Planning Under Uncertainty with Sensor Occlusions in Urban Driving Scenarios

Ross Alexander
Department of Aeronautics and Astronautics
Stanford University
Stanford, CA 94305
rbalexan@stanford.edu

***Abstract*—Safe and efficient autonomous driving in urban scenarios requires a policy that is responsive to uncertainty in the environment and sensors. Partial observability of the environment due to sensor occlusions has led to conservative manually-designed collision avoidance policies that are inefficient. Since collision avoidance is a sequential decision-making problem under uncertainty, it can be posed as a partially observable Markov decision process (POMDP), which can be solved to generate an optimal policy. We leverage the POMDP formulation to generate approximately optimal collision avoidance policies and evaluate the safety and efficiency of the resulting driving behaviors. We find that the policies outperform a random policy and are generally highly-performant and robust to uncertainty in the environment and sensors.**

I. Introduction

Autonomous driving in urban scenarios requires navigating safely and efficiently in light of significant uncertainty presented by occlusions. In dense urban settings, buildings, signs, cars, and other physical obstacles frequently occlude other road users from the field of view of the sensors. Planning in these occluded scenarios requires judicious safety thresholds in order to ensure collision-free driving. In this paper, we focus on developing safe and efficient policies for collision avoidance while passing a crosswalk with pedestrians emerging from an occluded region.

We consider the occluded crosswalk scenario depicted in Figure 1. The scenario consists of an autonomous vehicle driving along a roadway and a pedestrian crossing the roadway at a crosswalk. The ego vehicle's sensors are occluded by a static obstacle that prevents the pedestrian from being observed.

Urban driving scenarios are often heavily occluded, which presents significant challenges in achieving fully autonomous driving. Occlusions in urban scenarios can arise as a result of *static obstacles*, such as buildings, signs, and parked cars, but can also arise as a result of *dynamic obstacles*, such as moving pedestrians and moving cars. Above all, the autonomous vehicle must navigate safely and avoid collision with pedestrians. However, the autonomous vehicle must also navigate efficiently and cannot be paralyzed by the significant uncertainty in the environment and sensors. A pedestrian's pose and velocity may be completely unobservable due to a sensor occlusion, or at best, partially observable due to sensor uncertainty. As a result, the occluded crosswalk scenario strongly captures the trade-off between safety and efficiency in autonomous driving. Optimizing this trade-off and generating a driving policy that is safe and robust to uncertainty in the environment and sensors is a critical step in urban autonomous driving.

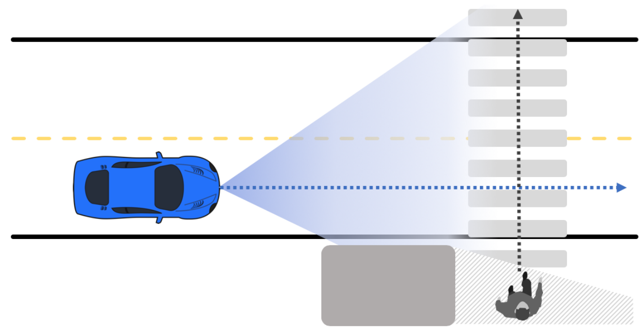

Fig. 1. The occluded crosswalk scenario. The ego vehicle (in blue) is driving along the roadway and the pedestrian (in gray) is crossing the crosswalk. The static obstacle occludes the ego vehicle's sensors and prevents observation of the pedestrian.

Planning in occluded scenarios faces two main challenges. First, the planning algorithm must be able to handle partial observability. Planning requires not only reasoning about an unoccluded pedestrian's pose and velocity, which can only be observed with finite accuracy and precision, but also reasoning about a potentially occluded pedestrian's pose and velocity, which cannot be observed. And second, real urban driving scenarios often contain multiple road users. While solving the occluded crosswalk problem with a single road user can be useful in theory, it is hardly useful in practice. Unfortunately, solving the occluded crosswalk problem with multiple road users can be intractable for many planning algorithms. Thus, a solution for planning in occluded scenarios must be capable of simultaneously handling partial observability of multiple road users in a way that maintains safety and efficiency, and finally, in a way that respects real-time decision-making constraints.

In this paper, we consider the problem of avoiding collisions with pedestrians crossing behind an occluded region on the side of the road and attempt to recreate and extend the results in [1]. The approach formulates the multiple road user collision avoidance problem as a partially observable Markov decision process (POMDP) with discrete state, action, and observation spaces. The POMDP is solved using offline techniques to obtain driving policies, which are evaluated in simulation and compared against a random policy on safety and efficiency metrics.

We discuss related work in Section II and describe the partially-observable Markov decision process (POMDP) and solution techniques for generating an approximately optimal policy in Section III. Section IV discusses how the collision avoidance problem can be posed as a POMDP and the solution approaches we take in obtaining an approximately optimal policy. Section V covers the experimental setup, analysis of the safety and efficiency criteria, and discussion of the resulting driving policies. Finally, we present our conclusions in Section VI and provide some areas for further research in Section VII.

II. Related Work

Several planning algorithms ensure collision-free paths when passing occluded regions. If objects emerging from occluded regions have a specified maximum velocity, the autonomous vehicle can follow a maximal velocity profile along its original planned path that ensures a collision-free trajectory [2]. Some approaches rely on the time-to-collision (TTC) metric and extensions of the TTC metric, which have shown to be reliable safety measures not only for vehicle collision avoidance, but also for pedestrian collision avoidance [3], [4]. Many of these planning frameworks are designed solely around the use of *autonomous emergency braking* (AEB) systems to perform drastic collision avoidance maneuvers. While AEB system planning frameworks are effective at ensuring essentially collision-free driving, they suffer from uncertainty in the environment and sensors that can create inefficient control policies likely to initiate unnecessary severe rapid braking.

In order to develop driving policies robust to uncertainty in the environment and sensors, others have chosen to model the collision avoidance problem as a partially observable Markov decision process (POMDP). POMDPs have been used to successfully generate autonomous driving policies in occluded scenarios. In particular, several approaches have leveraged POMDPs to develop longitudinal-control policies for unsignalized intersections [5], [6] and for unsignalized crosswalks [7]–[9]. Most of these approaches considered only one road user due to the increased algorithmic and computational complexity of considering multiple road users. While these approaches generated progress, their applicability is still limited since most urban scenarios contain multiple road users which limits their practical use.

A scalable approach to avoiding multiple road users in both the unsignalized intersection and unsignalized crosswalk was achieved through utility fusion [1]. The multiple road user collision avoidance problem was decomposed into single road user collision avoidance sub-problems enabling solution of the problem that scales linearly with the number of road users considered. The set of single road user optimal belief action utilities were synthesized to generate an approximation to the optimal belief action utility for all road users using a fusion function. The authors consider both sum and minimum fusion functions and find the minimum fusion function generates a more conservative policy. Others have expanded on the success of scalable longitudinal-control policies and developed scalable in-lane coupled longitudinal- and lateral-control policies augmented with AEB systems that maintain safety and improve scenario crossing speeds [10].

III. Background

A principled and general framework for planning under uncertainty is the partially-observable Markov decision process (POMDP). A POMDP is defined by the tuple $\langle \mathcal{S}, \mathcal{A}, \mathcal{O}, T, R, O, \gamma \rangle$, where $\mathcal{S}$ is the state space, $\mathcal{A}$ is the action space, $\mathcal{O}$ is the observation space, T is the transition model, R is the reward model, O is the observation model, and γ is the discount factor. A POMDP models the process where an agent in state $s \in \mathcal{S}$ takes action $a \in \mathcal{A}$ and transitions to a state $s' \in \mathcal{S}$ with probability $T(s, a, s') = \Pr(s' \mid s, a)$. Then, the agent observes $o \in \mathcal{O}$ with probability $O(o, s', a) = \Pr(o \mid s', a)$ and receives a real-valued reward $r = R(s, a)$. Rewards in the future are discounted by a factor $\gamma \in (0, 1)$, so that immediate rewards are valued more than future rewards.

In a POMDP, the state of the environment may not be fully observable, so the agent maintains a belief b about the underlying state, which is assumed to be a sufficient statistic for the agent's history of actions and observations. A belief is a categorical probability distribution over the state space that represents the probability of each state being the true underlying state. The belief b is updated after taking action a and observing o using the Bayesian belief update:

$$b'(s') \propto O(o \mid s', a) \sum_{s \in S} T(s' \mid s, a) b(s) \tag{1}$$

The solution to a POMDP is an optimal policy π^* that maps beliefs to actions. Following the optimal policy from any initial state maximizes the expected discounted sum of future rewards. Computing the exactly optimal policy is generally intractable in non-trivial POMDPs, so we must use approximation techniques to compute an approximately optimal policy [11]. There are a variety of approaches for generating approximately optimal policies that can be broadly split into two classes: offline methods and online methods [12]. Offline approaches involve computing the entire policy prior to execution, which is typically computationally expensive. Online approaches involve computing a policy from the current belief state over a specified planning horizon. Online approaches benefit from the ability to plan over the reachable belief space, which is typically a small subspace of the entire belief space, but suffer from reduced computational budget due to real-time, finite-memory computational constraints.

IV. Proposed Approach

A. POMDP Model

We describe in the following subsections how the single user occluded crosswalk problem can be formulated as a POMDP.

1) Environment and initial conditions: The roadway is $32\ \mathrm{m}$ long and crosswalk is located $20\ \mathrm{m}$ from the start of the roadway and is $10\ \mathrm{m}$ long, similar to the scenario depicted in Figure 1. In all scenarios, the ego vehicle starts at the beginning of the roadway with a velocity of $5\ \mathrm{m/s}$, and when a pedestrian is generated, the pedestrian starts at the beginning of the crosswalk with a velocity of $1\ \mathrm{m/s}$.

2) State space: For the occluded crosswalk scenario, we consider the positions and velocities of the ego vehicle and the pedestrian, which are expressed in the Frenet frame for generalization to arbitrary roadway configurations. The ego vehicle state tuple consists of the ego's position along the lane and velocity in the lane (s_{ego}, v_{ego}). Similarly for the pedestrian, the state tuple consists of the pedestrian's position along the crosswalk and velocity in the crosswalk (s_{ped}, v_{ped}). If there is no pedestrian, the pedestrian's state tuple is marked as absent (s_{absent}, v_{absent}).

3) Action space: The ego vehicle's actions consist of the set of accelerations $\mathcal{A} = \{-2\ \mathrm{m/s^2}, -1\ \mathrm{m/s^2}, 0\ \mathrm{m/s^2}, 1\ \mathrm{m/s^2}\}$. The accelerations are designed to map to standard strategic maneuvers – rapid deceleration, deceleration, continuation, and acceleration, respectively.

4) Observation space: The observation space is similar to the state space. The ego vehicle's state is assumed to be perfectly observable. If the pedestrian is occluded, the pedestrian's state tuple is unobserved $(o_{s_{absent}}, o_{v_{absent}})$. If the pedestrian is not occluded, the pedestrian's state is partially observable with the position and velocity following Gaussian distributions centered on the true value.

$$o_{s_{ped}} \sim \mathcal{N}(o_{s_{ped}} \mid \mu_s, \sigma_s) \tag{2}$$

$$o_{v_{ped}} \sim \mathcal{N}(o_{v_{ped}} \mid \mu_v, \sigma_v) \tag{3}$$

5) Transition model: The evolution of the ego vehicle from time step $t-1$ to t is given by a simple constant acceleration model shown in Equation (4). The ego vehicle's velocity is clipped below $0\ \mathrm{m/s}$ and above $8\ \mathrm{m/s}$ to model the inability to reverse on a roadway and to model a speed limit.

$$\begin{aligned} a_t &= a_{ego} \\ v_t &= \min(\max(v_{t-1} + a_t \Delta t, 0), 8) \\ s_t &= s_{t-1} + v_t \Delta t + \tfrac{1}{2} a_t \Delta t^2 \end{aligned} \tag{4}$$

The pedestrian evolves from time step $t-1$ to t according to a simple constant velocity model given in Equation (5). The velocity change at the following time step is sampled according to a uniform distribution over a set of velocities $\Delta V = \{-1\ \mathrm{m/s}, 0\ \mathrm{m/s}, 1\ \mathrm{m/s}\}$, which captures the possibility of the pedestrian changing speed during crossing. The pedestrian's velocity is also clipped below $0\ \mathrm{m/s}$ and above $2\ \mathrm{m/s}$ to model the intention of the pedestrian to eventually cross the road.

$$\begin{aligned} &a_t - \text{unmodeled} \\ &v_t = v_{t-1} + \min(\max(\Delta v \sim \mathcal{U}(\Delta v \mid \Delta v \in \Delta V), 0), 2) \\ &s_t = s_{t-1} + v_t \Delta t \end{aligned} \tag{5}$$

The state space is discretized into $1\ \mathrm{m}$ position increments and $1\ \mathrm{m/s}$ velocity increments to facilitate offline computation. The discretization of the state space gives 33 ego positions, 9 ego velocities, 11 pedestrian positions, and 3 pedestrian velocities, along with the absent pedestrian state tuple gives a state space cardinality of $|\mathcal{S}| = 10{,}098$.

6) Reward model: There is a reward of 1 for reaching the goal state and a collision penalty of -10. Rewards are discounted using a discount factor $\gamma = 0.9$, which encourages the agent to seek immediate rewards over future rewards.

B. POMDP Solvers

We use two offline methods, namely, QMDP [13] and SARSOP [14], to generate approximately optimal policies for the occluded crosswalk problem.

QMDP solves the POMDP by assuming full observability at the following time step. This enables the use of an approximate value iteration algorithm from which the optimal policy can be computed. However, since QMDP assumes full observability at the next time step, it suffers in scenarios that require information-gathering actions.

SARSOP, which stands for Successive Approximation of the Reachable Space under Optimal Policies, uses a set of belief points in the belief space to approximate the optimal policy. The SARSOP algorithm efficiently expands the set of belief points using reachability constraints, which leads to an effective policy representation with fewer belief points than standard point-based value iteration algorithms. SARSOP is more computationally complex than QMDP, but SARSOP incorporates the inherent uncertainty in the POMDP while QMDP eliminates it at the second step. Thus, we expect approximately optimal policies generated using SARSOP to be more robust to state uncertainty than those generate with QMDP.

V. Experiments & Results

We formulate the occluded crosswalk POMDP using a Julia package called `AutomotivePOMDPs.jl`[1], which is built on `POMDPs.jl` [15]. In each of the following experiments, the time step was 0.5 s and the probability of generating a pedestrian at each time step was 0.3.

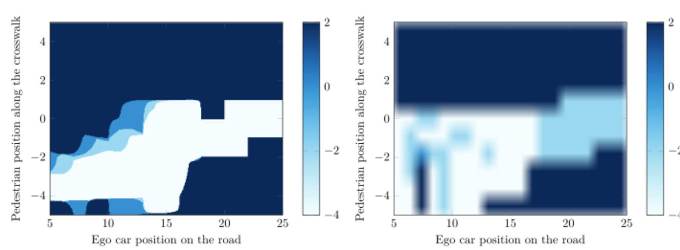

Fig. 2. Policy plots for the occluded crosswalk problem with the gradations representing accelerations in $\mathrm{m/s^2}$. There is no error in the sensing of the pedestrian's position and velocity.

A brief comparison of QMDP and SARSOP policies was conducted to demonstrate their differences. We analyzed the case of no sensor noise and examined the resulting policies,

[1] `https://github.com/sisl/AutomotivePOMDPs.jl`

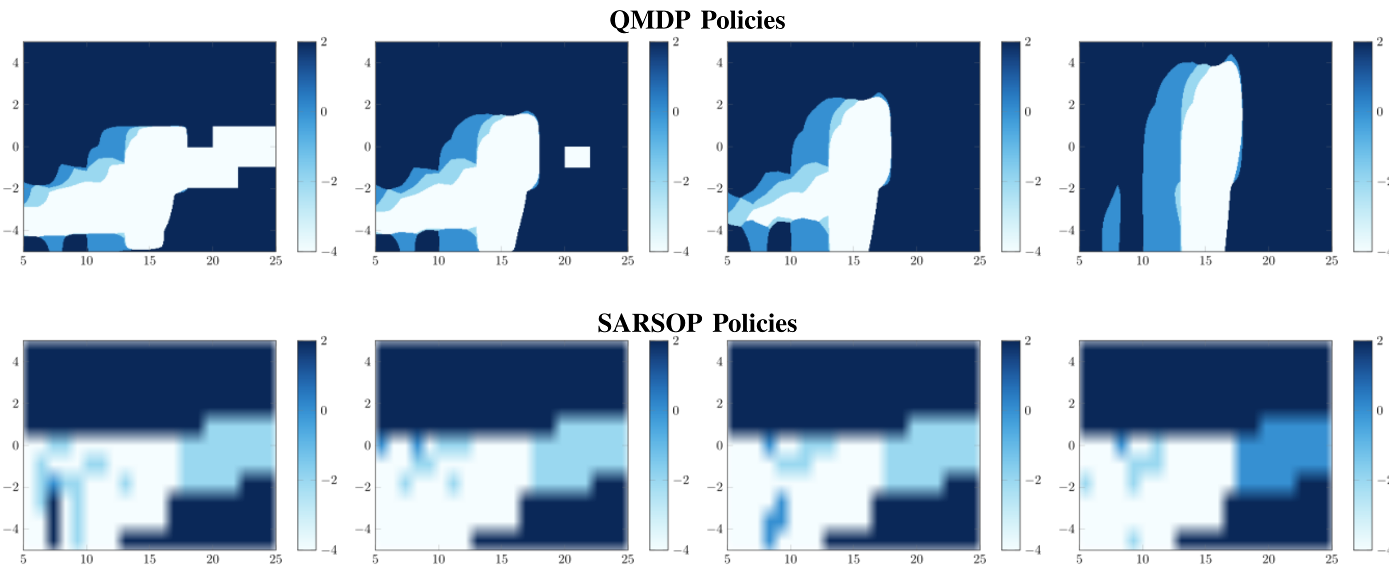

Fig. 3. Policy plots for the occluded crosswalk problem with the gradations representing accelerations in $\mathrm{m/s^2}$. The horizontal axis represents the position of the ego vehicle in the roadway in m, with the right edge of the plot representing the location of the crosswalk; the vertical axis represents the position of the pedestrian in the crosswalk in m. The ego vehicle is initialized at $5\ \mathrm{m/s}$ and the pedestrian is initialized at $1\ \mathrm{m/s}$. From left to right, the standard deviation of the sensor's position and velocity estimates increases ($\sigma = 0\ \mathrm{m}$, $\sigma = 0.5\ \mathrm{m}$, $\sigma = 1\ \mathrm{m}$, $\sigma = 2\ \mathrm{m}$), demonstrating the POMDP's robustness to uncertainty.

which are represented in Figure 2[2]. The ego vehicle is initialized at $5\ \mathrm{m/s}$ and the pedestrian is initialized at $1\ \mathrm{m/s}$.

The policies exhibit similar behavior, though the QMDP policy is more aggressive throughout, with aggressive braking for most of the ego vehicle's positions. The SARSOP policy appears similar, though the policy oddly switches to less aggressive braking as the ego approaches the pedestrian. This may be due to the $1000\ \mathrm{s}$ time-out termination condition we set for SARSOP. The policies are topologically similar to those presented in [1].

We also analyzed the effect of sensor uncertainty in each POMDP solver to examine the robustness to uncertainty in the environment and the sensors, which is shown in Figure 3.

As the sensor uncertainty grows, the QMDP policy evolves to maintaining acceleration and eventually braking as it approaches the crosswalk, regardless of the position of the pedestrian. This is a poor policy since the ego vehicle essentially comes to a complete stop, whether or not the vehicle and pedestrian would actually collide. The SARSOP policy however, adapts to the uncertainty and slows down early if a pedestrian is in the crosswalk. The SARSOP policy is also able to maintain most of the policy as acceleration when a collision with a pedestrian is not imminent. These results show similar behavior to those presented in [1].

The final experiments examine a real-world case. Following the choices of the authors in [1], in order to balance safety and efficiency, we set the collision penalty for QMDP to -1.6 and the collision penalty for SARSOP to -30. The position sensor uncertainty was set to $0.5\ \mathrm{m}$ and the velocity sensor uncertainty was set to $0.5\ \mathrm{m/s}$.

After obtaining the approximately optimal QMDP and SARSOP policies shown in Figure 4[3], we evaluated the policies in a higher-fidelity simulator with the time step decreased to $0.1\ \mathrm{s}$ and the probability of generating a pedestrian at each time step of 0.01. A random policy was evaluated alongside the QMDP and SARSOP policies. All three policies were compared on a metric of safety – the collision rate – and a metric of efficiency – the time to cross. The random policy results were computed over 10,000 simulations, while the QMDP policy and SARSOP policy results we computed on 100 simulations, due to the increased computational cost of performing the belief update. The SARSOP scenario was not finished since the belief update for SARSOP takes much longer than for QMDP. The results reported in [1] are reproduced alongside our results in Table I.

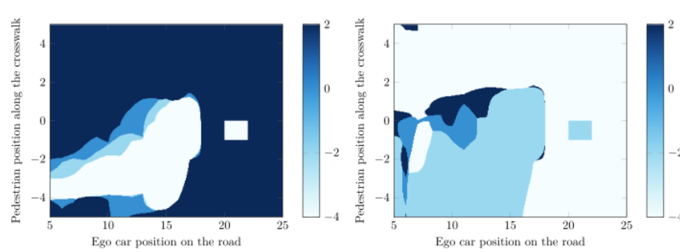

Fig. 4. Policy plots for the real-world occluded crosswalk problem with the gradations representing accelerations in $\mathrm{m/s^2}$.

Despite the incomplete results, we can still see the improvement QMDP makes over a random policy, greatly reducing the

[2]The SARSOP policy plots admittedly look different than the QMDP policy plots – the plotting routines for these were different when I generated them and I did not have time to re-run the script in order to make the formatting identical. The underlying notional policy representation, however, should be correct.

[3]I tried to fix the SARSOP plotting to match the QMDP plotting, but it isn't quite fixed.

TABLE I
RESULTS OF OCCLUDED CROSSWALK POLICIES

	Collision rate	Time to cross
Reported		
Random	$54.45 \pm 2.56\%$	12.03 ± 10.66 s
QMDP	$0.00 \pm 0.00\%$	10.61 ± 3.76 s
SARSOP	$0.00 \pm 0.00\%$	10.51 ± 4.44 s
Our Results		
Random	$18.65 \pm 38.95\%$	16.70 ± 7.72 s
QMDP	$4.00 \pm 19.69\%$	5.34 ± 0.98 s
SARSOP[4]	–	–

collision rate and the time to cross. The QMDP policy is much more reliable than the random policy. We would expect that the SARSOP policy would outperform both the random and QMDP policies, though not by that much for the QMDP policy. The results reported in [1] are close to our results, but not identical. We do not achieve the zero-collision-rate behavior of QMDP and SARSOP, likely due to differences in the many parameters governing the POMDP, but the trends show the expected behavior.

VI. CONCLUSION

In this paper, we presented the problems facing autonomous driving in occluded urban scenarios. We reviewed literature over solving the single road user occluded crosswalk and occluded intersection problems and discussed some recent work on solving multiple road user occluded problems through utility fusion. We framed for the single user occluded crosswalk problem as a partially observable Markov decision process (POMDP) and presented QMDP and SARSOP as our two approaches for solving the POMDP. We were generally able to recreated the results in [1], which highlight the strengths of the POMDP representation and power of the SARSOP algorithm in maintaining performance in the face of significant uncertainty in the environment and sensors by efficiently planning in the reachable belief space. There is still work to be done in full recreating the results of [1], which is addressed in the following section.

In the process of recreating these results, I learned a lot about `POMDPs.jl` and how to model autonomous driving as a POMDP, which will help me immensely in my future research.

VII. FUTURE WORK

We will continue to investigate some of the topics presented in this paper. It would be good to get a complete comparison of the random, QMDP, and SARSOP policies' collision rates and times-to-cross, possibly with more simulations to provide tighter error bounds. Additionally, a baseline manually-designed policy based on a time-to-collision (TTC) metric would also provide meaningful results for comparison.

The policies discussed in this paper were all generated using offline methods. A future research direction is to explore online methods while avoiding discretization of state, action, and observation spaces. A first step would be to use the partially observable Monte Carlo planning (POMCP) algorithm [16] or the determinized sparse partially observable trees (DESPOT) algorithm augmented with importance sampling to include rare, but critical events [17], [18].

Beyond this the POMDP could be considered in its natural continuous form and solved with a variety of algorithms based on progressive widening (PW), namely, the partially observable Monte Carlo planning with observation widening (POMCPOW) algorithm [19]. We expect these methods to be much more difficult to compute but possibly more efficient given the continuous nature of the state, action, and observation spaces.

It would also be interesting to see the performance of many of these policies in the presence of cyclists, who are important, but frequently neglected road users. Since cyclists typically have much higher velocities, we expect the policies learned from pedestrian behavior to fail, but future research could be done to develop hybrid policies that at least avoid a subset of cyclist collisions while maintaining or improving pedestrian safety.

REFERENCES

[1] M. Bouton, A. Nakhaei, K. Fujimura, and M. J. Kochenderfer, "Scalable Decision Making with Sensor Occlusions for Autonomous Driving," *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 2076–2081, 2018.

[2] R. Alami, T. Siméon, and K. Madhava Krishna, "On the influence of sensor capacities and environment dynamics onto collision-free motion plans," *IEEE International Conference on Intelligent Robots and Systems*, 2002.

[3] M. M. Minderhoud and P. H. Bovy, "Extended time-to-collision measures for road traffic safety assessment," *Accident Analysis and Prevention*, 2001.

[4] B. Volz, H. Mielenz, I. Gilitschenski, R. Siegwart, and J. Nieto, "Inferring Pedestrian Motions at Urban Crosswalks," *IEEE Transactions on Intelligent Transportation Systems*, 2019.

[5] S. Brechtel, T. Gindele, and R. Dillmann, "Probabilistic decision-making under uncertainty for autonomous driving using continuous POMDPs," in *2014 17th IEEE International Conference on Intelligent Transportation Systems, ITSC 2014*, 2014.

[6] C. Hubmann, N. Quetschlich, J. Schulz, J. Bernhard, D. Althoff, and C. Stiller, "A POMDP maneuver planner for occlusions in urban scenarios," in *IEEE Intelligent Vehicles Symposium, Proceedings*, 2019.

[7] S. M. Thornton, F. E. Lewis, V. Zhang, M. J. Kochenderfer, and J. Christian Gerdes, "Value Sensitive Design for Autonomous Vehicle Motion Planning," in *IEEE Intelligent Vehicles Symposium, Proceedings*, 2018.

[8] S. M. Thornton, B. Limonchik, F. E. Lewis, M. J. Kochenderfer, and J. C. Gerdes, "Toward Closing the Loop on Human Values," *IEEE Transactions on Intelligent Vehicles*, 2019.

[9] N. R. Kapania, V. Govindarajan, F. Borrelli, and J. Christian Gerdes, "A hybrid control design for autonomous vehicles at uncontrolled crosswalks," in *IEEE Intelligent Vehicles Symposium, Proceedings*, 2019.

[10] M. Schratter, M. Bouton, M. J. Kochenderfer, and D. Watzenig, "Pedestrian collision avoidance system for scenarios with occlusions," *IEEE Intelligent Vehicles Symposium, Proceedings*, vol. 2019-June, no. Iv, pp. 1054–1060, 2019.

[11] O. Madani, S. Hanks, and A. Condon, "On the undecidability of probabilistic planning and related stochastic optimization problems," *Artificial Intelligence*, 2003.

[12] M. J. Kochenderfer, *Decision Making Under Uncertainty: Theory and Application*. MIT Press, 2015.

[13] M. L. Littman, A. R. Cassandra, and L. P. Kaelbling, "Learning policies for partially observable environments: Scaling up," in *Machine Learning Proceedings 1995*, 1995.

[14] H. Kurniawati, D. Hsu, and W. S. Lee, "SARSOP: Efficient point-based POMDP planning by approximating optimally reachable belief spaces," in *Robotics: Science and Systems*, 2009.

[15] M. Egorov, Z. N. Sunberg, E. Balaban, T. A. Wheeler, J. K. Gupta, and M. J. Kochenderfer, "POMDPs.jl: A Framework for Sequential Decision Making under Uncertainty," *Journal of Machine Learning Research*, 2017.
[16] D. Silver and J. Veness, "Monte-Carlo planning in large POMDPs," in *Advances in Neural Information Processing Systems 23: 24th Annual Conference on Neural Information Processing Systems 2010, NIPS 2010*, 2010.
[17] N. Ye, A. Somani, D. Hsu, and W. S. Lee, "DESPOT: Online POMDP planning with regularization," *Journal of Artificial Intelligence Research*, 2017.
[18] Y. Luo, H. Bai, D. Hsu, and W. S. Lee, "Importance sampling for online planning under uncertainty," *International Journal of Robotics Research*, 2019.
[19] Z. N. Sunberg and M. J. Kochenderfer, "Online algorithms for POMDPs with continuous state, action, and observation spaces," in *Proceedings International Conference on Automated Planning and Scheduling, ICAPS*, 2018.