Supplemental Material
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 R&D, Ulm, Germany
{nicolas.schneider,fabian.flohr}@daimler.com
[2] Intelligent Systems Laboratory, Univ. of Amsterdam, The Netherlands
{J.F.P.Kooij,D.M.Gavrila}@uva.nl

This document contains supplemental material to our ECCV 2014 paper.

Section 1 shows that our method outperforms the the state-of-the-art Probabilistic Hierarchical Trajectory Matching (PHTM) method from [2] when comparing with the $predll$ measure. Section 2 contains more plots for the prediction $error$ that were omitted from our paper due to page constraints.

Recall that the prediction error, $error$, is the Euclidean distance between the true future position, and the expected future position predicted by a model (i.e. the expected value under the predicted future state distribution). The predictive log-likelihood, $predll$, is the log likelihood of the true future position under the predicted future state distribution. Thus, this measure is large when both the true position is close to the expected position (i.e. high accuracy) and the variance is low (i.e. high precision).

1 Comparison of $predll$ with PHTM

Table 1 in our paper submission shows the predictive log likelihood $predll$ (Eq. (22) in the submission) of the model variations at $t_p = 16$ time steps ($\sim 1\,s$) in the future, averaged over the second up to the moment that the pedestrian reaches the curb. For completeness, Table 1 here includes the predictive log likelihood for the PHTM model too, in the rightmost column.

In the PHTM model, the predictive distribution $\overline{P}_{t_p|t}(X_{t+t_p})$ for state X_{t+t_p} at time t is represented by a set of particles $\{X^{(s)}_{t+t_p}\}$. Each particle is associated with a trajectory snippet from training trajectories in the database, and the future position of the particle is obtained by following the path the trajectory that best matches the snippet (see [2] for a more detailed description of their approach). To compute the predictive likelihood $predll(t_p|t)$ of the PHTM model shown in the table, we used kernel density estimation. A Gaussian kernel was placed on each particle in $\{X^{(s)}_{t+t_p}\}$, and weighted by the particle weights $w^{(s)}$. Bandwidths for kernels were estimated using the data-driven linear diffusion approach proposed by [1] in *leave-one-out* cross-validation.

The results in Table 1 here show that PHTM is good in predicting stopping motion, similar to our proposed model, but the predictive likelihoods for crossing pedestrians are low compared to our proposed model. We observe from the output of PHTM that on the stopping sub-scenario, almost all predicted positions come from tracks in the database with stopping motion. This results in an accurate predictive distribution with

Table 1. Prediction log likelihood of the GT pedestrian position for $t_p = 16$ frames ($\sim 1\,s$) ahead, for different sub-scenarios (rows) and models (columns), for TTE $\in [-15, 0]$. The first four sub-scenarios contain "normal" pedestrian behavior. The fifth case is anomalous (*lower* likelihood is better). Model variations (best SLDS variant marked in bold): full context (SC+HSV+AC), no curb (SC+HSV), only head (HSV), only criticality (SC), no context (SLDS), KF (LDS). This table is the same as Table 1 of our ECCV paper, except for the last columns with the results for the PHTM model.

Sub-scenario	SC+HSV+AC	SC+HSV	HSV	SC	SLDS	LDS	**PHTM**
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
critical, vehicle seen, crossing	**-0.90**	-0.27	-0.15	-0.25	-0.13	-1.88	-0.80

low variance, and therefore high likelihood (i.e. -0.38). On the crossing sub-scenarios, however, the predicted positions are spread out, and the distribution accordingly has high variance. We found that PHTM matches snippets to tracks with many different walking speeds and variations in the successor parts of the trajectories. On the other hand, our approach uses linear dynamics to estimates walking speed, and can handle both stopping and crossing. Furthermore, the PHTM model has no additional context information and can only differentiate between the two motion types. Indeed, the table shows that the predictive log likelihood of PHTM is low for all crossing sub-scenarios.

In summary, our proposed approach outperforms the state-of-the-art PHTM method in all sub-scenarios when comparing the predictive log-likelihoods for $\sim 1\,s$ in future, especially on sub-scenarios with crossing pedestrians.

2 Comparison of model variations

In Fig. 3 in our paper, the plots in the right column illustrated the prediction error, for $\sim 1\,s$ in the future, of the model variations on the two critical scenarios. Each plot shows the mean and 1 std. dev. range (over all sequences in a particular sub-scenario) of a model's prediction error over time. For completeness, Figure 1 here shows plots of the prediction error of *all* normal scenarios (both non-critical and critical).

From the two figures of the non-critical situations (top-row), we see that our model and its variations perform all very similar, and make better predictions than the standard constant velocity Kalman Filter with acceleration noise. In the non-critical case where the pedestrian sees the vehicle and crosses (top right), one can observe that our proposed approach has a slightly higher error than the variations, which was also visible in the slightly lower likelihood for this case in Table 1. The reason for this is that in a few sequences, at some moments before the pedestrian reaches the curb, the model does find it probable that the situation is critical. Therefore, it is also probable that the pedestrian will stop, and the predictive positional distribution contains most mass at the curbside.

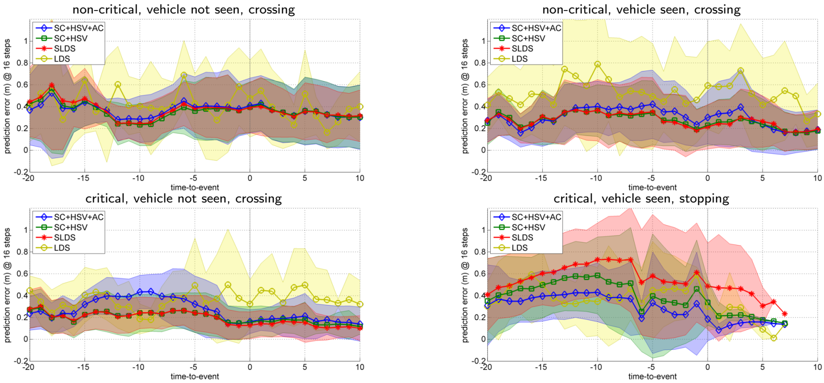

Fig. 1. Predictive error of all normal scenarios, similar to the right column of Fig. 3 in our paper.

References

1. Botev, Z.I., Grotowski, J.F., Kroese, D.P.: Kernel density estimation via diffusion. The Annals of Statistics 38(5), 2916–2957 (2010)
2. Keller, C., Gavrila, D.: Will the pedestrian cross? A study on pedestrian path prediction. IEEE Trans. ITS 15(2), 494–506 (2014)