Honda Research Institute EU
Honda Research Institute Europe GmbH
https://www.honda-ri.de/

Pedestrian Crossing Prediction using Multiple Context-based Models

Sarah Bonnin, Thomas H Weisswange, Franz Kummert, Jens Schmüdderich

2014

Preprint:

This is an accepted article published in 17th International IEEE Conference on Intelligent Transportation Systems. The final authenticated version is available online at: https://doi.org/[DOI not available]

Pedestrian Crossing Prediction using Multiple Context-based Models

Sarah Bonnin, Thomas H. Weisswange, Franz Kummert, and Jens Schmuedderich

Abstract— **In inner-city, most vehicle-pedestrian collisions occur when a pedestrian is crossing the road and the driver does not see or pay attention to him. Current ADAS (advanced driver assistance systems) warn the driver or apply the brakes shortly before the collision, but in some situations the collision cannot be fully avoided because most systems react only when the pedestrian is already in front of the vehicle. To fully avoid a collision, a driver should be warned earlier. Behavior prediction is a solution that can be used to warn a driver before the pedestrian starts crossing. In this paper, we propose a generic context based model to predict crossing behaviors of pedestrians in inner-city. We will show that our model provides accurate prediction at an early time. However, there are specific locations such as zebra crossings, where based on expert driving experience, one would expect that a prediction can be done even earlier. Therefore, we have developed an additional specific model fitted to the context of zebra crossings. The experiments show that this model produces both, better and earlier predictions in this specific context. Because our goal is to build a generic crossing prediction system, we finally apply the framework of the 'Context Model Tree' to combine the two models. We demonstrate that this multi-model system is well suited to provide early predictions for realistic data, including both, generic inner-city situations and zebra crossings.**

I. Introduction

Pedestrians are the most vulnerable road users and most accidents happen while they are crossing a road [1]. When compared to cars, pedestrians have a higher dynamic range, which might make them more difficult to predict.

Nowadays, pedestrian protection is a common function of current advanced driver assistant systems (ADAS). For example, the Collision Warning with Full Auto Brake and Pedestrian Detection system, that is commercialized by Volvo, alerts the driver or starts an automatic emergency braking when the time-to-collision (TTC) with a pedestrian in front drops below a threshold [2][3]. However, as it is mentioned in [4][5], such reactive systems, that use pedestrian detection only, are often not sufficient to fully avoid a collision. More advanced collision avoidance systems require behavior prediction using pedestrian dynamics and behavior analysis.

A simple prediction method uses the current position and velocity to estimate the most probable next location of a dynamic object. Because the method is based only on physical observation, it can be expected to be accurate for objects with a low dynamic range and/or for a very short time horizon. This method is mostly used in the prediction step of tracking algorithm [6]. But in [7], the authors use the estimation of the next position of a moving pedestrian to compute a probability of collision with the ego-vehicle (vehicle equipped with the sensors that we used for our recordings).

A pedestrian usually follows a particular path, either to cross a road, walk along the pavement or turn at an intersection. The different paths of a pedestrian are all composed of a succession of positions and velocities that can be learned [8]. A matching algorithm can then be used to compare the current trajectory of a pedestrian (in terms of spatial location, velocity or heading angle) with the learned ones. When the best match is found, the corresponding trajectory can be used to predict future positions of the pedestrian and estimate a risk of collision [9].

Tracking and trajectory prediction methods define a position of a pedestrian for any point in time. Because pedestrians can change their walking direction very suddenly, therefore the prediction error will grow fast with increasing time horizon. An alternative would be to look at the problem from a higher level, that is, to predict pedestrian behavior. A behavior describes a category of future movements, for example walking versus stopping. Behavior prediction tries to classify a category without defining every single position state. This is usually enough to decide if the pedestrian will potentially collide with a vehicle. The authors in [10][11] learn motion features to recognize the start of a crossing behavior. They define the first foot motion towards the road at a specific location as the start of such a behavior. When a pedestrian is classified as crossing the road, they use the current position and velocity of the ego-vehicle to evaluate the risk of collision and initiate a braking or evasion maneuver. However, they use a static sensor setup at a specific location where pedestrians usually cross the road, which limits the scope of such methods.

A pedestrian walking towards the road might cross, but in some situations, e.g. when a pedestrian traffic light is red, it might also stop. The authors in [12] propose two different approaches that combine trajectory prediction with behavior prediction to differentiate between crossing and stopping behavior. The first approach uses walking trajectories and stopping trajectories respectively to train two separate models. Both models learn pedestrian location information with extracted features from dense optical flow. They combine both models with a multi-model particle filter to compute the probability of a pedestrian's future state. In the second approach they learn different crossing and stopping trajectories using pedestrian positions and motion

S. Bonnin is with Cor-Lab, Bielefeld University, Germany sarah.bonnin@gmail.com

F. Kummert is with the Faculty of Technology, Bielefeld University, Germany Franz@techfak.uni-bielefeld.de

T.H. Weisswange and J. Schmuedderich are with the Honda Research Institute Europe GmbH, Offenbach am Main, Germany FirstName.LastName@honda-ri.de

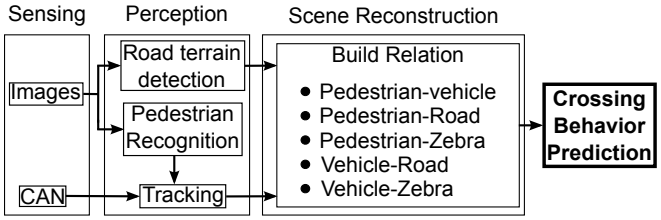


Fig. 1. Overview of the system architecture for crossing behavior prediction.

features in a single model (flow variations on the pedestrian legs and upper body). Afterward, they use a probabilistic matching algorithm to find the most probable future state of the pedestrian. This is then used to compute features like the time-to- stop or the time-to-curbstone to compute a probability of crossing or stopping. Both methods can distinguish a crossing behavior from a stopping behavior with 80% accuracy on average 0.23s before the pedestrian enters the egolane. For comparison, for the same accuracy, a physical prediction method predicts a crossing behavior only 0.09s in advance. They also show that a human has a prediction time horizon of 0.57s for the same data. This shows that distinguishing stopping from crossing behavior is not an easy task, even for a human [12], but he still generally outperforms the model.

In contrast to a motion based model, humans analyze the current situation, deducing what the pedestrian will do, based on knowledge and contextual information. Where and when a pedestrian will cross the road is highly related to its specific context location [13]. For example, a pedestrian walking towards a zebra crossing is more likely to cross. On the other hand, a pedestrian approaching the road at an intersection might turn left or right instead of crossing the road. A pedestrian with the intention to cross might not do so if there is a car approaching fast, but might cross if the car is still far away.

In this paper, we propose a generic context based model to predict crossing behaviors of pedestrians in inner-city. We will focus on those cases where the pedestrian actually crosses the path of the ego-vehicle. We show that our model performs generally well on a real inner-city dataset. In order to evaluate the influence of the environment, we test a model adapted to the specific stetting of a zebra location. In an earlier publication, we proposed a context model tree framework designed to combine multiple models fitted to different contexts [14][15]. We could show the benefit of it when applied to cut-in prediction on highways, where we combined a model for general highways with a model for entrance scenarios. Therefore, we also want to test this framework to combine the generic inner-city model with the zebra model to predict pedestrian crossing behaviors in inner-city.

The next section gives an overview of our system for behavior prediction. Section III presents the inner-city model with its features, explains the dataset used to train and test the classifier and proposes an event-based evaluation. Section IV presents the zebra model with its results. Section V introduces the context model tree used to combine the inner-city and zebra models and shows the results of the combination with an event-based evaluation. Finally, section VI concludes the paper and presents future work.

II. System Overview

Fig.1 provides an overview of the system architecture built to predict pedestrian behavior.

The sensing module represents the input of the system. Color images are received with a frame rate of 20Hz provided by a stereo camera mounted behind the windshield of a vehicle. We also record CAN data from the ego-vehicle.

The system uses the stereo-images to compute disparities. The rest of the perception module works on single images. One part does detection and recognition of pedestrians, using the method of [16], and assigns those to a 3D position using disparities. The tracking module [17] filters 3D positions, assigns an ID to each detected pedestrian and computes its velocity. Another module uses a spatial image recognition algorithm to detect different types of road terrain [18]. In this algorithm, the road is segmented into two parts, the egolane which is the area where the ego-vehicle drives, and the road area which is the terrain where vehicles drive. We use an extension of this algorithm to also detect zebra crossing areas.[1]

The scene reconstruction module uses all the information provided by the perception module to create the relations between traffic entities and contextual information. Fig.2 shows a scenario with a pedestrian (P1) and a zebra crossing. We compute, for example, the distances between all entities and the zebra crossing.

The prediction module uses this reconstructed scene to compute features which are used to predict if a pedestrian is going to cross in front of the ego-vehicle. This paper is mainly focusing on this last module. The output of this module could then be used to decide if warning the driver or reducing velocity is necessary.

III. Inner-City Model

A. Definition

When driving in inner-city, a driver has to care for pedestrians that can cross its path at any time. To prevent a car-pedestrian collision, we want to develop a generic inner-city model that covers usual behaviors of pedestrians in inner-city based on physical motion with additional knowledge of the context. Fig.2 shows a scene in inner-city where a pedestrian P1 is going to cross in front of the ego-vehicle. The arrows 1-9 represent features 1-9 of Table I that are used by the inner-city model.

The features have been designed based on expert knowledge. For example, a pedestrian who is going to cross will be close to the curbstone before entering the road. This behavior is described by the *distanceToCurbstone* feature. A feature

[1]This basically means training the same algorithm with annotated images including zebra crossing areas.

TABLE I

FEATURE DEFINITION

N°	Feature Name	Description and scaling bounds (IC: inner-city model, Z: zebra model)
1	distanceCollisionPoint	Lateral distance between the pedestrian and the collision point (CP), where pedestrian and ego-vehicle will intersect. IC:[-6;10]
2	timePedCollisionPoint	Time for the pedestrian to reach the collision point. IC:[0,6] Z:[0;12]
3	diffTimeColision	Time for the ego-vehicle to reach the collision point minus time for the pedestrian to reach the collision point. IC:[-8;10] Z:[-6;4]
4	distanceToCurbstone	Lateral distance between the pedestrian and the curbstone. IC:[-6;4] Z:[-2;8]
5	timeToCurbstone	Time for the pedestrian to reach the curbstone. IC:[-2;4] Z:[-1;11]
6	timeToEgoLane	Time for pedestrian to reach the egolane. IC:[-1;4] Z:[0;12]
7	globalOrientation	Moving direction as the angle between the pedestrian and the road.
8	isFacingRoad	Pedestrian is walking towards the road, parallel or away to it. IC:{0;1} Z:{0;1}
9	relativeOrientation	Moving direction as the angle between the pedestrian and the ego-vehicle. IC:[0;3.14] Z:[0;3.14]
10	distanceToEgoLane	Lateral distance between the pedestrian and the egolane. Z:[-7;9]
11	distanceToZebra	Lateral distance between the pedestrian and the zebra crossing. Z:[-4;10]
12	timeToZebra	Time for the pedestrian to reach the zebra crossing. Z:[-3;6]

alone does not suffice to predict a crossing behavior. For example, a very small value of *distanceToCurbstone* does not necessary mean that the pedestrian is going to cross, it could be just waiting for a bus. On the other hand, the same feature combined with the knowledge that the pedestrian is moving towards the road (*timeToCurbstone*) indicates a crossing behavior. To be able to handle many situations (including static pedestrians starting to cross) we use a multitude of features. Because the combination between features is not an easy task, we use a learning classifier. As we did in our previous work [14], we chose to train a single layer perceptron [19].

To fit the characteristics of a single layer perceptron, all features need to be scaled to a similar range. For each feature computed on the training set, we have separated the values into two groups: crossing and non crossing. When looking at the distribution between both groups, we observed that in some ranges the two groups overlap, whereas in some others they do not. It is easy to separate the two groups when they do not overlap. For that reason, we have chosen a function that preserves the values in the overlapping area. A feature like the *distanceToCurbstone* will originally be $[-\infty,\infty]$, but we observed that the *distanceToCurbstone* of a pedestrian which does cross the road varies between [-6, 4]. A minimum and maximum threshold are defined for each feature as shown in Table I. Afterward, we choose a fermi function to scale the features 1-6 and 9-12 because it limits the impact of the choice of the thresholds compared to, e.g., a piecewise linear function:

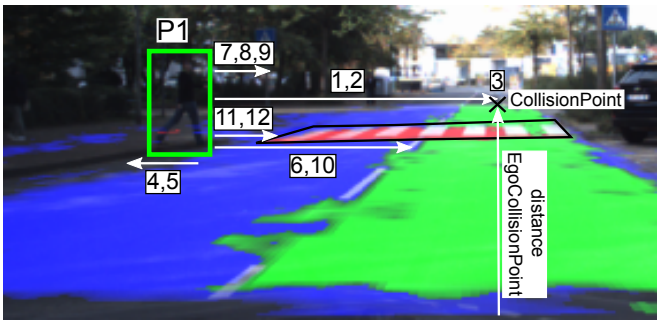


Fig. 2. Features computed to predict if pedestrian P1 is going to cross in front of the ego-vehicle. The green area represents the egolane. The blue area represents the road area. The red area withing the black rectangle is the zebra crossing. Each number is associated to a feature described in Table I.

$$\text{fermi}(x,min,\ max) = 1-(2/(\exp((x-\mu)/k)+1)) \quad (1)$$

The parameter $\mu = \min + (\max - \min)/2.0$ is used to center the results in the interval and the parameter $k = (\max - \mu)/log(1/0.9-1)$ changes the slope of the curve.

Features 7-9 are angles defined on $[-\pi,\pi]$. To extract the meaningful information without taking into account the cyclic dependencies, they have to be preprocessed. Before applying the fermi function to the feature *relativeOrientation*, we have applied the following formula: $p = abs(90-x)$. We do not apply the fermi function to the feature *globalOrientation*, instead we divide the feature value by π because there was no clear separation between the two groups. To compute the feature *isFacingRoad* as a boolean value, we first check the lateral position of the pedestrian. If the pedestrian is on the right of the ego-vehicle and the absolute value of the feature is above 90°, the pedestrian is facing the road and the scaled value of the feature is 1, otherwise it is 0. On the left of the ego-vehicle a pedestrian is facing the road if the feature value is below 90°.

B. Experiments

In this section, we present an evaluation of the performance and prediction time horizon of the generic inner-city model using real data. We show that the inner-city model performs generally well at predicting crossing behaviors, but in specific locations, the prediction starts later than what we could expect.

Recordings: All tests have been done using real data recorded by a camera mounted on the windshield of a vehicle on German inner-city roads (curved and straight). We recorded a single stream of one hour length in areas where pedestrians often cross the road [2], with schools, shops, bus stops and zebra crossings. An initial recording showed that in one hour, 5 to 6 pedestrians were crossing (out of

[2]Offenbach-Bieber, Germany, latitude 50.084697°N, longitude 8.827502°E

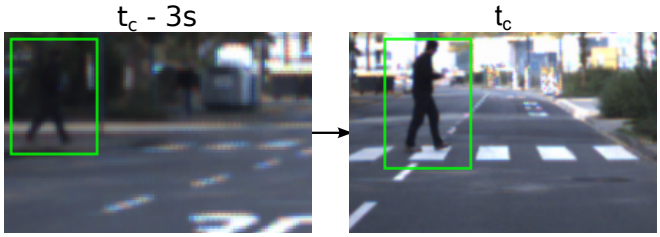


Fig. 3. Example of a crossing pedestrian. The right image shows a pedestrian just entering the egolane. This point in time t_c is defined as the beginning of the crossing behavior. The left image shows the same pedestrian at time $t_c - 3s$ which we define as the start of positive prediction event annotation.

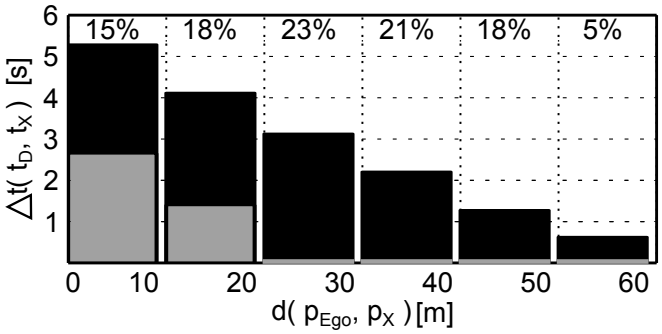


Fig. 4. Time between the first pedestrian detection t_D and the start of a crossing behavior t_X versus the distance between the ego-vehicle's position p_{Ego} and the crossing point p_X. Grey: pedestrians that are tracked by the original system. Black: pedestrians that are detected by the algorithm merged with manual pedestrian annotations. Percentage values on top: ratio of pedestrians that cross in the corresponding distance range.

83 overall). To increase the size of the positive dataset, we have instructed seven people with a list of tasks to perform while the ego-vehicle was approaching. Every time we were approaching one of them, they freely chose a task of the list. They could e.g. walk on the pavement, wait for the bus, cross at a zebra crossing, or cross in a non zebra crossing area, but we never knew in advance what they were going to do. They have been told to act naturally, without taking any risk when crossing the road. Our recording contains normal pedestrian crossings as well as several crossings of these instructed people.

In this paper, we define the start of a pedestrian crossing when its foot touches the egolane[3] as shown in Fig.3 and our goal is to predict a crossing behavior before this point. Fig.4 shows that on average, 15% of the pedestrians in our dataset start crossing at below 10m from the ego-vehicle, 18% of the pedestrians start crossing between 10m and 20m, and 63% of the pedestrians start crossing above 20m. Due to the camera's opening angle and resolution used for recording our dataset, the detection range of the pedestrian recognition system is limited to approximately 20m in longitudinal and 15m in lateral distance. On average, a pedestrian that starts crossing the road below 10m can be detected 2.65s before a crossing. Between 10m and 20m, the first detection of a crossing pedestrian starts 1.4s before a crossing. 67% of the crossing pedestrians are not detected at all due to the limited detection range. To be able to predict pedestrians earlier and at a further distance, we have decided to manually annotate pedestrians that are above 20m. The annotated 2D positions of pedestrians were mapped to 3D using the disparity map and then fed to the tracking algorithm. The black bars in Fig.4 show that with annotated pedestrians up to 5.3s and up to 60m distance. This "assisted" pedestrian detection lies within the range of state of the art laser scanners [20].

Dataset: We use a supervised learning method to predict behaviors, so we need to build one dataset for training the classifier and one dataset for testing. To train and test a classifier we need to compute the input features and provide the groundtruth at each timestep. That means that each ID has a corresponding annotation as a negative or positive example for each timestep.

[3]We use the egolane and not the road area to concentrate on those cases that really cross the path of the ego-vehicle.

In inner-city, we are interested in predicting if a pedestrian will cross the future path of the ego-vehicle. A pedestrian is considered as crossing when its foot touches the egolane as shown in Fig.3. When a crossing is detected, each timestep from that moment to three seconds before is annotated as a positive example (annotated to 1). Three seconds before that and while the pedestrian is on the egolane are ignored (annotated to -1) in the training because there is no obvious start or end point for a prediction. Everything else is annotated as negative example (annotated to 0).

The dataset contains 322 pedestrians with 36 of them crossing the road which shows a realistic ratio of crossing versus non-crossing, despite the use of instructed pedestrians. We excluded wrong pedestrian detections (e.g. bikes, trees, traffic signs) from the dataset in order to focus the evaluation on the prediction module rather than the general system.

Out of the dataset, we created a subset 'zebra' containing all the pedestrians that are within 10m distance of a zebra (18% cases). It might be interesting to separately test on this dataset, because we observed that the ratio of crossing versus non-crossing is very different from the rest of the dataset (26% crossing at zebra crossings versus 9% crossing in inner-city without priority). We also constructed a subset 'intersection' containing all pedestrians at intersections (26%). These cases will be very difficult, if the ego-vehicle turns, because they do not cross the ego-path despite being in front of the car. To deal with these pedestrians, we expect additional information about the ego-path to be required, therefore we do not use them in the following. We created a third subset 'inner-city' containing all other pedestrians.

The inner-city model is trained and tested on the 'inner-city' dataset and the 'zebra' dataset (the 'ICZ dataset').

Event Based Evaluation: To evaluate the inner-city model, we do an event based evaluation. We consider the consecutive timesteps in which for each pedestrian the prediction is above a threshold θ as one event (prediction gaps of size smaller than 1s are grouped into the same event). Following the principle of an ROC curve, θ is sampled from 0 to 1 and the TPR and FPR are computed for each setting of theta. To do

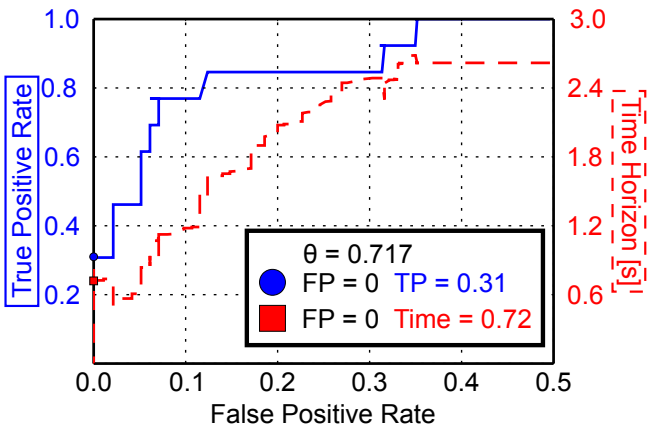


Fig. 5. Event based ROC curve of the generic inner-city model tested on the 'ICZ dataset'. The solid line shows True Positive Rate versus False Positive Rate. The dashed line represents an average of the PTH versus the False Positive Rate.

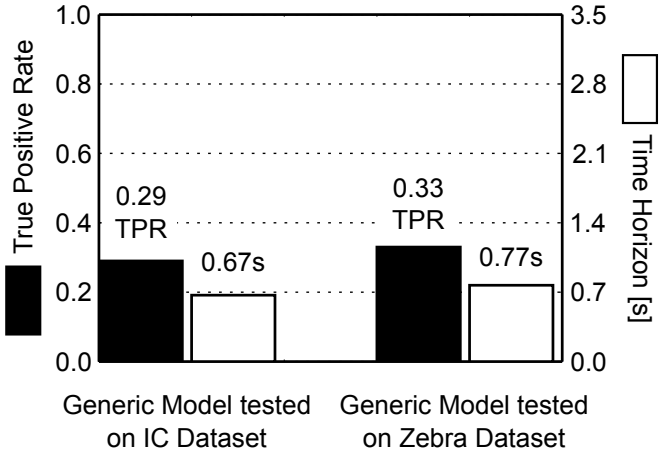


Fig. 6. Performance of inner-city model. Black and white bars show the TPR and PTH respectively. The left shows the performance on the 'inner-city' set. The right shows the performance on the 'zebra' set.

so, we compute the events and then compare those with the groundtruth annotation at the same timesteps. If the event overlaps with any timestep labeled to 1, it is counted as a true positive (TP). If it overlaps with any timestep labeled to 0, it is counted as a false positive (FP). For more details on event based evaluation seeing [15].

We compute the prediction time horizon (PTH) as the difference between the time of the first prediction of a TP event with the time t_c when the pedestrian starts crossing (Fig.3).

Fig.5 shows an ROC curve with the true positive rate (TPR) versus the false positive rate (FPR) of the inner-city model tested on the 'ICZ dataset'. The ROC curve shows that for a threshold θ_j of 0.717, the model has a TPR of 31% for a FPR of 0.0%. At this threshold, the model can predict a crossing 0.72s before it starts[4].

This ROC curve shows that by lowering the threshold, it would be possible to have a higher TPR with a better prediction time horizon. However, this comes at the cost of making prediction errors by increasing the FPR. We can see in the dataset that there are much fewer people crossing than non crossing (only 16% versus 84%). For that reason, we target for a minimum of errors, and choose θ_j for the following experiments.

We are interested in seeing if the model shows differences in performance between the two important types of context. We therefore tested it with the 'inner-city' and the 'zebra' dataset separately. Fig.6 shows that the inner-city model, can predict non priority inner-city crossings with a TPR of 29% on average 0.67s in advance without making errors. It can predict zebra crossing behaviors with a TPR of 33%, a FPR of 0% on average 0.77s in advance.

Discussion: These numbers show that the inner-city model can predict both types of crossings equally well. However, we expect that prediction at zebra crossings should be easier because a pedestrian at zebras is usually detected for a much longer time and its intention to cross is often more visible. We have developed a model specific to zebra crossings to test if it can produce even better results.

[4]The PTH versus FPR curve sometimes shows a reduction of time horizon despite lowering the threshold. This occurs because we use only true positive events to compute the time horizon, and a new TP usually starts with a small PTH.

IV. Zebra Model

A. Definition

This model is designed to predict crossing behaviors of pedestrians at zebra crossings, where pedestrians have the priority to cross the road. The features used by the pedestrian crossing node are similar to the ones in the inner-city node, however, the feature ranges learned by the classifier are different. For example, pedestrians crossing at zebras are usually closer to the ego-vehicle. A pedestrian standing at the side of the road is more likely to cross in the vicinity of a zebra crossing than in other areas. Fig.2 shows a scene where a pedestrian P1 is going to cross in front of the ego-vehicle at a zebra crossing. This model uses features 2-9 as described above. Additionally, it uses features 10-12 of Table I that are specific to zebra crossings. As for the inner-city model, we use a single layer perceptron as classifier and a fermi function to scale the features. For each feature, Table I also shows the corresponding min and max thresholds for scaling.

B. Experiments

In this section, we present an evaluation of the performance and prediction time horizon of the zebra model. We show that a zebra model, that uses specifically tuned contextual information, performs better at locations than a generic inner-city model, where more pedestrians are expected to cross. However, such a model does not generalize well to other scenarios.

Event Based Evaluation: We did an event based evaluation of the zebra model tested with the 'zebra' dataset. For this model, we chose a threshold which provides a good TPR and a long prediction time horizon. We have chosen the threshold

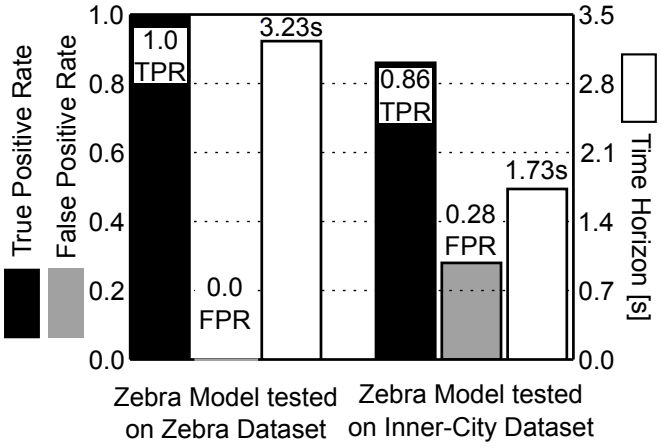


Fig. 7. Performance of zebra model. Black, grey and white bars show the TPR, FPR and PTH respectively. The left shows the performance on the 'zebra' set. The right shows the performance on the 'inner-city' set.

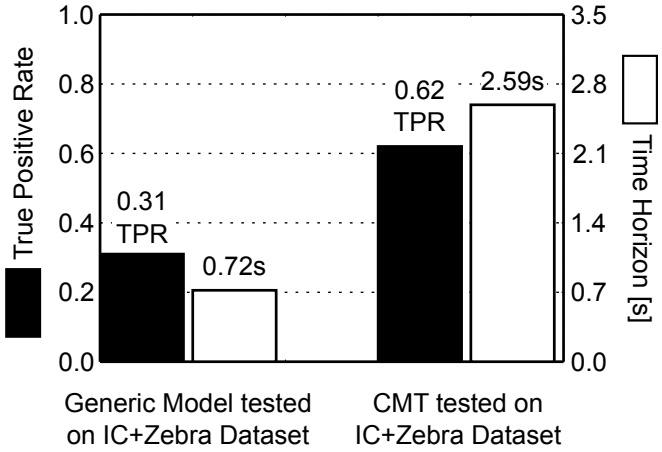


Fig. 8. Performance of inner-city model (left) versus CMT (right). Black and white bars show the TPR and PTH on the 'ICZ dataset'.

($\theta_i = 0.34$) that allows the model to have a good prediction quality for no prediction errors. Fig.7 shows that the zebra model, tested at zebra crossings, can predict all crossing pedestrians at zebras on average 3.23s in advance without doing any FP. These numbers confirm our hypotheses that pedestrians at zebra crossings can be predicted more easily and much earlier.

We have also tested the zebra model on the 'inner-city' dataset.[5] Fig.7 shows that the model, at θ_i, can predict crossings in inner-city with a TPR of 86% for an average of 1.73s in advance. But, the model also produces a FPR of 28%.

Discussion: From Fig.7, we can see that the model performs well at predicting non priority crossings in inner-city with a very good PTH, but it does a lot of errors when predicting non crossing pedestrians. At zebras, many more pedestrians cross the road. The zebra model has therefore learned a tendency to over-predict crossing behaviors and makes many FP errors when applied to a different scenario.

As we showed in our previous work [15], combining multiple models can really improve the quality of the prediction. For that reason, the next section proposes to combine the inner-city model with the zebra model.

V. Context Model Tree

A. Definition

Despite getting a good performance for a single generic model, we see that this might no be the best possible performance for all situations. Therefore we want to combine the two models. We previously introduced a framework to combine multiple models of different specificity [15]. This framework, which is called context model tree (CMT), combines models hierarchically by ordering them from generic to specific. The top node is a generic model which is always activated. In our application, the inner-city model will be the top node of the CMT. The zebra model is a more specific model and will be a sub-node of the inner-city node. This node gets activated only when a zebra crossing is detected at up to 40m distance, and a pedestrian is within 10m distance of the zebra area. In case a zebra crossing is not detected, the zebra node will not be activated. By always having the top node active, the CMT ensures that a traffic participant still gets always predicted.

There can be situations where for a same traffic participant, the two nodes provide opposite prediction. In such a case, the CMT uses a competition function to ensure an accurate final prediction. For each pedestrian that is predicted by the inner-city model, the classifier produces a confidence output c_j. When the sub-node is activated, it applies a competition function between the confidence c_i of its classifier and the confidence c_j sent by the top node.

The confidence values c_j and c_i do not have the same meaning because they are produced by independent classifiers. To make the two models comparable, we first scale each confidence output by dividing it by the best threshold of the corresponding model. The competition function is then a maximum function between both scaled confidences. The sub-node outputs the result of the maximum function as the final prediction:

$$c_f = max((c_j/\theta_j),(c_i/\theta_i)) \quad (2)$$

We predict if a pedestrian is crossing if $c_f >= 1$.

B. Experiment

Dataset: The CMT is tested on the 'ICZ dataset'.

Event Based Evaluation: Fig.8 shows that, at its best threshold, the generic inner-city model has a TPR of 31% and can predict crossing behavior 0.72s in advance (see also Fig.5). In comparison, for the same test dataset, the CMT has a TPR of 62% and can predict crossing behavior 2.59s in advance without any errors.

Fig.9 shows four scenes from our dataset with the corresponding perception results and pedestrian predictions. In the scenes shown in Fig.9a, Fig.9b, and Fig.9d pedestrians are

[5]For this evaluation, the features specific to the zebra model were set to 0 when no zebra crossing was detected.

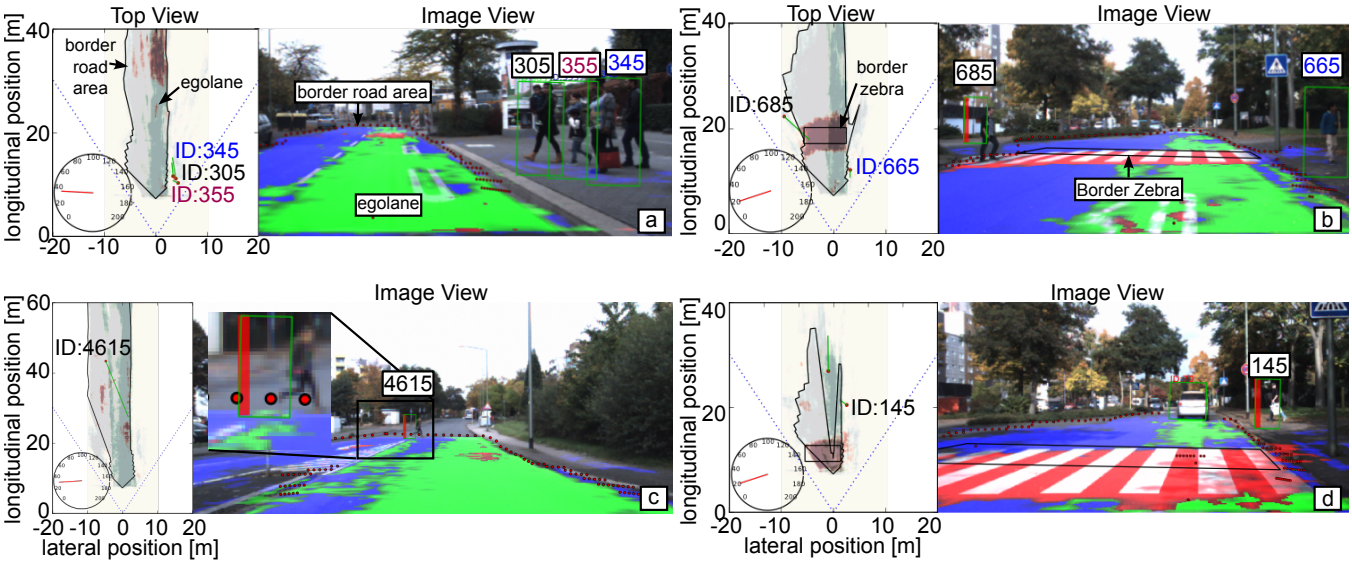


Fig. 9. Four examples of prediction results. The top view (left) shows the ego-vehicle at the origin with its corresponding velocity. The dark blue represents the egolane. The black lane represents the border of the road area. The black rectangle around the red area represents the border of zebra crossings. Red dots represent pedestrians and the green line attached to them their relative velocity. The image view (right) shows the detected pedestrians inside rectangles. The vertical bar in a rectangle means that the pedestrian will cross the road. Pedestrian 685 in b) is an instructed person, all others are regular pedestrians.

detected by the pedestrian recognition algorithm. In Fig.9c, the pedestrian which is about 40m away from the ego-vehicle was manually annotated. In Fig.9a, pedestrians are walking towards the road but they will not cross because they are waiting for a bus to arrive. The CMT correctly predicts that the pedestrians are not going to cross the road. In Fig.9b, the CMT correctly predicts that pedestrian 685 is going to cross the road in front of the ego-vehicle at a zebra crossing and pedestrian 665 is not. In Fig.9c, the CMT predicts that pedestrian 4615 is going to cross the road without priority. This is an example which shows that pedestrians that do not have any priority usually cross the road at a further distance of the ego-vehicle. In Fig.9d, the CMT predicts that pedestrian 145 is going to cross the road. In this scene, the CMT does a wrong prediction because it is an intersection where the ego-vehicle is turning left before the pedestrian intersects the ego-path.

As explained earlier, perception errors have been excluded from the dataset for these experiments. However, we have also evaluated our CMT on the entire dataset containing all excluded examples. The CMT has a TPR of 44% and on average it can predict crossings 2.72s in advance. It also produces prediction errors with a FPR of 6%, with 60% of the false positives due to a wrong pedestrian detection (tree, traffic sign), and 40% of the FPs occurring at intersections. To reduce the latter errors, one possibility would be to create a model specific to intersections and place it as a sub-node of the inner-city node in the CMT.

Discussion: These numbers show that using a CMT to combine multiple contextual models is a very suited solution. It can predict crossing pedestrians in inner-city with very few errors as well as predicting pedestrian at zebra crossings much earlier.

In reality, people do not cross often where they do not have the priority to do so. The CMT allows the combination between a generic model that is only predicting clear cases due to the higher impact of a FP with one that really focuses on predicting crossing behaviors as early as possible.

VI. Conclusion and Outlook

In this paper, we proposed a generic model designed to predict crossing behavior in inner-city and show tests on real inner-city data. Because the generic model had a limited prediction time horizon in specific areas where many people usually cross the road, we have additionally developed a zebra crossing model using more specific contextual information. We have shown that this specific model can predict crossing behaviors with a very good accuracy and a long prediction time horizon. However, it has a limited scope. We applied our context model tree framework to combine both models. The results of this combination of context based models show the benefit of using this framework to improve the accuracy of the prediction as well as the prediction time horizon. Compared with the state of the art, we show an improved time horizon for a good performance. On similar data as [12], we can reach a performance comparable to humans (PTH 0.71s versus 0.5s for the 'non priority' dataset).

We have shown an event based evaluation of the different models. In contrast to the commonly applied frame-based evaluation, an event-based evaluation considers the temporal grouping of system signals and the relevance of a predicted event for the driver of the ego vehicle. Our evaluation was based on the fraction of true positives/false positives. However, in order to analyze the online performances of our system and the acceptability to a driver, we plan to do an evaluation using the frequency of events (FP/TP per hour) (see also [15]). For such experiments, we will do

new recordings in order to build a dataset that contains the real frequency of crossing pedestrians. We might also use a different sensing setup to be able to detect pedestrians at a further distance without the need to manually annotate them.

In order to reduce prediction errors, we will also develop a model to predict crossing behaviors at intersection and add it to the CMT. With an extended dataset, we might find other additional contexts for which prediction could be improved by designing specific models.

As the authors explain in [21], a pedestrians also take into account their distance to surrounding vehicles as well as their respective velocities. In our feature set, we only include the relations to the ego-vehicle, but in the future we want to extend this to other cars in the scene.

REFERENCES

[1] A. Martin, "Factors influencing pedestrian safety: a literature review," Tech. Rep., 2006.

[2] E. Coelingh, A. Eidehall, and M. Bengtsson, "Collision warning with full auto brake and pedestrian prediction - a practical example of automatic emergency braking," in *Proc. IEEE Intelligent Transportation Systems*, 2010, pp. 155–160.

[3] MobilEye Ltd., "C2-270 collision prevention system user manual," Tech. Rep., 2007.

[4] T. Gandhi and M. M. Trivedi, "Pedestrian collision avoidance systems: A survey of computer vision based recent studies," in *Proc. IEEE Intelligent Transportation Systems*, 2006, pp. 976–981.

[5] ——, "Pedestrian protection systems: Issues, survey, and challenges," in *Proc.IEEE Intelligent Transportation Systems*, 2007, pp. 413–430.

[6] B. Musleh, F. Garca, J. Otamendi, J. M. Armingol, and A. De la Escalera, "Identifying and tracking pedestrians based on sensor fusion and motion stability predictions," *Sensors*, vol. 10, no. 9, pp. 8028–8053, 2010.

[7] Y. Abramson and B. Steux, "Harware-friendly pedestrian detection and impact prediction," in *Proc. IEEE Intelligent Vehicles Symposium*, 2004, pp. 590–595.

[8] F. Large, D. Vasquez, T. Fraichard, and C. Laugier, "Avoiding cars and pedestrians using velocity obstacles and motion prediction," in *Proc.IEEE Intelligent Vehicles Symposium*, 2004, pp. 375–379.

[9] Z. Chen and N. H. C. Yung, "Improved multi-level pedestrian behavior prediction based on matching with classified motion patterns," in *Proc.IEEE Intelligent Transportation Systems*, 2009, pp. 1–6.

[10] S. Koehler, M. Goldhammer, S. Bauer, K. Doll, U. Brunsmann, and K. . Dietmayer, "Early detection of the pedestrian's intention to cross the street," in *Proc. IEEE Intelligent Transportation Systems*, 2012, pp. 1759–1764.

[11] S. Koehler, B. Schreiner, S. Ronalter, K. Doll, U. Brunsmann, and K. Zindler, "Autonomous evasive maneuvers triggered by infrastructure-based detection of pedestrian intentions," in *Proc. IEEE Intelligent Vehicles Symposium*, 2013, pp. 519–526.

[12] C. G. Keller and D. M. Gavrila, "Will the pedestrian cross? a study on pedestrian path prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, pp. 494–506, 2014.

[13] M. C. Montel, T. Brenac, M. A. Granié, M. Millot, and C. Coquelet, "Urban environments, pedestrian-friendliness and crossing decisions," in *Proc. Transportation Research Board 92nd Annual Meeting*, 2013, pp. 1–13.

[14] S. Bonnin, T. H. Weisswange, F. Kummert, and J. Schmuedderich, "Accurate behavior prediction on highways based on a systematic combination of classifiers," in *Proc. IEEE Intelligent Vehicles Symposium*, 2013, pp. 242–249.

[15] ——, "General behavior prediction by a combination of scenario-specific models," *IEEE Transactions on Intelligent Transportation Systems*, vol. in press, pp. 1–11, 2014.

[16] A. Gepperth, M. Garcia Ortiz, and B. Heisele, "Real-time pedestrian detection and pose classification on a GPU," in *Proc. IEEE Intelligent Transportation Systems*, 2013, pp. 348–353.

[17] T. H. Weisswange, B. Bolder, J. Fritsch, S. Hasler, and C. Goerick, "An integrated adas for assessing risky situations in urban driving," in *Proc. IEEE Intelligent Vehicles Symposium*, 2013, pp. 292–297.

[18] J. Fritsch, T. Kuehnl, and F. Kummert, "Monocular road terrain detection by combining visual and spatial information," *IEEE Transactions on Intelligent Transportation Systems*, vol. in press, pp. 1–11, 2014.

[19] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2007.

[20] U. Lages, "Automatic scenario generation by advanced offline processing for groundtruth evaluation," Tech. Rep., 2013.

[21] S. Schmidt and B. Faerber, "Pedestrians at the kerb recognising the action intentions of humans," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 12, no. 4, pp. 300–310, 2009.