2018 IEEE Intelligent Vehicles Symposium (IV)
Changshu, Suzhou, China, June 26-30, 2018

Predicting trajectories of vehicles using large-scale motion priors

Suraj M S[1,2], Hugo Grimmett[1], Lukáš Platinský[1] and Peter Ondrúška[1]

Abstract— **We present a simple yet effective paradigm to accurately predict the future trajectories of observed vehicles in dense city environments. We equipped a large fleet of cars with cameras and performed city-scale structure-from-motion to accurately reconstruct 10M positions of their trajectories spanning over 1000h of driving.**

We demonstrate that this information can be used as a powerful high-fidelity prior to predict future trajectories of newly observed vehicles in the area without the need for any knowledge of road infrastructure or vehicle motion models. By relating the current position of the observed car to a large dataset of the previously exhibited motion in the area we can directly perform prediction of its future position.

We evaluate our method on two large-scale data sets from San Francisco and New York City and demonstrate an order of magnitude improvement compared to a linear-motion based method. We also demonstrate that the performance naturally improves with the amount of data and ultimately yields a system that can accurately predict vehicle motion in challenging situations across extremes in traffic, time, and weather conditions.

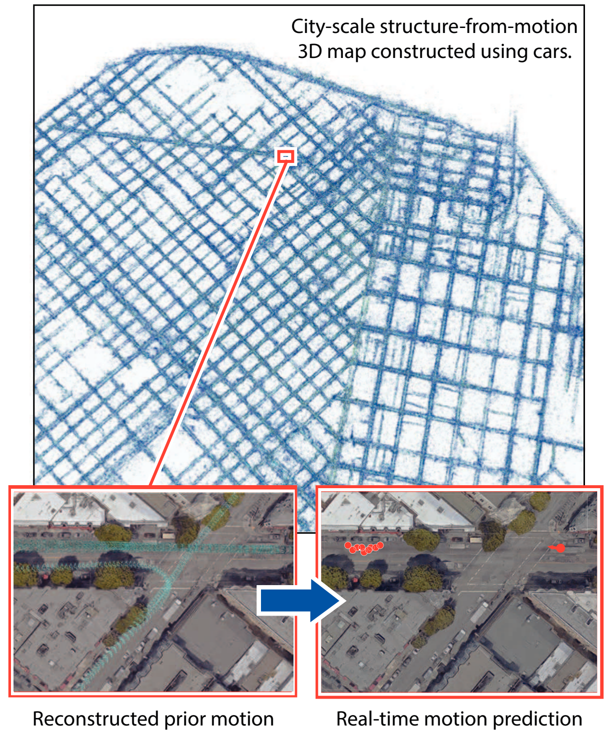


Fig. 1: Our system can accurately predict the future position of a vehicle based only on the previous motion in the area. This prior motion was crowd-sourced by a fleet of mobile-phone-equipped cars.

I. Introduction

For autonomous vehicles, critical tasks like path planning and obstacle avoidance require the ability to predict the future evolution of the environment around the robot. Complex environments such as urban city traffic present significant challenges when it comes to such planning and perception. In the near future, these methods will play a significant role in reducing road accidents.

Movement predictions in these semi-structured environments are usually based on assumed motion dynamics of the vehicles around the car, for example by using a Kalman Filter. A common disadvantage of these models is that they can generalize badly to the vast complexity of real world scenarios such as busy intersections or turns. The exhibited motion of vehicles in these situations usually cannot be predicted reliably using simple motion models like linear extrapolation, especially if the prediction horizon is larger than a few seconds. Another approach is to annotate the road infrastructure in the form of semantic map, capturing traffic rules. This has a benefit that it can extrapolate the expected motion of a car provided that it follows these rules. However, the amount of work needed to produce such reliable maps and then to keep them updated is a disadvantage.

In this work we propose an alternative approach. Inspired by the impact of large-scale data sets on computer vision [1] we utilise a large amount of crowd-sourced high-quality motion data to drive the motion prediction. We collected it by equipping a large vehicle fleet with cameras and performing structure-from-motion at city-scale to accurately reconstruct their trajectories. We use these data as samples from the underlying motion distribution in the area and show how these data can be used for future motion prediction of newly observed cars. It has the benefits of needing zero human annotation, it implicitly captures modelled and unmodelled aspects of the vehicle motion, scales to large city-scale scenarios and improves with time as the amount of data increases.

This system can be used universally as a motion-prediction step in various vehicle-tracking systems for the purpose of vehicle safety and autonomy.

We evaluate the method on two city-scale data sets specifically collected for this experiment. We demonstrate that the system can be used to drive motion prediction at

[1]Blue Vision Labs, London, UK {mssuraj,hugo,lukas,peter}@bluevisionlabs.com
[2]Georgia Institute of Technology, Atlanta, US mssuraj@gatech.edu

large scale in a variety of traffic environmental conditions. To our knowledge this is the first published approach to be constructed and evaluated at this scale. Specifically, we present:

- A simple way to create large-scale accurate priors of vehicle motion as a by-product of building a crowd-sourced city-scale 3D map of the environment.
- A method for accurately predicting a vehicle's future position using the extracted prior from the area.
- A comprehensive evaluation on two city-scale data sets in San Francisco and New York City containing millions of samples. We show that the method vastly improves the precision over a baseline method and also demonstrate continuously improving performance as the amount of prior data grows.

The rest of this paper is organized as follows. In the following section we revisit some of the key works in the area of motion prediction and on usage of large-scale data sets. In Sec. III we detail how an unstructured motion prior can be constructed and used to accurately predict a vehicle's future position. We evaluate the method in Sec. IV and conclude in Sec. V.

II. Related Work

Various methods have been proposed over years to understand and model vehicle motion dynamics, driver intent and vehicle interactions with the environment and neighboring agents. In most cases, motion prediction involves relying fully or partly on a vehicle dynamics model. [2] compare and evaluate several motion models for tracking vehicles. They conclude that constant turn-rate and acceleration model(CTRA) perform the best among constant turn-rate and velocity(CTRV), constant steering angle and velocity(CSAV), constant curvature and acceleration (CCA) and purely linear motion models such as constant velocity(CV) or constant acceleration(CA). Such models are usually combined with Kalman filtering [3] or Bayesian filtering [4] for path prediction. However, these approaches are able to perform predictions only for a very short window into the future. [5] combine a constant yaw-rate and acceleration model with a maneuver classifier to predict vehicle trajectories. But their methods are restricted to limited scenarios and constrained by the number of maneuvers.

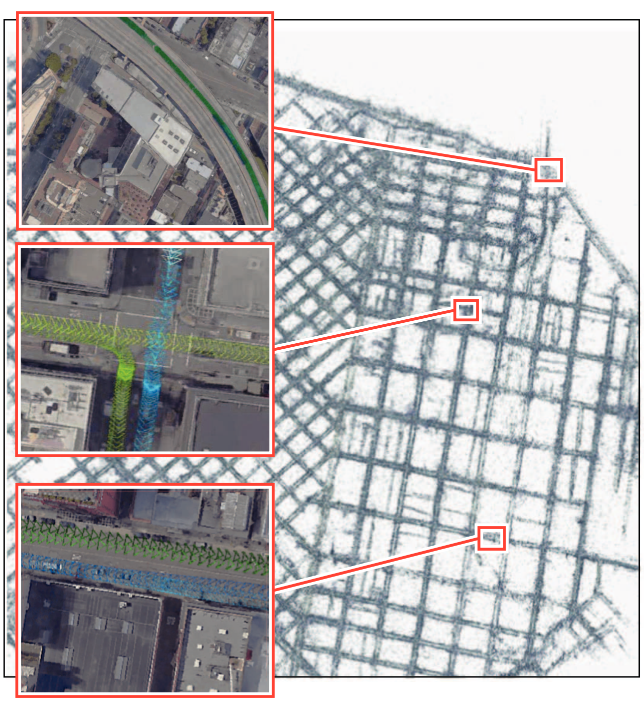


Fig. 2: Examples of the prior trajectories in San Francisco data set as generated by a randomized fleet of camera-equipped cars and reconstructed by large-scale structure-from-motion.

Recently, the focus has shifted to data-driven approaches to learn vehicle dynamics rather than explicitly crafting them. These usually employ Dynamic Bayesian networks [6], Gaussian mixture models [7] [8] [9], Hidden Markov models [8], Neural networks [10] [11] or a combination of these. These achieve better performance than pure vehicle dynamics based approaches. However, they are either trained for specific limited scenarios like highways or tend to learn a general model that do not utilize environment-specific cues such as traffic pattern in the area, changes in the environment structure, etc.

Our approach, on the other hand, utilizes location specific information for accurate predictions. Instead of learning a global model, we rely on the historical vehicle trajectories in the locality to perform on-the-fly prediction. Additionally, our system decouples the prediction system and environment knowledge thereby enabling easy update to environment priors.

We leverage the pervasive nature of cheap cameras to collect city-scale motion patterns and environment information and show that this information can be effectively used for trajectory prediction without any explicit modelling.

III. Large-Scale Motion Prior

The presented method estimates a car's future position in 3D space given its current observed position using a motion prior G in the area. Formally, given the car's observed state $s_0 = (x_0, r_0, v_0)$ consisting of position $x_0 \in R^3$, rotation $r_0 \in SO(3)$, and velocity $v_0 \in R$ in the space we aim to predict its state after $t > 0$ seconds:

$$p(s_t | s_0, G). \tag{1}$$

The motion prior we propose to use consists of a large set of individual trajectory samples that contain accurate 3D positions and rotations of vehicles driven through the city in the past:

$$G = \{G^1, G^2, ..., G^N\}, \tag{2}$$

where each trajectory $G^i = \{s^i_1, s^i_2, ..., s^i_m\}$ is a sequence of observed positions, rotations, and velocities of the car at regular time intervals $t = 1, 2, 3, ..., m$ as the car had been driven around the city.

As discussed in Section IV this method was successfully applied in a city-scale scenario where such a prior can be automatically extracted as a by-product of building a large-scale crowd-sourced 3D map of the environment. Specifically, we do not use any manual or semantic annotations of the environment or any knowledge of traffic rules. Instead, we assume each trajectory implicitly captures all relevant information in the exhibited behaviour. Some examples of these trajectories are displayed in Figure 2.

(a) $\sigma_v = 5$

(b) $\sigma_v = 25$

Fig. 3: Effect of σ_v on sampling. With increased σ_v we draw samples from a wider range of prior states with different velocities.

To predict the future position of a vehicle at a time t, we hypothesize that the car is following the same trajectory pattern as one of the cars in the past at the same location had followed. Specifically, for each prior state s_j^i we assume that the vehicle is going to follow the same motion pattern as the vehicle that generated the prior trajectory continuing from that state. Under this assumption the pose of the car in the future is going to be

$$s_t = s_{j+t}^i + \epsilon, \tag{3}$$

where s_{j+t}^i is the observed pose of the vehicle previously driven through the area after t seconds after the queried state and ϵ is a random noise modelling the fact that the trajectory can slightly differ.

The distribution of the future pose is then a weighted sum of these individual distributions

$$p(s_t|s_0, G) = \frac{1}{Z}\sum K(s_j^i, s_0)p(s_t|s_{j+t}^i, \epsilon), \tag{4}$$

where Z is a normalization factor

$$Z = \sum K(s_j^i, s_0), \tag{5}$$

and $K(s_j^i, s_0)$ measures the similarity of the prior state to the current state capturing the likelihood that it can indeed follow the exhibited prior motion pattern.

We model this likelihood as the sum of similarities of individual factors

$$K(s_j^i, s_0) = \exp\{-\frac{\|x_j^i - x_0\|^2}{\sigma_x^2} - \frac{\|r_j^i - r_0\|^2}{\sigma_r^2} - \frac{\|v_j^i - v_0\|^2}{\sigma_v^2}\} \tag{6}$$

where $\|x_j^i - x_0\|^2$ is the euclidean distance of the sample position in the 3D space, $\|r_j^i - r_0\|^2$ is the relative difference of the heading angles, and $\|v_j^i - v_0\|^2$ is the difference in the linear speed. The parameters σ_x, σ_r and σ_v model the relevance of the individual components. Note that we do not

Algorithm 1 Motion prior sampling

Input

s_0: initial state (position, rotation, velocity)
t: time horizon
G: motion prior

Output

$\hat{s}_{1:N}$: samples of the predicted future car state

```
Z ← Σ_{i,j} K(s_i^j, s_0)
for i = 1, 2, 3, ..., m, j = 1, 2, 3, ..., N do
Compute each prior state's relevance to initial state
    μ_i^j ← (1/Z) K(s_i^j, s_0)
end for
Generate N future state samples
for k = 1, 2, 3, ..., N do
Sample a prior state according to relevance
    s_j^i ← MultinomialSample(G, μ)
    Compute a future state sample
    s_t^k ← s_{j+t}^i + ε
end for
return ŝ_{1:N}
```

Fig. 4: Vehicle motion predictions at intersections. The green circle represents the query position and velocity at time t. The red icons represent the distribution of predicted samples at $t+5$. Note that the road ahead is a one-way route in opposite direction. Our prior implicitly capture this information without any manual annotation.

need to manually choose a subset of the dataset as the motion prior in a location for measuring similarity. The exponential term in Equation 6 lets us use all the trajectories in the dataset for a city as the prior G and requiring us to set only the hyper-parameters $\sigma_x, \sigma_r, \sigma_v$ for the entire city.

The probability density function $p(s_t|s_0, G)$ can be evaluated explicitly in a closed form. Moreover, a sampling procedure can be implemented efficiently by first sampling the corresponding prior state s^i_j according to relevance factor K , performing table look-up for s^i_{j+t} and adding noise. The algorithm is summarized in Algorithm 1

Examples of samples drawn from the distribution S_t are displayed in Figure 3. As shown, sampling follows the previously observed trajectories of prior motion in the area while parameters σ model the relevance of the individual components of the state. For example, a small value of σ_v results into predictions matching the current velocity while large σ_v results into predictions sampled using a wider variety of the previously observed initial velocities. This might be useful if the initial velocity is uncertain or unknown (e.g., when the car is observed for the first time).

In Section IV we evaluate the sampling accuracy on a large data set using this strategy. Despite the fact that the method is rather simple, it is very effective in practice once the prior G covers the area well.

IV. Experiments

In this section we evaluate the performance of the method described in Sec. III for predicting the future trajectories of cars in the city. In particular, we measure the likelihood of observing a car at its future position predicted by our method and also compare it to a simpler linear-motion model commonly used in many approaches. We demonstrate that after collecting a sufficient amount of data the described data-driven method can accurately predict the future position of a car while also significantly outperforming a simpler linear motion model. At the end of the section we explore the failure cases of the method and suggest possible directions for its improvement.

A. Data sets

To evaluate the method we collected a data set of 10M images captured by a fleet of 50 drivers using camera-equipped mobile phones capturing imagery data at regular 3Hz intervals in Downtown San Francisco and New York City. Examples from this data set are displayed in Figure 5. In total, this data set covers almost 1000h of driving at different time, weather, and traffic conditions experienced by the fleet during its operation. Next, we performed a large-scale structure-from-motion reconstruction to recover both the 3D map of the city and accurate 3D trajectories taken by the individual drivers. The resulting trajectories capture the motion pattern exhibited by the drivers in the fleet precisely localised in 3D space.

We use a one-off split method: when evaluating a trajectory we remove it from the data set and use the rest of data from the city as a motion prior G. We then compare the true trajectory with that predicted by the model.

To measure the effect of data set size, we uniformly sample a subset of 1M, 2M, 3M , 5M and 7M of the full 10M dataset such that the relative density of trajectories across locations remains the same. The results in Table II shows the effect of dataset size (and thus prior size) on prediction error.

B. Motion prediction accuracy

To evaluate the motion prediction accuracy we measure the probability of predicting the correct position $p(s_t|s_0, G)$ as defined in Eq. (4). We report two metrics:

1) The average negative log likelihood $-logp(s_t|s_0, G)$ for different values of t. This metric directly captures the probability of observing a car at a particular position, as defined by the predictive distribution. Intuitively, a more accurate method would have a lower negative log likelihood of the observed data than a less accurate method.
2) The fraction of the predictive distribution that lies within a radius of d meters of the correct position for various values of d. This captures the likelihood that a randomly sampled position from the predictive distribution is going to be within this limit.

In addition to absolute numbers we compare the accuracy against a simpler linear-motion model commonly used in various Kalman-filter based methods. We use a noisy linear model defined as

$$v_t \sim v_0 + \epsilon_v, \quad (7)$$

$$r_t \sim r_0 + \epsilon_r, \quad (8)$$

$$x_t \sim x_0 + r_t * v_t + \epsilon_x, \quad (9)$$

and fit the noise constants $\epsilon_v, \epsilon_r, \epsilon_x$ to minimise the negative log-likelihood of the data set.

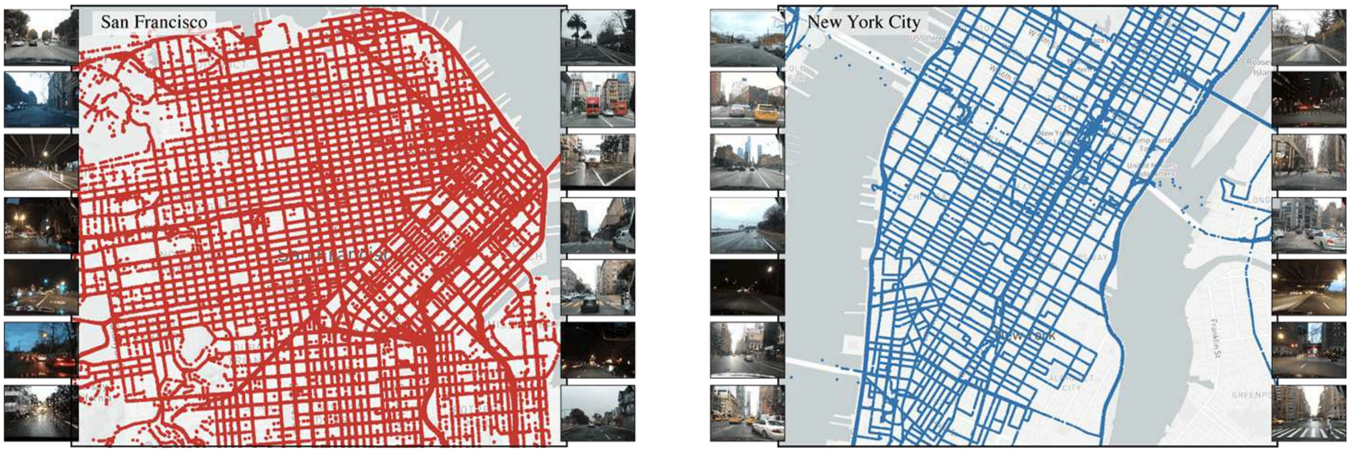


Fig. 5: Dataset used to evaluate the method. We have collected over 10M images in San Francisco and New York using dashcam-mounted mobile phones. These images were used to perform large-scale structure-from-motion to reconstruct accurate vehicle trajectories in the city over a period of several weeks.

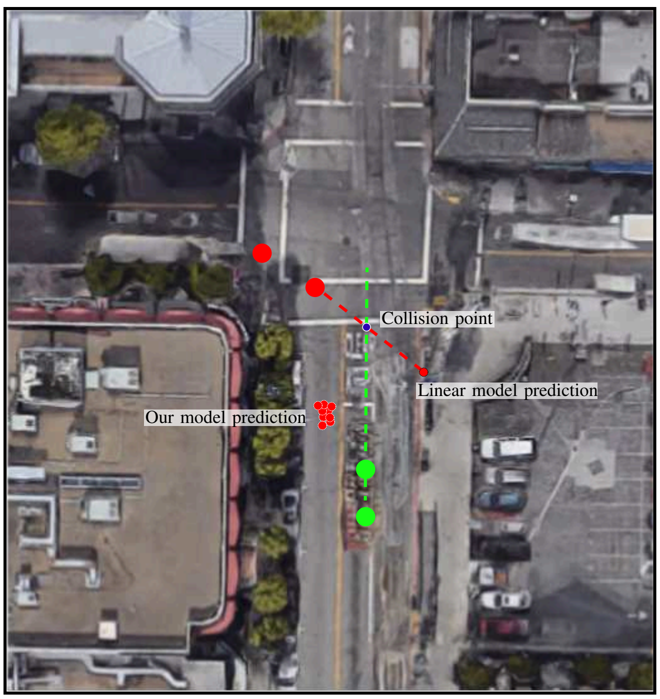


Fig. 6: A comparison to the linear model. The big red circles indicate the position of an oncoming car at t and $t+1$. Under the linear model, it is expected to collide with our vehicle. Our model corrects for this error using the motion prior.

Data	Model	1	2	3	4	5
All	Motion prior	28.74	30.56	32.80	37.55	41.48
All	Linear model	5.30	24.21	116.19	302.90	383.74
Intersection	Motion prior	27.61	32.67	33.56	40.34	44.34
Intersection	Linear model	5.23	25.12	144.90	355.35	441.22

TABLE I: Average negative log likelihood for different prediction horizons (in seconds). As the prediction horizon increases the predictive power decreases. The prior-based method, however, degrades more gracefully than the linear-motion method.

The results are summarized in Tab. I, Tab. II and Fig. 7. While the linear model prediction error grows dramatically with the look ahead time window, our model's error is much slower to grow. This makes the method suitable for predicting motion in long occlusions. A typical scenario of when using motion prior leads to better results is shown in Figure 6. While the linear model predicts a motion leading to collision the prior-based method predicts previously experienced curved motion.

Finally, Fig. 8 shows one of the failure modes of the method. It depicts a segment of the map with insufficient priors to account for all possible trajectories, thus resulting in incorrect predictions. This is a typical performance when not enough data from the area was collected and improves itself as all possible motions are exhibited over time.

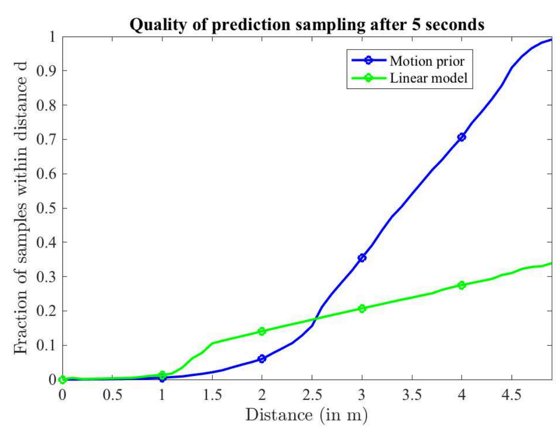
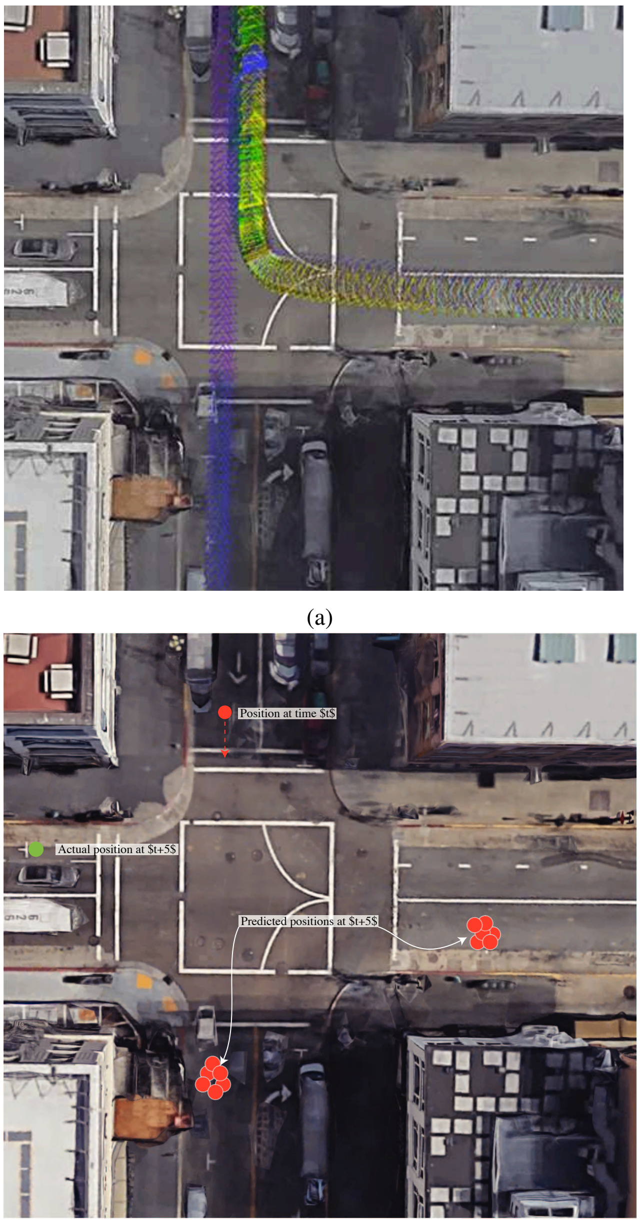


Fig. 7: The average fraction of the predictive distribution within distance dm. As the predictive distance increases, the motion-prior model quality increases rapidly, while the linear-motion model is slow to improve.

Dataset size	1M	2M	3M	5M	7M
Prediction error	67.63	58.66	52.75	45.26	42.77

TABLE II: Average negative log likelihood for 5 second prediction horizon as the size of the dataset increases.



(a)

(b)

Fig. 8: A failure mode due to lack of observed data. In (a) we show the observed prior trajectories in the map, and in (b) we show the incorrect prediction for a car to turn right due to lack of observed right turns.

V. Conclusion

We presented a non-parametric method predicting future poses of vehicles in urban environments leveraging motion data which can be collected efficiently through crowd-sourcing at city-scale. Unlike parametric based approaches the method improves over time as the amount of samples increases avoiding the need for extensive model complexity and tuning.

The method is basic step behind many situation awareness and tracking systems in self-driving vehicles. In the future we would like to explore its integration with more complex camera or lidar-based systems.

There are also many ways for its improvement. For example, to increase precision, the relevance function can be extended to include another factors beyond similarity in position, rotation and velocity to previous samples. These can be time of day to capture different traffic flows or semantic information, such as positions of other cars.

References

[1] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A Large-Scale Hierarchical Image Database," in *CVPR09*, 2009.

[2] R. Schubert, E. Richter, and G. Wanielik, "Comparison and evaluation of advanced motion models for vehicle tracking," in *Information Fusion, 2008 11th International Conference on*. IEEE, 2008, pp. 1–6.

[3] S. Ammoun and F. Nashashibi, "Real time trajectory prediction for collision risk estimation between vehicles," in *Intelligent Computer Communication and Processing, 2009. ICCP 2009. IEEE 5th International Conference on*. IEEE, 2009, pp. 417–422.

[4] N. Kaempchen, K. Weiss, M. Schaefer, and K. C. Dietmayer, "Imm object tracking for high dynamic driving maneuvers," in *Intelligent Vehicles Symposium, 2004 IEEE*. IEEE, 2004, pp. 825–830.

[5] A. Houenou, P. Bonnifait, V. Cherfaoui, and W. Yao, "Vehicle trajectory prediction based on motion model and maneuver recognition," in *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on*. IEEE, 2013, pp. 4363–4369.

[6] T. Gindele, S. Brechtel, and R. Dillmann, "Learning driver behavior models from traffic observations for decision making and planning," *IEEE Intelligent Transportation Systems Magazine*, vol. 7, no. 1, pp. 69–79, 2015.

[7] J. Wiest, M. Höffken, U. Kreßel, and K. Dietmayer, "Probabilistic trajectory prediction with gaussian mixture models," in *Intelligent Vehicles Symposium (IV), 2012 IEEE*. IEEE, 2012, pp. 141–146.

[8] S. Sivaraman, B. Morris, and M. Trivedi, "Learning multi-lane trajectories using vehicle-based vision," in *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*. IEEE, 2011, pp. 2070–2076.

[9] N. Deo, A. Rangesh, and M. M. Trivedi, "How would surround vehicles move? A Unified Framework for Maneuver Classification and Motion Prediction," *ArXiv e-prints*, Jan. 2018.

[10] B. Kim, C. M. Kang, S. H. Lee, H. Chae, J. Kim, C. C. Chung, and J. W. Choi, "Probabilistic vehicle trajectory prediction over occupancy grid map via recurrent neural network," *arXiv preprint arXiv:1704.07049*, 2017.

[11] A. Khosroshahi, E. Ohn-Bar, and M. M. Trivedi, "Surround vehicles trajectory analysis with recurrent neural networks," in *Intelligent Transportation Systems (ITSC), 2016 IEEE 19th International Conference on*. IEEE, 2016, pp. 2267–2272.