2018 21st International Conference on Intelligent Transportation Systems (ITSC)
Maui, Hawaii, USA, November 4-7, 2018

Learning in the Curbside Coordinate Frame for a Transferable Pedestrian Trajectory Prediction Model

Nikita Jaipuria*
Dept. of Mechanical Engineering
Massachusetts Institute of Technology
Cambridge, USA
nikitaj@mit.edu

Golnaz Habibi*
Dept. of Aeronautics and Astronautics
Massachusetts Institute of Technology
Cambridge, USA
golnaz@mit.edu

Jonathan P. How
Dept. of Aeronautics and Astronautics
Massachusetts Institute of Technology
Cambridge, USA
jhow@mit.edu

***Abstract*—This paper presents a novel framework for accurate pedestrian trajectory prediction in intersection corners or near crosswalks. Given prior knowledge of curbside geometry (i.e. angle made by intersecting curbs at the corner point of interest and the coordinates of the corner itself), the presented framework can accurately predict pedestrian trajectories even in new, unseen intersections. This is achieved by learning motion primitives in a common frame, called the *curbside coordinate frame*. A key insight in developing this common frame is to ensure that trajectories from intersections with different geometries, representing the same behavior, are spatially similar in the common frame. Motion primitives learned in such a common frame, can then be easily generalized to predict in new intersections, with different geometries than the ones trained on. We test our algorithm on real pedestrian trajectory datasets collected at two intersections, with distinctly different curbside and crosswalk geometries. A comparison of our algorithm with [1] demonstrates improved prediction accuracies of pedestrian trajectory prediction in the case of same training and test intersections, and the improvement of accuracy in the most different training and test intersections scenarios. The result also shows additional context, such as information about pedestrian traffic lights, if available, can be easily incorporated in our prediction model for further improvement in prediction accuracy.**

***Index Terms*—Pedestrian motion prediction, skewed coordinate system, Contravariant components, affine transformation, motion primitives, common frame, transferable model, Gaussian Process, sparse coding**

I. Introduction

Increased safety of road travelers and a consequent reduction in road accident fatality rate has been the main driver of research on advanced driver-assistance systems (ADAS) and self-driving cars. Recent advances in computation power and an increase in the amount of publicly available training datasets provided a boost to the application of state-of-the-art machine learning approaches in this field.

Safe and reliable operation of self-driving cars in busy, urban scenarios requires interaction with multiple moving agents like cars, cyclists and pedestrians. Motion prediction of pedestrians is more challenging than that of cars (and to some extent, cyclists) because of the absence of uniform pedestrian "rules of the road" like staying within road boundaries, following lanes etc. The few pedestrian "rules" that exist are often unclear and frequently violated. The problem is further complicated in intersection scenarios, where additional context, such as tightly packed sidewalks and traffic lights or stop signs, influences pedestrian movement. Furthermore, intersection geometry, such as position and shape of curbsides, also significantly influences pedestrian movement. It is practically impossible to collect data, and train on every different type of intersection geometry. Therefore, there exists a need for a general prediction algorithm, which when trained on one intersection, can be generalized to

*These authors contributed equally

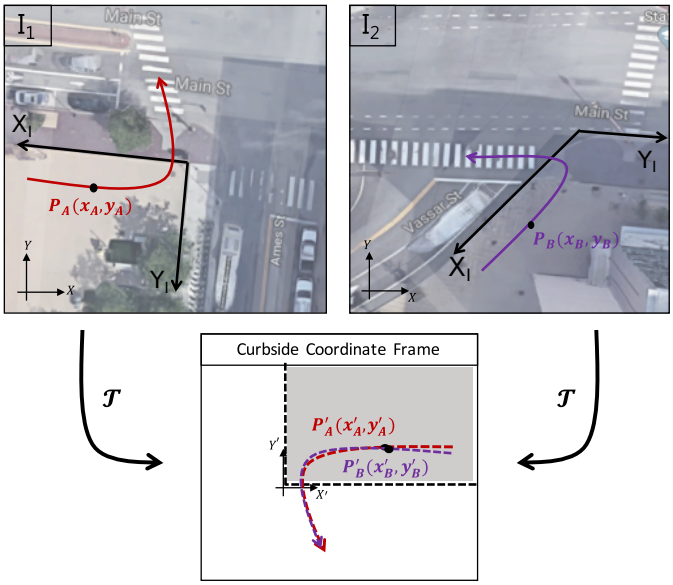


Figure 1: An illustration to show how points $P_A(x_A, y_A)$ on the red trajectory in intersection $\mathbf{I_1}$ and $P_B(x_B, y_B)$ on the purple trajectory in intersection $\mathbf{I_2}$, under the transformation $\mathscr{T}$, map to points $P'_A(x'_A, y'_A)$ and $P'_B(x'_B, y'_B)$ in the curbside coordinate frame. We show that $\mathscr{T}$ is in general an affine transformation. Since pedestrian trajectories in urban intersections are significantly constrained by the curbsides, transforming them into the curbside coordinate frame using an affine transformation, intuitively would map trajectories with similar pedestrian motion behavior approximately on top of each other in the curbside coordinate frame. This insight helps in developing a general trajectory prediction model.

978-1-7281-0323-5/18/$31.00 ©2018 IEEE

predict in new, unseen intersections, with similar situational context, but varying geometries.

Chen et al. [2] combine the merits of Markovian-based and clustering-based techniques to show significant improvement over state-of-the-art clustering methods for pedestrian trajectory prediction. However, their approach fails to incorporate context and is based on motion primitives learned using spatial features (x,y position in the car frame) specific to the training environment. Most of the previous work on context-based pedestrian intent recognition is limited to the identification of stopping versus crossing intent [3]–[9], as opposed to long term trajectory prediction. The latter can can provide additional, useful information to the self-driving vehicle for planning its future course of action. Such as how much time would it take for someone to cross the road, which crosswalk/path would be used, etc. Furthermore, the use of spatial context features like orthogonal distance to curbside [6]–[8] makes these intent classification models directly dependent on the specific training intersection geometry, and prevents generalization to new intersections with varying curbside geometries. Bonnin et al. [10] developed a more generic, context-based, multi-model system for predicting crossing behavior in inner-city situations and zebra crossings. However, the output of their prediction model is again a crossing probability as opposed to predicted future trajectory.

Coscia et al. [11] forecast long-term behavior of pedestrians by making use of past observed patterns and semantic segmentation of a bird's eye view of the scene. Such an approach, when applied in the real world, would require accurate high definition semantic maps of each scene, which are expensive to create and maintain. It is also unclear if their prediction model can be generalized to new, unseen scenes. Ballan et al. [12] and Sadeghian et al. [13] follow a similar approach to path prediction while also demonstrating the ability to "transfer knowledge". However, a prior bird's eye view of the scene is needed for both these approaches as well. Our approach encodes situational context and provides a transferable prediction model, which can be generalized to predict in corners of new, unseen intersections, without needing a high-definition prior map. The key idea that enables this generalization is the use of a simple prior on curbside geometry (i.e. angle made by intersecting curbs at the corner point of interest and the coordinates of the corner) to construct a common "curbside coordinate frame", such that trajectories with similar intent are spatially similar in this common frame (see Fig. 1). While high-definition prior maps are not a limiting constraint for the application of the proposed method, if available, the presented framework is general enough to incorporate context information embedded in such maps.

[1] introduced a transferable pedestrian prediction model at intersections using inverse reinforcement learning (IRL). In their work, the trajectory is predicted by inferring the pedestrian intent at intersection. However, goal locations as the potential intents need to be selected based on the semantic contexts. Moreover, this approach requires a map including some semantic contexts around the intersection to construct the feature vector for the learning model. We evaluate our proposed algorithm in two different intersections and compare it with previous work [1]. Our algorithm outperforms the previous approach when the trained and test environment are the same. When the trained and test intersections are different, our approach performs better than [1] except one case.

The main contributions of this work are as follows: (i) Introduction of the **"curbside coordinate frame"** as a common frame in which spatially dissimilar trajectories from different intersections, representing the same underlying pedestrian intent, are spatially similar (see Fig. 1); (ii) Introduction of a **novel representation of distance to curbside** as the contravariant components of pedestrian positions in the curbside coordinate frame (can be orthogonal or skewed). This representation ensures that distance to curbside, as a context feature, is independent of intersection geometry (as opposed to other representations such as orthogonal distance to curbside); (iii) Proof of the fact that the **transformation of trajectories from the car frame, into the curbside coordinate frame, is affine**. Such a transformation, therefore, preserves properties such as collinearity, parallelism etc. across intersections while encoding context (see Fig. 1); (iv) **Transferable Augmented Semi Nonnegative Sparse Coding (TASNSC)**, as a context-based pedestrian trajectory prediction model for accurate, long term (≈ 5 seconds) prediction in corners of new, unseen intersections with similar semantic cues as the ones that the model is trained on.

II. Preliminaries

In this section, we first briefly review the Augmented Semi-Nonnegative Sparse Coding (ASNSC) algorithm. Out of other prior approaches [14]–[17], ASNSC was chosen for learning motion primitives in this work, because of its ease of direct application to the task of pedestrian trajectory prediction [2], [18]. This is followed by a review of covariant and contravariant components of a position vector in a skewed coordinate system.

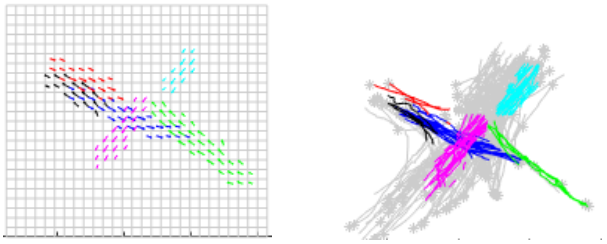
A. Augmented Semi-Nonnegative Sparse Coding (ASNSC)

Given a training dataset of pedestrian trajectories in a discretized world, i.e. $\mathscr{D} : \{t_i\}$, ASNSC learns a set of L motion primitives, i.e. $B : \{\mathbf{m}_1, \ldots, \mathbf{m}_L\}$ (see Fig. 2(a)). Here, a 'trajectory' is a sequence of two-dimensional position measurements, taken at a fixed time interval. t_i is the vectorized representation of the i-th trajectory, s.t. $t_i \in \mathbb{R}^K$, where K is the number of cells in the discretized world, and the j-th element of t_i, is its average velocity in the j-th cell of the discretized world.

As shown in Fig. 2(b), B is used to segment the original training trajectories into clusters, where each cluster is best explained by one of the learned motion primitives. A transition matrix ($\mathbf{R}$) is thus created, where $\mathbf{R}(i, j)$ is the set of trajectories exhibiting a transition from $\mathbf{m}_i$ to $\mathbf{m}_j$ for off-diagonal elements, and the set of trajectories ending in $\mathbf{m}_i$ for the diagonal elements. Each transition, i.e. $\{\mathbf{m}_i, \mathbf{m}_j | \mathbf{R}(i, j) \neq \emptyset\}$, is modeled as a two-dimensional Gaussian Process (GP) flow field [19], [20]. In particular, two independent GPs, (GP_x, GP_y), are used to learn

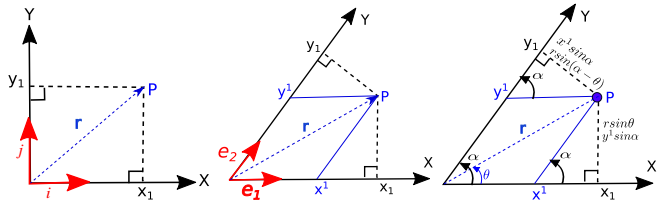
a mapping from the two-dimensional position features to the x and y velocities respectively, for each transition.

Given an observed trajectory (t_o), if the unitary motion pattern that most likely generated t_o is given by $\{\mathbf{m}_p|\mathbf{R}(p,p) \neq \emptyset\}$, a set of predicted trajectories, along with the likelihood of each, is given by the motion patterns starting in $\mathbf{m}_p$, i.e. $\{\mathbf{m}_p, \mathbf{m}_q|\mathbf{R}(p,q) \neq \emptyset\} \ \forall \ q$.



(a) (b)

Figure 2: (a) Each color represents a motion primitive ($\mathbf{m}_k$) learned using ASNSC in a discretized world; (b) Segmentation of training trajectories (in gray) into clusters, where each cluster is best explained by the motion primitive of the same color in (a).



(a) (b) (c)

Figure 3: (a) Orthogonal coordinate system; (b) Skewed coordinate system; (c) Calculation of contravariant components in a skewed coordinate system using trigonometry

B. Covariant and contravariant components of vectors in a skewed coordinate system

As shown in Fig. 3(a) and Fig. 3(b), a coordinate system can be either orthogonal (represented by unit vectors $\hat{i}, \hat{j}$) or skewed (represented by unit vectors $\hat{e}_1, \hat{e}_2$). In an orthogonal coordinate system, covariant and contravariant components of a position vector are perfectly aligned. A position vector in such a system has only one representation i.e. $\vec{r} = x_1\hat{i} + y_1\hat{j}$ (see Fig. 3(a)). In a skewed coordinate system, the covariant components (x_1, y_1) and contravariant components (x^1, y^1) of a position vector do not align. The same position vector, in such a system, can be represented using both its covariant and contravariant components. Representing it using the contravariant components is more standard since this representation is compatible with the rule of vector sums, i.e. $\vec{r} = x^1\hat{e}_1 + y^1\hat{e}_2$ (see Fig. 3(b)). However, since $(\hat{e}_1 \cdot \hat{e}_2) \neq 0$ in a skewed coordinate system, $r^2 \neq (x^1)^2 + (y^1)^2$ in general. As shown in Fig. 3(c), basic trigonometric identities can instead be used for computing the contravariant components of a position vector in a skewed coordinate system (Note that $0 < \alpha < \pi$ in the following equations).

$$x^1 = r\sin(\alpha - \theta)/\sin\alpha \tag{1}$$

$$y^1 = r\sin\theta/\sin\alpha \tag{2}$$

To meet the objective of pedestrian trajectory prediction in urban intersections, where curbside geometry significantly constraints pedestrian motion, learning motion primitives and their transition in the curbside coordinate frame $X'Y'$ (instead of an arbitrarily placed, car frame XY, as shown in Fig. 1), can help improve prediction accuracies because of the addition of context. Furthermore, the curbside coordinate frame provides a *common frame* in which trajectories from different intersections, with the same underlying intent, are spatially similar. Thus, learning motion primitives in this common frame aids in developing a context-aware prediction model that can be generalized to predict in any intersection.

III. Algorithm

As discussed earlier, designing a general, transferable prediction model requires encoding features that are independent of the specific training intersection geometry. In this section, we show that any point on a pedestrian trajectory, when mapped from the original, arbitrarily placed car frame, into the common curbside coordinate frame, using a transformation/mapping function as defined in the following text, undergoes an affine transformation. The choice of the curbside coordinate frame, as the common frame into which trajectories from different intersections are mapped, can be justified by the fact that pedestrian trajectories are significantly constrained by curbsides in intersections. Since an affine transformation preserves properties such as collinearity, ratios of distances, parallelism etc., the situational context of pedestrian trajectories, i.e. shape and relative distance with respect to curbsides, is preserved under such a transformation (see Fig. 4 and Fig. 5).

Definition 1. *A coordinate frame with its origin at the intersection corner of interest, and its axes along the two curbsides intersecting at the chosen corner, is defined as the "intersection frame" (X_IY_I in $\boldsymbol{I_1}$ and $\boldsymbol{I_2}$ in Fig. 1). An intersection frame can be either orthogonal or skewed.*

Definition 2. *Given a point $P(x,y)$, on an observed trajectory t_o, in the arbitrarily placed car frame (i.e. XY frame in $\boldsymbol{I_1}$ and $\boldsymbol{I_2}$ in Fig. 1), let us define a transformation $\mathscr{T} : t_o \rightarrow t'_o$ in a common frame C (see Fig. 1), s.t. $P(x,y) \rightarrow P'(x',y')$, where x', y' are the contravariant components of P' in the intersection frame. We call this common frame C as the "curbside coordinate frame".*

Lemma 1. *$\mathscr{T}$ is an affine transformation*

Proof. Given the original, orthogonal car frame O and an intermediate, helper coordinate system H (also orthogonal but with its origin at the intersection corner and its x-axis parallel to the x-axis of intersection corner axis I) T_O^H and T_H^C

represent the coordinate transformation from O to H and H to C respectively, then $\mathscr{T} = T_H^C T_O^H$.

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \mathscr{T} \begin{pmatrix} x \\ y \end{pmatrix} = T_H^C T_O^H \begin{pmatrix} x \\ y \end{pmatrix} \tag{3}$$

Since, T_O^H is simply a combination of rotation and translation, it is an affine transformation. Let us now assume that the original point $P(x,y)$ in O maps to $P^*(x^*,y^*)$ in H, and r is the distance of P^* from the origin: $(x^*)^2 + (y^*)^2 = r^2$. Note that, by definition, the origin and x-axis of H overlap with the origin and x-axis of C. From Fig. 3(c), if θ is the angle made by the position vector with the x-axes,

$$x^* = r\cos\theta, y^* = r\sin\theta \tag{4}$$

Therefore, from (1), (2) and (4), if α, $0 < \alpha < \pi$ is the angle between the intersecting curbsides, $P'(x',y')$ can be written as

$$x' = (r\cos\theta\sin\alpha - r\sin\theta\cos\alpha)/\sin\alpha \tag{5}$$

$$\implies x' = x^* - y^*/\tan\alpha \tag{6}$$

$$y' = r\sin\theta/\sin\alpha = y^*/\sin\alpha \tag{7}$$

Note that (6), (7) can be combined and written in matrix form as

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = T_H^C \begin{pmatrix} x^* \\ y^* \end{pmatrix} = \begin{pmatrix} 1 & -1/\tan\alpha \\ 0 & 1/\sin\alpha \end{pmatrix} \begin{pmatrix} x^* \\ y^* \end{pmatrix} \tag{8}$$

For a given α, T_H^C linearly maps (x^*,y^*) to (x',y') and hence, is an affine transformation. Furthermore, since T_O^H and T_H^C are both affine transformations, $\mathscr{T}$ is also an affine transformation by (3). For the special case which of an intersection with orthogonal curbsides, $\alpha = \pi/2$, T_H^C is the identity matrix, and hence, transformed position vector components in the common curbside frame are the same as the original position vector components in the car frame. □

Since $\mathscr{T}$ is affine, all general properties of an affine transform hold under $\mathscr{T}$, i.e.

- Collinearity is preserved
- Parallel lines remain parallel
- Convexity of sets is preserved
- Ratios of distances are preserved i.e. the midpoint of a line segment remains the midpoint of the transformed line segment

As discussed earlier, since the objective of this paper is pedestrian motion estimation in urban intersections, which is highly constrained by curbside geometry, mapping pedestrian trajectories into the curbside coordinate frame helps in representing trajectories from intersections with different geometries in a common frame. This aids in building a context-aware, general prediction model.

Algorithm 1 describes TASNSC as a transferable version of the ASNSC algorithm. Given the curbside unit vectors $(\hat{e_1},\hat{e_2})$ of the training intersection, $\mathscr{T}$ is used to map the training trajectories from the local, arbitrary placed car frame into the common curbside coordinate frame. Motion primitives are then learned in the curbside coordinate frame using ASNSC (line 4). For trajectory prediction in an unseen intersection, first the observed trajectory is transformed into the common curbside coordinate frame using $\mathscr{T}$ (line 5). Motion primitives and their transitions, learned in the common curbside coordinate frame, are then used for prediction, followed by a transformation of the predicted trajectory into the original, car frame of the test intersection (line 6). Algorithm 2 describes the procedure for transformation of pedestrian trajectories under $\mathscr{T}$. Fig. 4 and Fig. 5 show the transformation of trajectories into the curbside coordinate frame under $\mathscr{T}$ for an orthogonal and skewed coordinate system respectively.

Algorithm 1: Transferable ASNSC (TASNSC)

input : curbside unit vectors of training intersection $(\vec{e_1},\vec{e_2})$, set of training trajectories in car frame $(\mathscr{D}_c)$, curbside unit vectors of test intersection $(\vec{e_1'},\vec{e_2'})$ and observed trajectory t_0

output : predicted the future trajectory in car frame (t_p)

/* Training Phase */

1 $\mathscr{D} = \{\}$

2 **for** $\forall\ t_i \in \mathscr{D}_c$ **do**

/* map training trajectories to curbside frame */

3 $t_i' = \mathscr{T}(\vec{e_1},\vec{e_2},t_i)$ $\mathscr{D} \leftarrow \{\mathscr{D}, t_i'\}$

/* learn motion primitives in curbside frame */

4 $\{B,S\} = \text{ASNSC}(\mathscr{D})$

/* Prediction Phase */

/* map observed trajectory to curbside frame */

5 $t_o' = \mathscr{T}(\vec{e_1'},\vec{e_2'},t_o)$

/* set of predicted trajectories in curbside frame */

6 $t_p' = \text{predict}(B,t_o')$ $t_p = \mathscr{T}^{-1}(t_p')$

7 **return** t_p

Algorithm 2: Transformation $\mathscr{T}$

input : curbside unit vectors $(\vec{e_1},\vec{e_2})$ and trajectory in car frame t_i

output : trajectory in common curbside frame t_i'

1 $\alpha \leftarrow cos^{-1}(\vec{e_1}.\vec{e_2})$; // angle between $\vec{e_1}$ and $\vec{e_2}$

2 **for** $\forall\ P_j(x_j,y_j) \in t_i$ **do**

3 ${x_j}' \leftarrow r\sin(\alpha-\theta)/\sin\alpha$; // Fig. 3(c), $0 \le \theta \le 2\pi$

4 ${y_j}' \leftarrow r\sin\theta/\sin\alpha$; // $0 < \alpha < \pi$

5 **return** $t_i' = \{({x_j}',{y_j}')\}$

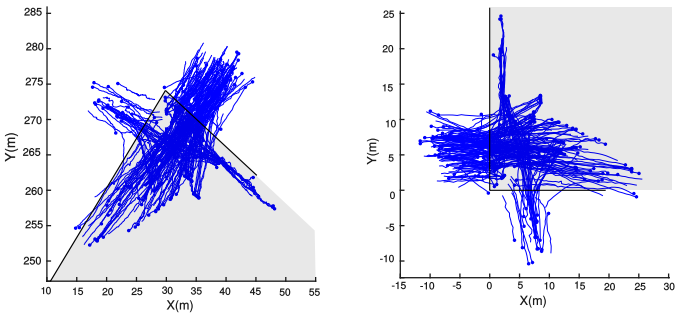


Figure 4: Original (left) and transformed trajectories in the curbside coordinate frame (right) under the transformation $\mathscr{T}$, when the curbs are orthogonal to each other. Trajectories are shown in blue and shaded gray area denotes the sidewalk.

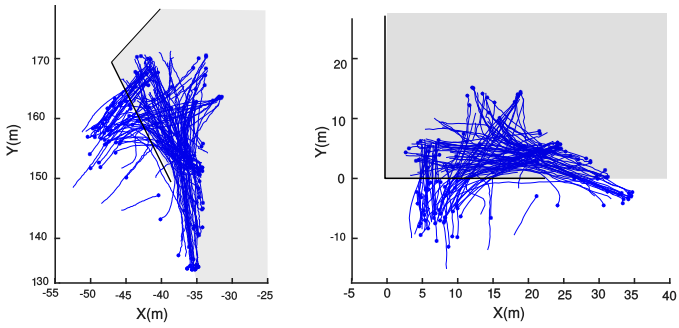


Figure 5: Original (left) and transformed trajectories in the curbside coordinate frame (right) under the transformation $\mathscr{T}$, when the curbs are skewed. Trajectories are shown in blue and shaded gray area denotes the sidewalk.

IV. Results

A. Dataset description

We test our algorithm on real pedestrian data collected by a Polaris GEM vehicle equipped with three Logitech C920 cameras and a SICK LMS151 LIDARa [21], [22]. A prior occupancy grip map of the environment, created using the on-board LIDARs, is used to extract curbside boundaries. However, as long as the intersection corner is not crowded by obstructions such as trees, it is possible to detect the curbside online. Real pedestrian trajectories are collected in two different intersections (see Fig. 7). The dataset collected in intersection $\mathbf{I_1}$, with nearly orthogonal curbsides, consists of 186 training and 32 test trajectories while that collected in intersection $\mathbf{I_2}$, with skewed curbsides, consists of 114 training and 22 test trajectories. An observation history of 2.5 seconds prior to the pedestrian entering the intersection is used to predict 5 seconds ahead in time.

B. Experiment details

Two experiments were conducted for evaluating the prediction performance of TASNSC. In the first experiment, the training and test intersections are the same. While in the second experiment, the training and test intersections are different. The prediction performance of TASNSC in both these experiments is compared with a prior, inverse reinforcement learning based prediction model that has been shown to generalize to new, unseen intersections [1]. (We call this model 'Transferable IRL' in the rest of the paper for brevity).

Fig. 6 shows qualitative results of prediction performance of TASNSC in different scenarios. The output of TASNSC, similar to ASNSC, is a set of predicted trajectories with the likelihood of each. Such an output is desired for application of state-of-art uncertainty aware probabilistic planners [23]. However, in all the plots in Fig. 6, only the predicted trajectory with the maximum likelihood is shown for clarity.

We use three metrics to evaluate prediction performance: (i) computation time for predicting a new trajectory; (ii) *Modified Hausdorff Distance (MHD)* [24] is an object-matching metric and is used to compare predicted trajectories with ground truth. More simply, it measures the average distance between points on the predicted trajectory and actual observed trajectory, for the same time stamps; (iii) *distLast* represents the distance between the last point on the prediction trajectory and the ground truth.

Since TASNSC provides a set of predicted trajectories with the likelihood of each, for both *MHD* and *distLast*, if a set of n trajectories is predicted as $\{\mathbf{t}_1,\ldots,\mathbf{t}_n\}$, with their likelihood of prediction given by $\{l_1,\ldots,l_n\}$, and their prediction error (either MHD or distlast) denoted by e_i, the weighted prediction error is computed as follows:

$$\text{Weighted prediction error} = \frac{\sum_{i=1}^{n} l_i e_i}{\sum_{k=1}^{n} l_k} \tag{9}$$

As is clear from the comparison in Table I, TASNSC outperforms Transferable IRL [1] in all scenarios, when trained and tested on the same intersection. Furthermore, TASNSC performs better even when trained and tested on different intersections, in all scenarios except for one, in which the training environment is an orthogonal intersection ($\mathbf{I_1}$) and the model predicts on a skewed intersection ($\mathbf{I_2}$). This is a challenging scenario and the better performance of Transferable IRL in this case can be attributed to its use of a set of semantic labels as context features. While TASNSC, on the other hand, predicts with a prior on curbside geometry only. In all scenarios, TASNSC is 4-10 times faster than Transferable IRL in predicting the future trajectory.

Algorithm	Train on	Test on	MHD (m)	Time (sec)
Transferable IRL	$\mathbf{I_1}$	$\mathbf{I_1}$	1.44	0.61
TASNSC	$\mathbf{I_1}$	$\mathbf{I_1}$	0.65	0.08
Transferable IRL	$\mathbf{I_2}$	$\mathbf{I_1}$	1.67	0.61
TASNSC	$\mathbf{I_2}$	$\mathbf{I_1}$	1.28	0.06
Transferable IRL	$\mathbf{I_1}$	$\mathbf{I_2}$	1.70	0.35
TASNSC	$\mathbf{I_1}$	$\mathbf{I_2}$	1.79	0.07
Transferable IRL	$\mathbf{I_2}$	$\mathbf{I_2}$	2.03	0.35
TASNSC	$\mathbf{I_2}$	$\mathbf{I_2}$	1.27	0.09

Table I: Quantitative performance comparison of TASNSC with transferable IRL [1]

Algorithm	Train on	Test on	MHD (m)	distLast (m)
TASNSC	$\mathbf{I_1}$	$\mathbf{I_2}$	1.72	4.41
(with traffic light)	$\mathbf{I_2}$	$\mathbf{I_1}$	0.80	2.14
TASNSC	$\mathbf{I_1}$	$\mathbf{I_2}$	1.79	4.63
(without traffic light)	$\mathbf{I_2}$	$\mathbf{I_1}$	1.28	2.98

Table II: Quantitative performance comparison of TASNSC with and without pedestrian traffic light added as a context feature

C. Pedestrian traffic Light as context

Adding context, such as pedestrian traffic light status, in the GP based motion patterns representing transition between motion primitives, boosts prediction performance of TASNSC [25].

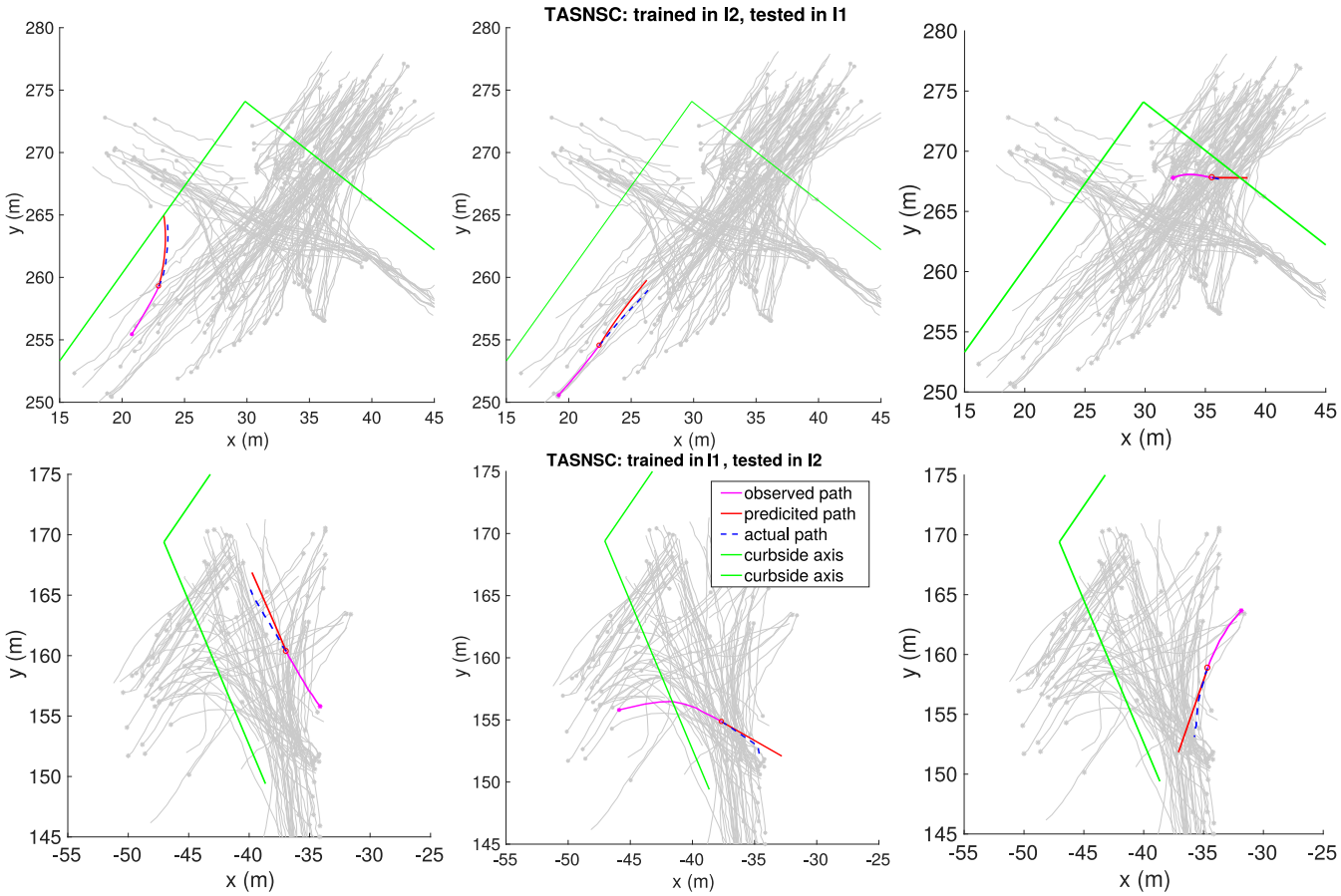


Figure 6: Prediction results of TASNSC in intersection $\mathbf{I_1}$ when the model is trained in $\mathbf{I_2}$ (top) and results of TASNSC in intersection $\mathbf{I_2}$ when the model is trained in $\mathbf{I_1}$ (bottom), Ground truth is shown in dotted blue, observed trajectory in pink & predicted trajectory in red. The result shows TASNSC performs well in different scenarios. In Figure top right scenario, TASNSC correctly estimates the direction of future trajectory. However, TASNSC does not predict the pedestrian stops at intersection as it does. In general TASNSC only learns the orientation of the velocity, not velocity magnitude. Considering the velocity profile can improve the prediction which is left for the future work (see section V)

Furthermore, the best prediction performance, in terms of both *MHD* and *distLast* is achieved by TASNSC trained in intersection $\mathbf{I_1}$ and tested on $\mathbf{I_2}$ and vice versa (see Table II).

V. Conclusion

This paper has introduced a common frame for pedestrian trajectory prediction, called the "curbside coordinate frame". Pedestrian trajectories from intersections of different shapes, representing the same behavior, are spatially similar when mapped into this common frame. We leverage curbside coordinate frame to develop TASNSC as a general, accurate pedestrian trajectory prediction model for urban intersections. Training trajectories are first mapped into the proposed common frame. The ASNSC framework is then used for learning motion primitives and subsequently, modeling the transition between these learned primitives in the common frame. The motion primitives and their transition, thus learned, not only encode situational context in the form of distance to curbside, but are also agnostic to the specific training intersection geometry. Such motion primitives, can therefore, be used for prediction in new, unseen intersections with different curbside geometries by mapping the observed trajectory into the common frame. We test our algorithm on two different intersections, one with almost orthogonal curbsides and the other with skewed curbsides. TASNSC significantly outperforms IRL approach [1], when trained and tested on the same intersection. When trained and tested on different intersections, TASNSC has a better prediction than IRL in all scenarios except for one. Addition of traffic light as an additional context feature in the GP based transition models helps further boost prediction performance of TASNSC.

Our approach is limited by the need for a prior on curbside geometry. Vectorized road maps are an easy and low-maintenance source of obtaining the required prior. Additionally, one might also argue that curbsides can be detected on-line as the vehicle approaches an intersection corner of interest, observability can be an issue because of occlusions and/or a limited field of view of on-board perception sensors. Therefore, there is a need to explore the robustness of our prediction

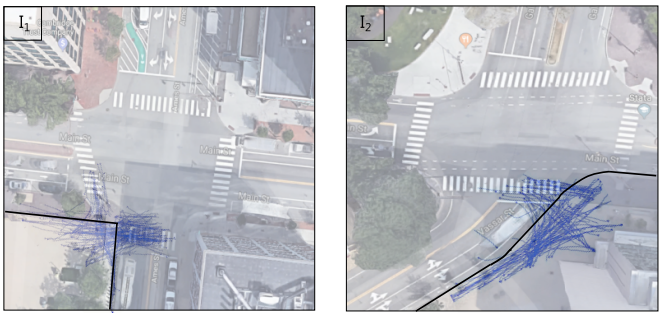
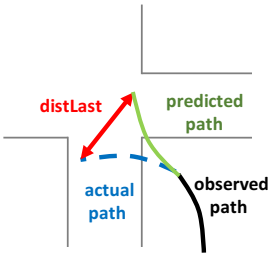


Figure 7: An overhead snapshot of intersection $\mathbf{I_1}$ with orthogonal curbsides (left) and intersection $\mathbf{I_2}$ with skewed curbsides (right). The training dataset, shown in blue, consists of pedestrian trajectories collected using a 3D LIDAR and camera on-board a Polaris GEM vehicle parked at the intersection corners.

Figure 8: An illustration of distLast which is defined as the distance between the end point of ground truth and predictied trajectory



models to uncertainty in curbside geometry. Furthermore, interaction among pedestrians is not considered in the presented TASNSC framework and will be part of future work. Lastly, incorporating the speed profile of pedestrians would help further improve prediction performance, and will again be part of future work.

Acknowledgment

Special thanks to Anthony Colangeli, Jhustin Miller, and Michael Everett for their tremendous help in collecting and annotating data. This project is funded by Ford Motor Company.

References

[1] M. Shen, G. Habibi, and J. P. How, "Transferable Pedestrian Motion Prediction Models at Intersections," *ArXiv e-prints*, Mar. 2018.

[2] Y. F. Chen, M. Liu, and J. P. How, "Augmented dictionary learning for motion prediction," in *Robotics and Automation (ICRA), 2016 IEEE International Conference on*. IEEE, 2016, pp. 2527–2534.

[3] A. T. Schulz and R. Stiefelhagen, "Pedestrian intention recognition using latent-dynamic conditional random fields," in *Intelligent Vehicles Symposium (IV), 2015 IEEE*. IEEE, 2015, pp. 622–627.

[4] A. J. Gonzalez, W. J. Gerber, R. F. DeMara, and M. Georgiopoulos, "Context-driven near-term intention recognition," *The Journal of Defense Modeling and Simulation*, vol. 1, no. 3, pp. 153–170, 2004.

[5] F. Schneemann and P. Heinemann, "Context-based detection of pedestrian crossing intention for autonomous driving in urban environments," in *Intelligent Robots and Systems (IROS), 2016 IEEE/RSJ International Conference on*. IEEE, 2016, pp. 2243–2248.

[6] J. F. P. Kooij, N. Schneider, F. Flohr, and D. M. Gavrila, "Context-based pedestrian path prediction," in *European Conference on Computer Vision*. Springer, 2014, pp. 618–633.

[7] B. Völz, K. Behrendt, H. Mielenz, I. Gilitschenski, R. Siegwart, and J. Nieto, "A data-driven approach for pedestrian intention estimation," in *Intelligent Transportation Systems (ITSC), 2016 IEEE 19th International Conference on*. IEEE, 2016, pp. 2607–2612.

[8] B. Völz, H. Mielenz, G. Agamennoni, and R. Siegwart, "Feature relevance estimation for learning pedestrian behavior at crosswalks," in *Intelligent Transportation Systems (ITSC), 2015 IEEE 18th International Conference on*. IEEE, 2015, pp. 854–860.

[9] Y. Hashimoto, Y. Gu, L.-T. Hsu, and S. Kamijo, "A probabilistic model for the estimation of pedestrian crossing behavior at signalized intersections," *2015 IEEE 18th International Conference on Intelligent Transportation Systems*, pp. 1520–1526, 2015.

[10] S. Bonnin, T. H. Weisswange, F. Kummert, and J. Schmüdderich, "Pedestrian crossing prediction using multiple context-based models," in *Intelligent Transportation Systems (ITSC), 2014 IEEE 17th International Conference on*. IEEE, 2014, pp. 378–385.

[11] P. Coscia, F. Castaldo, F. A. Palmieri, A. Alahi, S. Savarese, and L. Ballan, "Long-term path prediction in urban scenarios using circular distributions," *Image and Vision Computing*, vol. 69, pp. 81–91, 2018.

[12] L. Ballan, F. Castaldo, A. Alahi, F. Palmieri, and S. Savarese, "Knowledge transfer for scene-specific motion prediction," in *European Conference on Computer Vision*. Springer, 2016, pp. 697–713.

[13] A. Sadeghian, F. Legros, M. Voisin, R. Vesel, A. Alahi, and S. Savarese, "Car-net: Clairvoyant attentive recurrent network," *arXiv preprint arXiv:1711.10061*, 2017.

[14] P. P. Choi and M. Hebert, "Learning and predicting moving object trajectory: a piecewise trajectory segment approach," *Robotics Institute*, p. 337, 2006.

[15] C. Sung, D. Feldman, and D. Rus, "Trajectory clustering for motion prediction," in *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on*. IEEE, 2012, pp. 1547–1552.

[16] J.-G. Lee, J. Han, and K.-Y. Whang, "Trajectory clustering: a partition-and-group framework," in *Proceedings of the 2007 ACM SIGMOD international conference on Management of data*. ACM, 2007, pp. 593–604.

[17] C. Vollmer, S. Hellbach, J. Eggert, and H.-M. Gross, "Sparse coding of human motion trajectories with non-negative matrix factorization," *Neurocomputing*, vol. 124, pp. 22–32, 2014.

[18] Y. F. Chen, "Predictive modeling and socially aware motion planning in dynamic, uncertain environments," *Thesis*, 2017.

[19] J. Joseph, F. Doshi-Velez, A. S. Huang, and N. Roy, "A bayesian nonparametric approach to modeling motion patterns," *Autonomous Robots*, vol. 31, no. 4, p. 383, 2011.

[20] G. S. Aoude, B. D. Luders, J. M. Joseph, N. Roy, and J. P. How, "Probabilistically safe motion planning to avoid dynamic obstacles with uncertain motion patterns," *Autonomous Robots*, vol. 35, no. 1, pp. 51–76, 2013.

[21] J. Miller and J. P. How, "Predictive positioning and quality of service ridesharing for campus mobility on demand systems," in *Robotics and Automation (ICRA), 2017 IEEE International Conference on*. IEEE, 2017, pp. 1402–1408.

[22] J. Miller, A. Hasfura, S.-Y. Liu, and J. P. How, "Dynamic arrival rate estimation for campus mobility on demand network graphs," in *Intelligent Robots and Systems (IROS), 2016 IEEE/RSJ International Conference on*. IEEE, 2016, pp. 2285–2292.

[23] D. Bernardini and A. Bemporad, "Scenario-based model predictive control of stochastic constrained linear systems," in *Decision and Control, 2009 held jointly with the 2009 28th Chinese Control Conference. CDC/CCC 2009. Proceedings of the 48th IEEE Conference on*. IEEE, 2009, pp. 6333–6338.

[24] M.-P. Dubuisson and A. K. Jain, "A modified hausdorff distance for object matching," in *Pattern Recognition, 1994. Vol. 1-Conference A: Computer Vision & Image Processing., Proceedings of the 12th IAPR International Conference on*, vol. 1. IEEE, 1994, pp. 566–568.

[25] N. Jaipuria, G. Habibi, and J. P. How, "Casnsc: A context-based approach for accurate pedestrian motion prediction at intersections," in *NIPS Machine Learning for Intelligent Transportation Systems Workshop (MLITS)*, 2017.