DAIMLER

Context-based Pedestrian Path Prediction

Julian F. P. Kooij[1,2], Nicolas Schneider[1,2], Fabian Flohr [1,2], and Dariu M. Gavrila[1,2]

[1]Environment Perception, Daimler AG R&D, Ulm, Germany [2]Intelligent Autonomous Systems Group, Univ. of Amsterdam, The Netherlands

Task

- Driver assistance in intelligent vehicles
- Scenario: pedestrian approaches curb in front of vehicle
- Where will pedestrian be in a second? Will (s)he cross?
- Improve path prediction by considering the **context**:

*In a **critical situation**, the pedestrian will **likely stop at the curb** if (s)he **is aware** of the approaching vehicle.*

Computer Vision can help determine this context

- Is the situation critical, are we on collision course?
- Has pedestrian seen the vehicle?
- Has pedestrian reached the curbside?

Dynamic Bayesian Network

Pedestrian dynamics as Switching Linear Dynamical System (SLDS)

- x_t : pedestrian's lateral position
- y_t: (noisy) lateral position measurement
- v_t : velocity, selected by switching state M_t

$$x_t = x_{t-\Delta t} + v_t \Delta t + \varepsilon_t \Delta t$$
$$y_t = x_t + \eta_t$$
$$\varepsilon_t \sim \mathcal{N}(0, Q) \qquad \eta_t \sim \mathcal{N}(0, R)$$

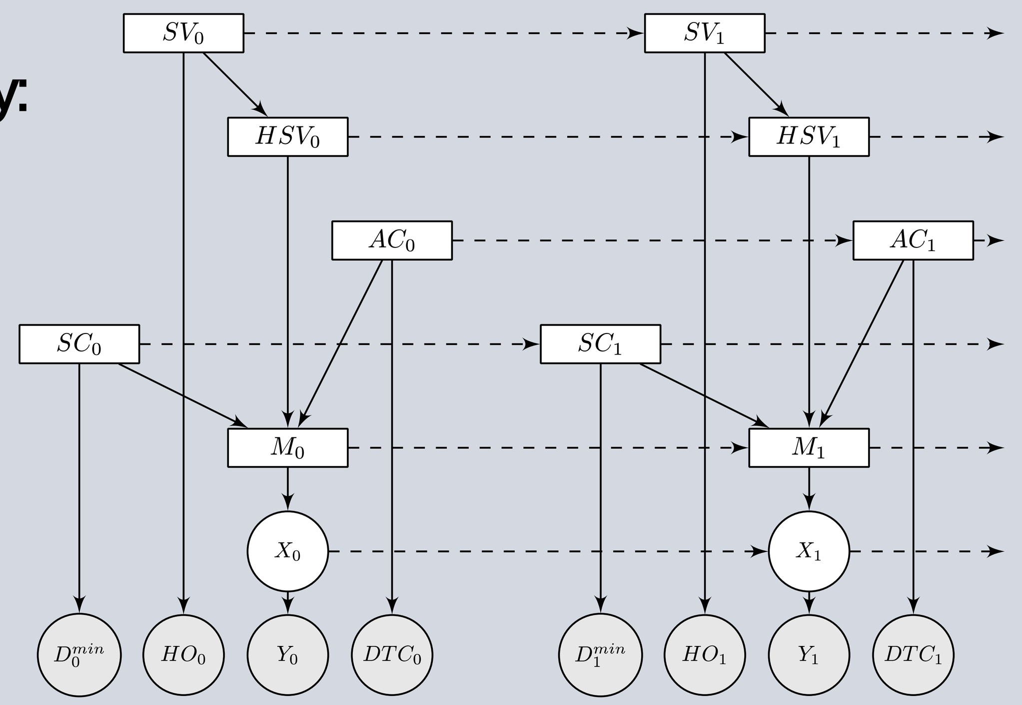
$$v_t = \begin{cases} 0 & \text{if } M_t = m_s \quad \text{(i.e. standing)} \\ v^{m_w} & \text{if } M_t = m_w \quad \text{(i.e. walking)} \end{cases}$$

- Here, v^{m_w} is the walking velocity, which we filter as part of the LDS state $X_t = \begin{bmatrix} x_t \\ v^{m_w} \end{bmatrix}$

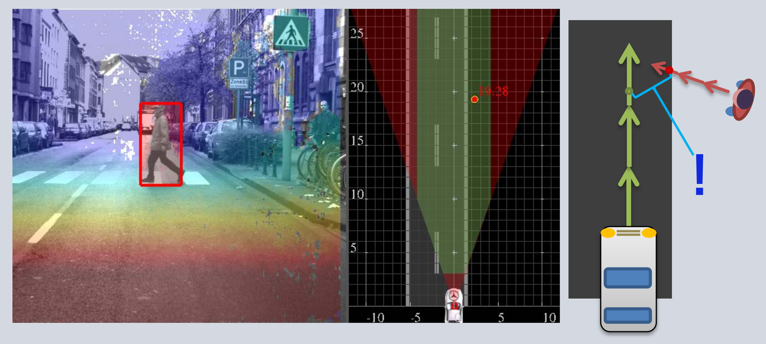
Boolean context nodes affect SLDS walking/ standing transition probability:

- SC: Situation is Critical
- SV: pedestrian Sees Vehicle
- HSV: pedestrian Has Seen the Vehicle
- AC: pedestrian is At the Curb

Context nodes define **distributions over context observables**



Context observables

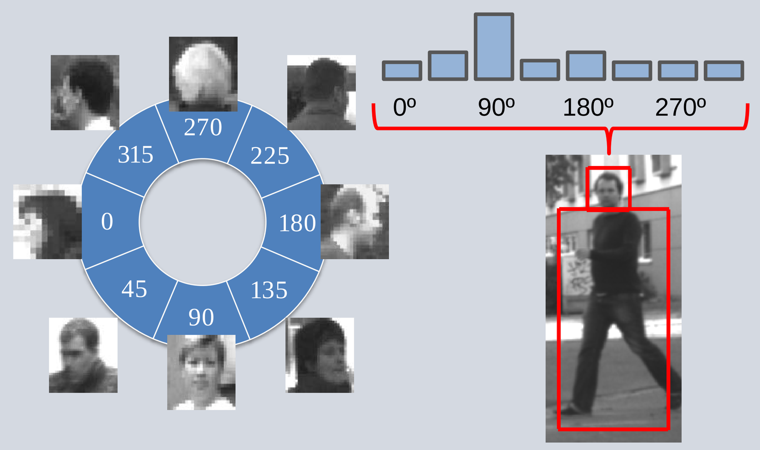


Lateral pedestrian position Y_t

- 2D tracking with stereo vision

Minimum distance | Situation Critical

- $\Gamma(D_t^{min} \mid SC_t)$
- Point of closest approach of vehicle and pedestrian paths
- Low distance indicates collision



Head orientation | Sees vehicle

- Multinomial$(HO_t \mid SV_t)$
- LRF features + Neural Network classifiers for 8 orientation classes
- Feature vector of classifier outputs

Distance-To-Curb | At Curb

- $\mathcal{N}(DTC_t \mid AC_t)$
- Difference of curb position and Y_t
- Lateral position of curb measured with Hough transform
- ROI determined by map location

Probabilistic inference

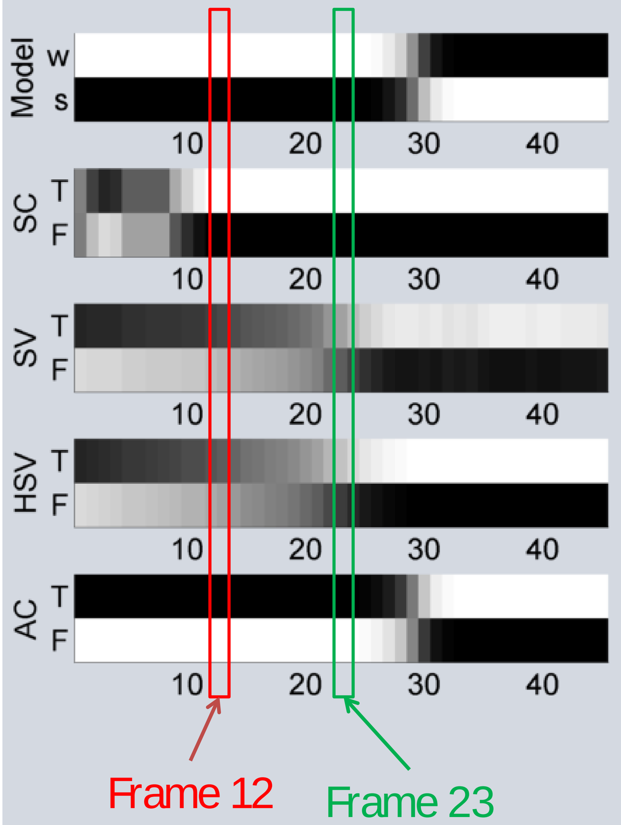
Assumed Density Filtering (ADF)

1. Predict $\overline{P}(Z_t, Z_{t-1} \mid O_{1:t-1}) = P(Z_t \mid Z_{t-1}) \tilde{P}(Z_{t-1} \mid O_{1:t-1})$
2. Update $\hat{P}(Z_t, Z_{t-1} \mid O_{1:t}) = P(O_t \mid Z_t) \overline{P}(Z_t, Z_{t-1} \mid O_{1:t-1})$
3. Collapse $\tilde{P}(Z_t \mid O_{1:t}) = \sum_{Z_{t-1}} \hat{P}(Z_t, Z_{t-1} \mid O_{1:t})$

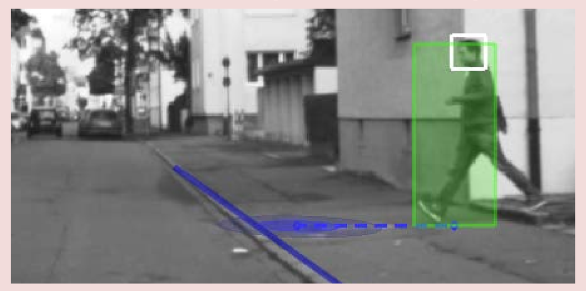
- Approximate and fast: limited number of modes in posterior
- Path prediction: perform "predict"/ "collapse" steps without "update"

Example of inference: Critical, Vehicle seen, Stopping (compared to other DBNs)

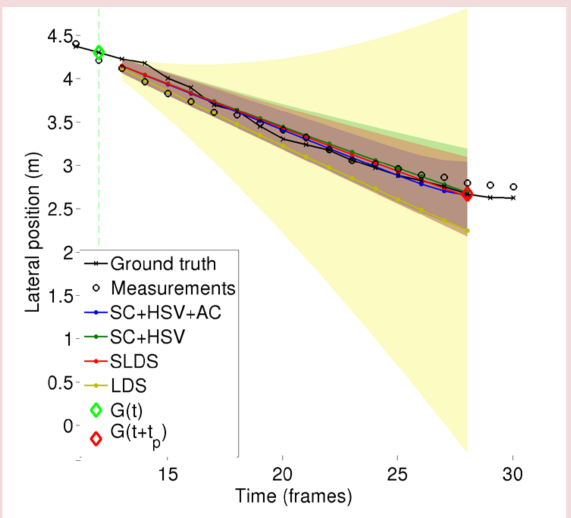
Posterior of latent nodes
w = walking, s = standing
T = true, F = false



Frame 12: vehicle not yet seen, all methods expect no stop

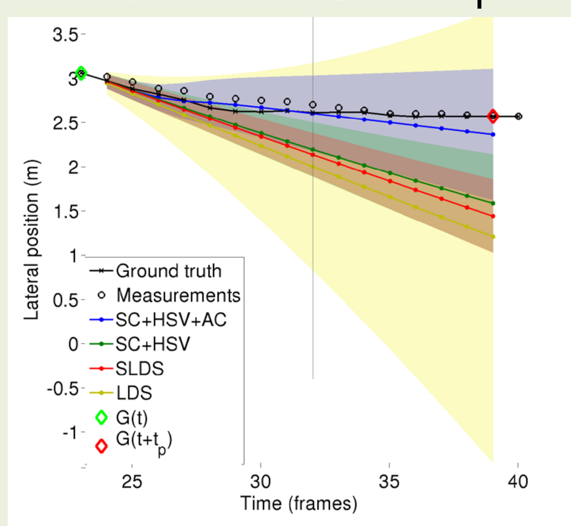


Predictive distributions of path



Frame 23: vehicle seen, our model predicts stop at curb

Predictive distributions of path



Evaluation

Comparison of our DBN with full context (SC+HSV+AC) against

- DBN with less context (e.g. SC+HSV only), or no context (SLDS)
- Standard Kalman Filter for position and velocity (LDS)
- Probabilistic Hierarchical Trajectory Matching (PHTM)†.

Leave-one-out Cross validation

- 58 sequences @16 fps (4 normal, 1 anomalous sub-scenario)
- Parameters estimated from ground truth
- Manual annotations of ground truth for context variables

Predictive Log-likelihood @1 sec time horizon

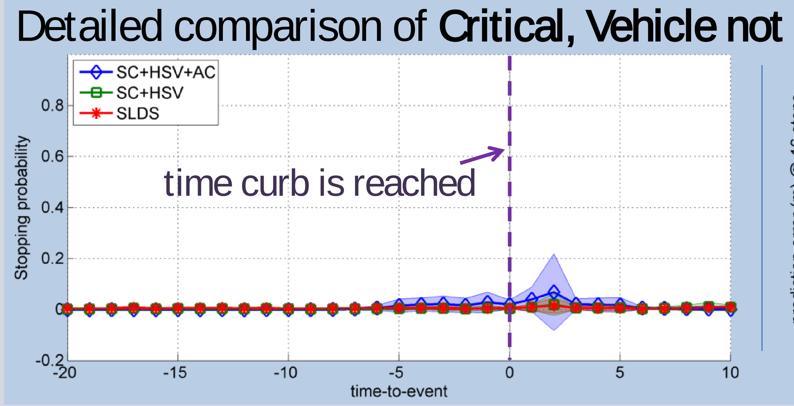
	Sub-scenario			SC+HSV+AC	SC+HSV	HSV	SC	SLDS	LDS	PHTM†
4 normal cases (higher is better)	Non-critical,	Vehicle not seen,	Crossing	-0.61	-0.53	**-0.52**	-0.59	-0.59	-1.90	-0.78
	Non-critical,	Vehicle seen,	Crossing	-0.53	**-0.45**	-0.46	-0.47	-0.49	-1.93	-0.75
	Critical,	Vehicle not seen,	Crossing	-0.48	-0.34	**-0.17**	-0.59	-0.33	-1.88	-0.97
	Critical,	Vehicle seen,	Stopping	**-0.33**	-0.70	-1.13	-0.80	-1.26	-1.88	-0.38
1 anomalous case (lower is better)	Critical,	Vehicle seen,	Crossing	**-0.90**	-0.27	-0.15	-0.25	-0.13	-1.88	-0.80

Computational costs

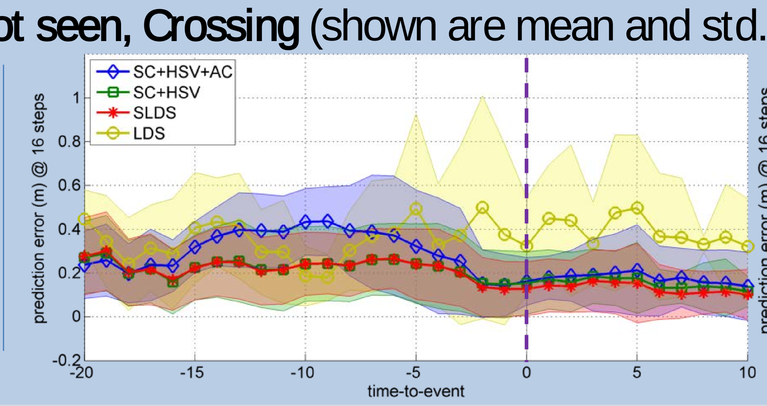
avg. per frame, in *milliseconds* (lower is better)

Approach	Observables	State est. & prediction	Total
SC+HSV+AC	160	40	200
SLDS	60	10	70
LDS	60	0.4	60
PHTM†	70	600	670

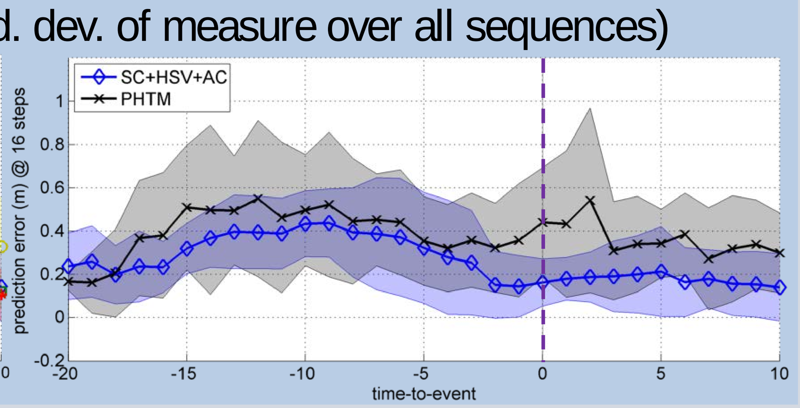
Detailed comparison of **Critical, Vehicle not seen, Crossing** (shown are mean and std. dev. of measure over all sequences)



Stopping probability

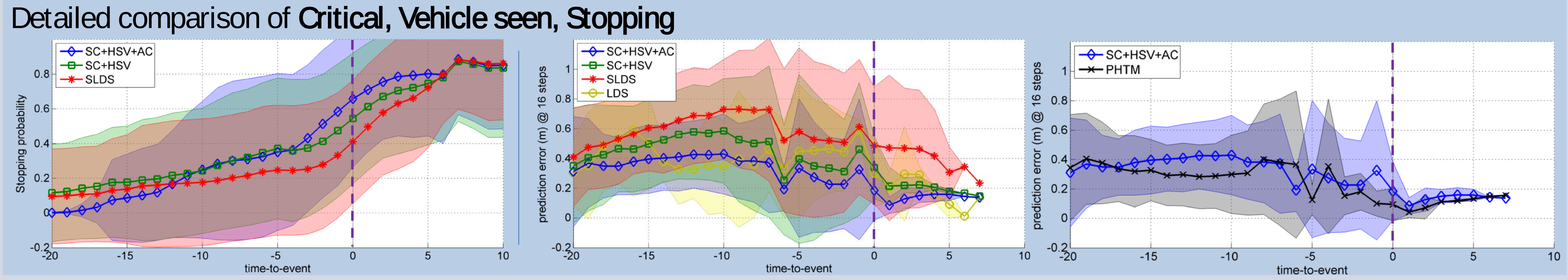


Predictive spatial error @1 sec horizon compared to DBN variants (lower is better)



Predictive spatial error @1 sec horizon compared to PHTM (lower is better)

Detailed comparison of **Critical, Vehicle seen, Stopping**



Conclusions

- SLDS (no context) can only react to change in dynamics after they occur
- Computer vision can give useful cues (pedestrian awareness, scene layout)
- Proposed context-based SLDS reduces uncertainty if/ when change will occur, and improves path prediction
- Enables early system reaction (warning/ breaking), can save lives
- Low computational cost (vs. PHTM)

†) Keller, C.G., Gavrila, D.M.: Will the pedestrian cross? A study on pedestrian path prediction. IEEE Trans. ITS 15(2), 494–506 (2014)