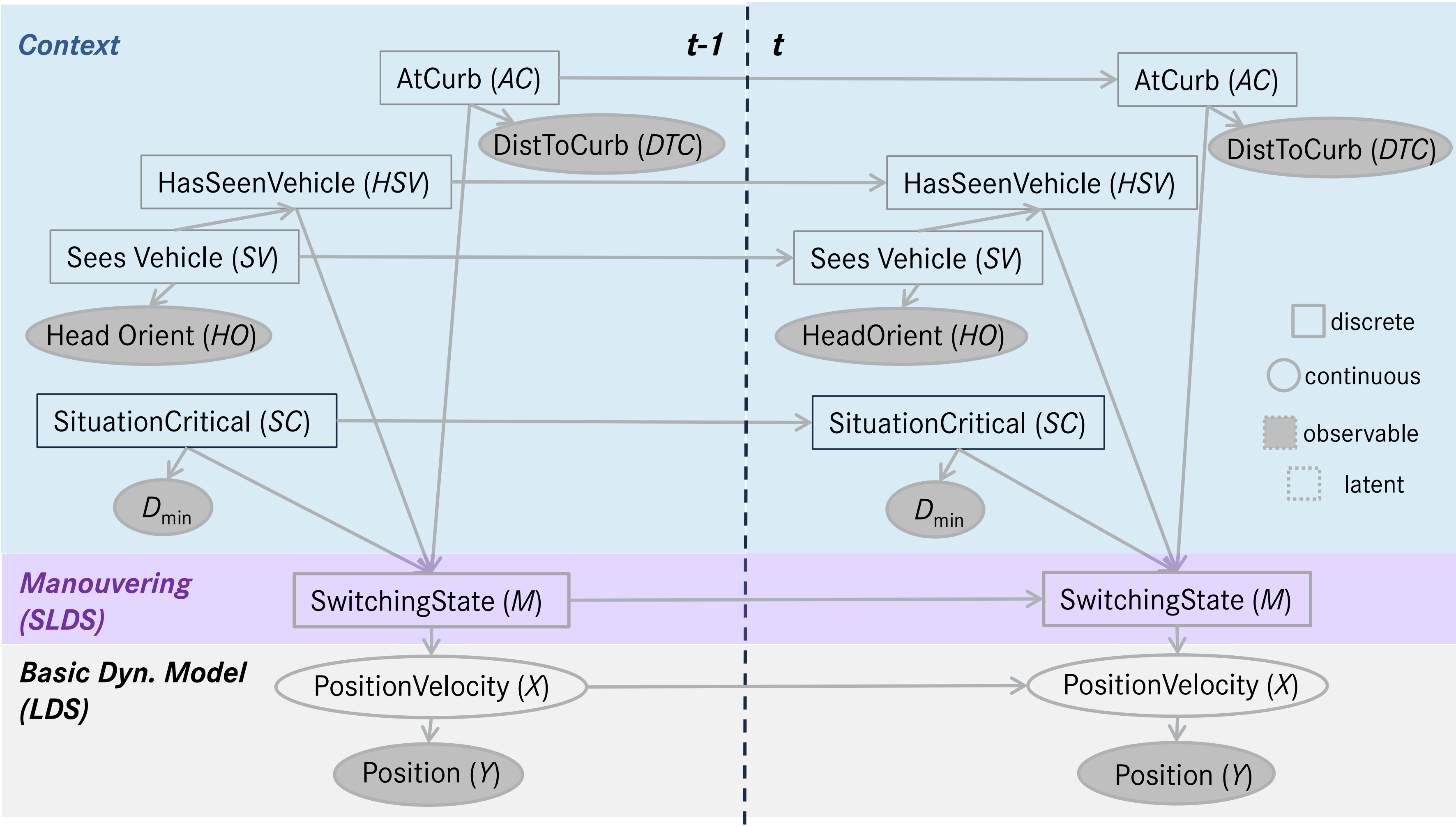
Context Matters



Context-based Switching Linear Dynamical System (SLDS)

Dynamical Bayesian Network

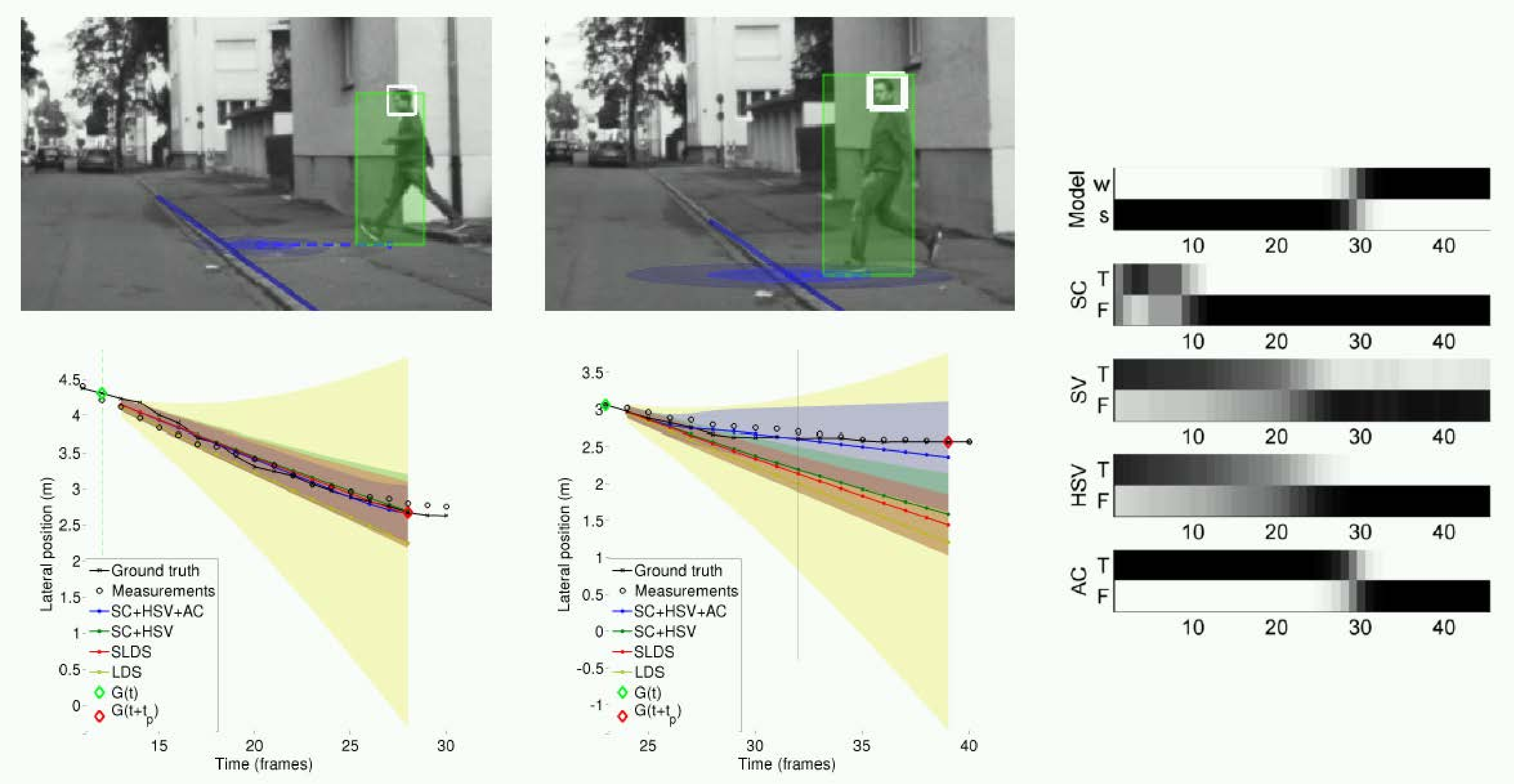
Inference by Assumed Density Filtering



J. P. F. Kooij, N. Schneider, F. Flohr and D. M. Gavrila. Context-based pedestrian path prediction. *Proc. ECCV*, 2014.

State Estimation & Path Prediction

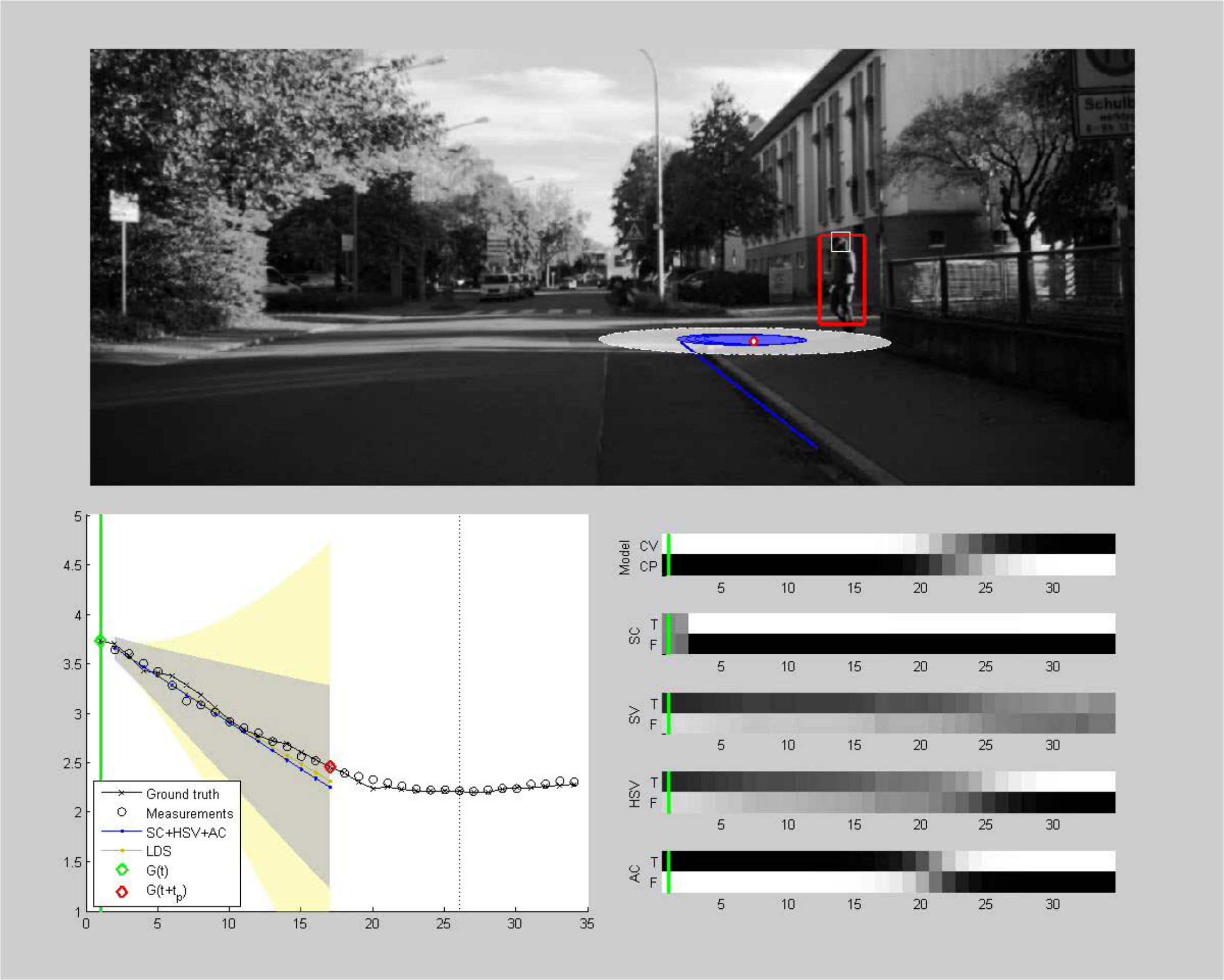
Scenario: pedestrian sees vehicle and stops (two snapshots of a run)



J. P. F. Kooij, N. Schneider, F. Flohr and D. M. Gavrila. Context-based pedestrian path prediction. *Proc. ECCV*, 2014.

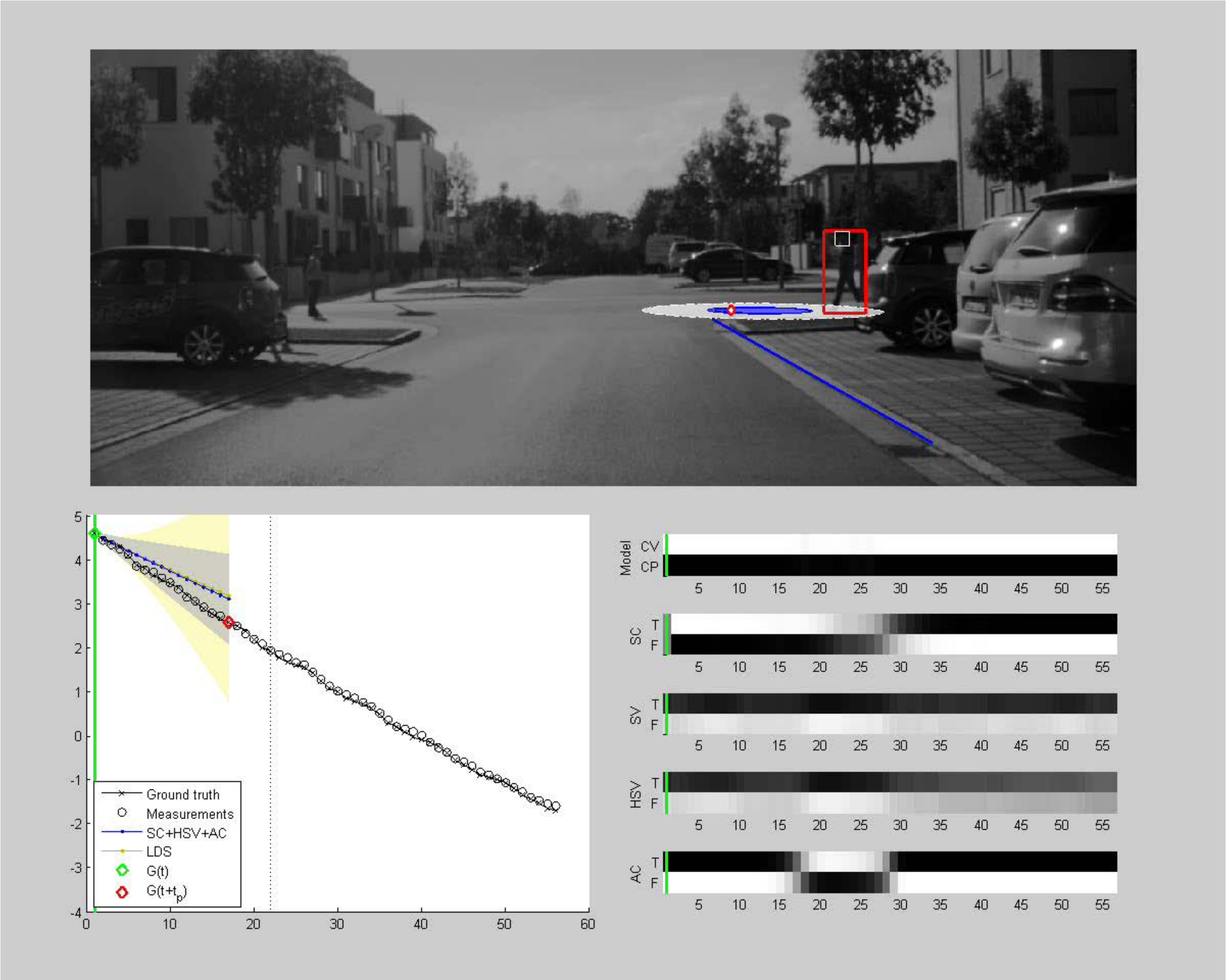
LDS (white) vs. context-based SLDS (blue), 1 s ahead, ±σ

Scenario: Stopping, situation critical, sees vehicle



LDS (white) vs. context-based SLDS (blue), 1 s ahead, ±σ

Scenario: Crossing, situation critical, not sees vehicle



Performance Evaluation

Metric

- Log likelihood of ground truth position *G* under the predictive distribution

$$predll(t_p|t) = \log\left[\overline{P}_{t_p|t}(G_{t+t_p})\right]$$

Results (*predll* 16 frames ahead ~ 1 s)

Higher (i.e. less negative) values are better	LDS (Really bad)	SLDS	**C-SLDS**	PHTM [1] (Other state-of-art is worse too)
SceneCritical=0, HasSeenVehicle=0, Crossing	-1.90	-0.59	-0.61	-0.78
SceneCritical=0, HasSeenVehicle=1, Crossing	-1.93	-0.49	-0.53	-0.75
SceneCritical=1, HasSeenVehicle=0, Crossing	-1.88	-0.33	-0.48	-0.97
SceneCritical=1, HasSeenVehicle=1, Stopping	-1.88	-1.26	-0.33	-0.38

Similar

Benefit of context

[1] C. Keller, C. Hermes and D.M. Gavrila. Will the pedestrian cross? Probabilistic Path Prediction based on Learned Motion Features. *Proc. of the DAGM.* (DAGM 2011 Prize)

J. P. F. Kooij, N. Schneider, F. Flohr and D. M. Gavrila. Context-based pedestrian path prediction. *Proc. ECCV*, 2014.

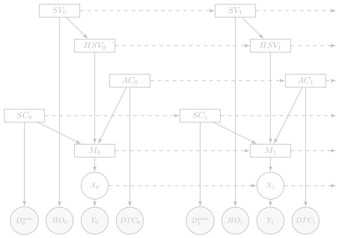
Context-based Pedestrian Path Prediction

Outline

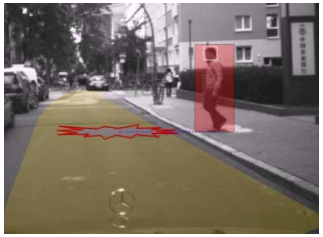
Regions of Interest
(Stereo, Motion, Geometry)

Object
Classification

Pose Estimation
& Tracking



Behavior Modeling
& Path Prediction



Collision Risk Assessment

Driver Warning / Vehicle Control

Vehicle Demo – Midterm BMWi UR:BAN Project (May 2014)

Scenario: Crossing, situation critical, does not see vehicle

„New“ warning comes 1 s earlier !

High tone: „state-of-the-art“ warning (Kalman Filter, based on current position estimate)

Low tone: „new“ warning (context-based SLDS, based on position predicted 1 s ahead)

Vehicle Demo – Midterm BMWi UR:BAN Project (May 2014)

Scenario: Stopping, situation critical, sees vehicle

No false alarm(s) !

High tone: „state-of-the-art“ warning (Kalman Filter, based on current position estimate)

Low tone: „new“ warning (context-based SLDS, based on position predicted 1 s ahead)

Automatic Evasion

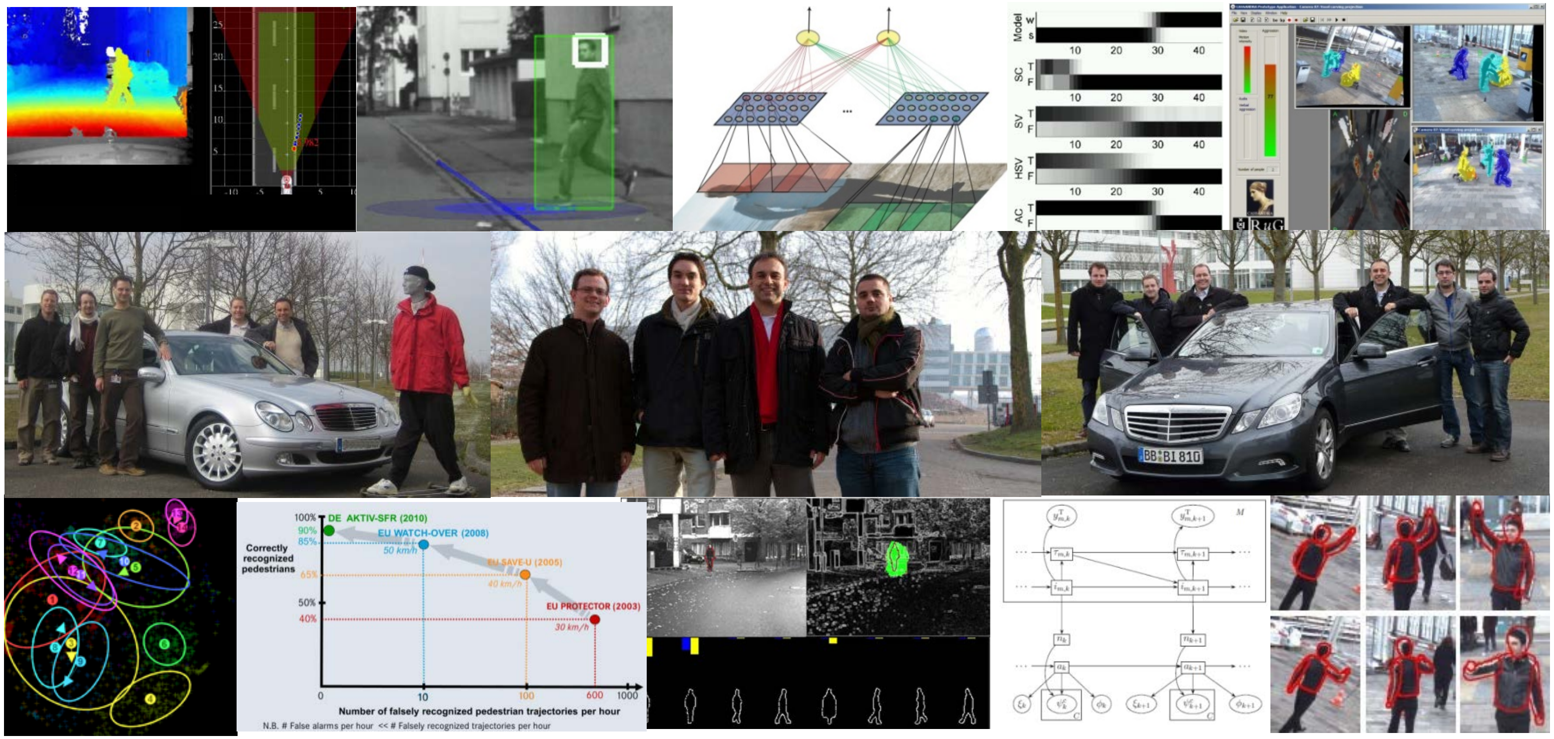
C. Keller, T. Dang, A. Joos, C. Rabe, H. Fritz, and D.M. Gavrila. Active Pedestrian Safety by Automatic Braking and Evasive Steering, IEEE Trans. on Intelligent Transportation Systems, 2011

300 ms from first sight of pedestrian to initiation of vehicle maneuver (braking or evasion)

Autonomous Driving

Mercedes-Benz Autonomous Driving on the Memorial Bertha-Benz Route, 09-2013
(Mannheim - Pforzheim, ~100 km on secondary and urban roads)

These are interesting times ...



... but the best is yet to come!

Credits

M. Enzweiler, F. Flohr, J. Giebel, M. Hofmann, C. Keller, J. Kooij, S. Munder, N. Schneider, and others ...