

Agent-Aware State Estimation: Effective Traffic Light Classification for Autonomous Vehicles

Shane Parr^{1*}, Ishan Khatri^{1*}, Justin Svegliato¹, and Shlomo Zilberstein¹

Abstract—Autonomous systems often operate in environments where the behavior of all agents is mostly governed by the perception of a specific feature of the environment. When an autonomous system cannot recover this feature, there can be disastrous consequences. We introduce a novel framework for *agent-aware state estimation* that exploits the dependency of all agents’ behavior on a feature to better indirectly observe the feature. To allow for fast and accurate inference, we provide a mapping of our framework to a dynamic Bayesian network and show that speed of inference scales favorably with the number of agents in the environment. We then apply our approach to traffic light classification, focusing on instances where direct vision of the light may be obstructed by glare, heavy rain, vehicles, or other environmental factors. Finally, we show that agent-aware state estimation outperforms prevailing methods that only use direct image data of the traffic light on a real-world autonomous vehicle data set of challenging scenarios.

I. INTRODUCTION

Autonomous systems often operate in high-stakes environments where the behavior of every agent is overwhelmingly governed by the perception of a specific feature. This feature can refer to any aspect of the environment that causes a specific behavior as an immediate reaction, such as an obstacle or a sign. We refer to this feature as a *global signal*: *global* in that it is shared among all agents currently being reasoned about and *signal* in that it strongly informs the decision making of each agent. This kind of signal is often seen in a wide range of practical applications, including autonomous driving [1], [2], [3], space exploration [4], [5], and search and rescue [6], [7], [8].

In scenarios where a system cannot always perceive the global signal, the consequences can be disastrous. There are several methods that can be used to mitigate such catastrophic scenarios. If human assistance is available, the agent may perform metareasoning and choose to transfer control to a human [9]. When fully autonomous, information gathering actions can be taken based on a belief that accounts for all current uncertainty to avoid myopic, over-confident behavior [10]. In settings with multiple collaborative agents, cost-effective communication is desirable to ensure that if one agent observes the global signal, it can be transmitted

This work was supported in part by an NSF Graduate Research Fellowship DGE-1451512 and the NSF grants IIS-1724101 and IIS-1813490.

*Both authors contributed equally.

¹College of Information and Computer Sciences, University of Massachusetts Amherst, MA, USA. Emails: sparr@umass.edu, {ikhatri, jsvegliato, shlomo}@cs.umass.edu

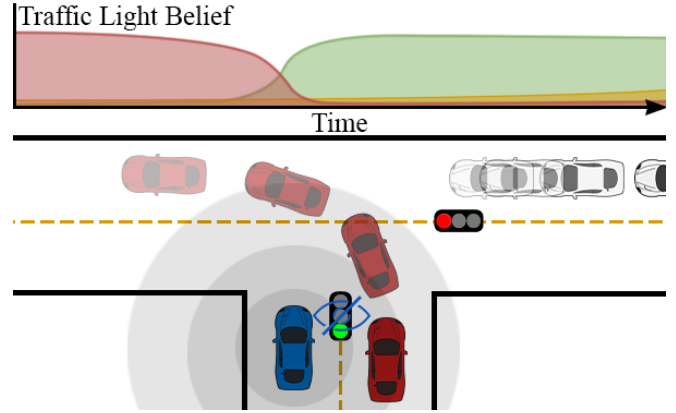


Fig. 1: By observing the behavior of intelligent agents, it is possible to estimate global signals. Here, by watching other vehicles in the environment, the blue autonomous vehicle can form a coherent belief of the state of a traffic light without observing it directly.

so other agents can still act favorably [11]. However, in high-stakes, partially observable multi-agent environments where human intervention, information gathering, and inter-agent communication is limited, state estimation *must* be robust enough to recover global signals when perception fails.

Figure 1 depicts an example of the problem at hand. A *blue* autonomous vehicle, the neutral observer, is approaching a three-way intersection with two other agents: the *red* vehicle and the *white* vehicle that are each operated by a human. Given the limitations of its sensors, the blue autonomous vehicle cannot observe the state of its traffic light, the global signal, due to an obstruction like glare. However, by observing the red vehicle turning and accelerating and the white vehicle decelerating, the blue autonomous vehicle can properly recover the state of its traffic light.

A simple approach to this form of state estimation is to use hidden Markov models (HMM) or other graphical models to smooth missing or erroneous observations [12]. Although HMMs can smooth a signal in the presence of noise, total obstruction of the signal for long periods of time is a considerable problem [13]. It is also difficult to incorporate complex information into HMMs because the number of states is in general exponential in the number of variables [14]. Thus, HMM approaches typically avoid using additional information representing the behavior of other agents. However, because HMMs work well for temporal smoothing, there have been many applications [15], including in multi-agent scenarios like traffic light classification [16].

Alternatively, state estimation and decision making can be tackled concurrently by using a partially observable Markov decision process (POMDP) [10]. The agent maintains a distribution over possible states as a belief and updates it with exact or approximate belief updates based on the current observation and action. This can be applied to a decentralized collaborative setting with multiple agents by using a Decentralized POMDP (Dec-POMDP) [17]. Although the belief can be updated in a provably correct way, there are practical considerations that prevent POMDPs and Dec-POMDPs from gaining widespread use. Principally, they are challenging to solve exactly or even approximately in some cases. Dec-POMDPs, without strict simplifying assumptions, are particularly hard to solve—they are NEXP-complete in the worst case with finite horizons [18]. Additionally, a Dec-POMDP model requires an accurate joint model of all plausible agent interactions that may not be available.

We propose a novel framework for state estimation of a global signal in a multi-agent system that merges the practicality of an HMM approach with the theoretical advantages of a Dec-POMDP approach. Using this framework and approximate single-agent behavior models, we can leverage observations of other agents as they perform decentralized planning and reconstruct a desired global signal. To do this, we use an approximate single-agent behavior model for each agent that models the dependence of their behavior on the global signal. Using a dynamic Bayesian network [19] where each node and edge corresponds to a component of a transition-independent Dec-MDP, we can improve a direct estimate the signal in a robust, decision theoretic manner. In short, by providing a tractable graphical model representation for inference, we are competitive with the speed and reliability of prevailing methods, while drastically expanding the scope and effectiveness of estimation.

Our main contributions are formalizing a framework for agent-aware state estimation problems by framing the scenario as a Dec-MDP and then demonstrating how to solve for the state given a policy. Additionally, we provide an exact Dynamic Bayesian Network (DBN) representation that facilitates fast and accurate inference (Figure 2). We apply this approach to the problem of traffic light classification by an autonomous vehicle when the traffic light may be obscured. We provide results on a collection of challenging traffic light classification examples [20], where the behavior of other vehicles is used to obtain additional information about the state of the intersection. Our results show that our approach outperforms prevailing approaches that only use direct image data of the traffic light.

II. BACKGROUND

We base our representation of an agent-aware state estimation problem on a *factored n-agent decentralized partially observable Markov decision process* (Dec-POMDP). A factored n-agent Dec-POMDP is a formal decision making framework for reasoning in partially observable, stochastic

environments with n agents using a factored state representation [18], [21]. Despite not explicitly using the decision making component of this formulation, it has a useful belief update step that when combined with a known policy allows for accurate and efficient state estimation.

A factored n -agent Dec-POMDP can be defined by a tuple $\langle \mathcal{I}, \mathcal{S}, \mathcal{A}, \mathcal{T}, \mathcal{R}, \Omega, O \rangle$, where:

- \mathcal{I} is a set of n agents.
- $\mathcal{S} = \mathcal{S}_0 \times \mathcal{S}_1 \times \dots \times \mathcal{S}_n$ is a finite set of factored states: a global state space \mathcal{S}_0 and a local state space $\mathcal{S}_{i>0}$ for each agent $i \in \mathcal{I}$.
- $\mathcal{A} = \mathcal{A}_1 \times \dots \times \mathcal{A}_n$ is a finite set of joint actions: an action set \mathcal{A}_i for each agent $i \in \mathcal{I}$.
- $\mathcal{T} : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \rightarrow [0, 1]$ is a transition function that represents the probability $\mathcal{T}(s'|s, \vec{a}) = Pr(s'|s, \vec{a})$ of reaching factored state $s' \in \mathcal{S}$ after performing joint action $\vec{a} \in \mathcal{A}$ in factored state $s \in \mathcal{S}$.
- $\mathcal{R} : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \rightarrow \mathbb{R}$ is a reward function that represents the expected immediate reward $R(s, \vec{a}, s')$ of reaching factored state $s' \in \mathcal{S}$ after performing joint action $\vec{a} \in \mathcal{A}$ in factored state $s \in \mathcal{S}$.
- $\Omega = \Omega_1 \times \dots \times \Omega_n$ is a finite set of joint observations: an observation set Ω_i for each agent $i \in \mathcal{I}$.
- $O : \mathcal{A} \times \mathcal{S} \times \Omega \rightarrow [0, 1]$ is an observation function that represents the probability $O(\vec{z}|\vec{a}, s) = Pr(\vec{z}|\vec{a}, s)$ of joint observation $\vec{z} \in \Omega$ after taking joint action $\vec{a} \in \mathcal{A}$ and ending up in factored state $s \in \mathcal{S}$.

If the current state $s \in \mathcal{S}$ can be determined by the latest joint observation $\vec{z} \in \Omega$, a Dec-POMDP is called a Dec-MDP.

A factored n -agent Dec-MDP is *locally fully observable* if each agent i 's observation fully determines their current local state s_i and global state s_0 .

A factored n -agent Dec-MDP is *transition independent* if the transition function \mathcal{T} can be represented by a tuple of transition probabilities $(\mathcal{T}_0, \mathcal{T}_1, \dots, \mathcal{T}_n)$ such that

$$\mathcal{T}(s, \vec{a}, s') = \mathcal{T}_0(s'_0|s_0) \prod_{i=1}^n \mathcal{T}_i(s'_i|s_i, a_i).$$

Transition independent Dec-MDPs are simpler because they obviate joint models.

III. AGENT-AWARE STATE ESTIMATION

We now introduce our approach to agent-aware state estimation. In the *agent-aware state estimation problem*, a *neutral observer* must recover a *signal* that may not always be directly observable. In addition to the neutral observer, there are multiple *agents* completing independent tasks in an environment. The environment is composed of a *global state* that is observed by all agents—which includes the signal—and a *local state* that is observed by each agent. By making noisy observations of the agents behaving in the environment, the neutral observer must recover the signal.

To perform estimation, we start with a specific kind of Dec-MDP, pair it with agent models, convert the problem

Algorithm 1 A method for agent-aware state estimation

```
1: function INITIALIZE(signalized Dec-MDP  $\mathcal{M}$ , policy
   set  $\Pi$ , observation set  $\mathcal{O}$ , observation function  $\mathcal{Z}$ )
2:   AASEP  $A \leftarrow \text{CREATEAASEP}(\mathcal{M}, \Pi, \mathcal{O}, \mathcal{Z})$ 
3:   DBN  $D \leftarrow \text{MAPAASEPTODB}(A)$ 
4:   return  $D$ 

5: function INFERENCE(DBN  $D$ , observation history  $\mathbf{H}$ ,
   time limit  $T$ , state prior  $P$ )
6:   state estimate  $\tilde{S} \leftarrow P$ 
7:   for  $t$  from 1 to  $T$  do
8:      $\tilde{S} \leftarrow \text{UPDATEESTIMATE}(D, \tilde{S}, \mathbf{H}_t)$  (sum product)
9:     signal estimate  $\tilde{G}_t \leftarrow \tilde{G}$  from  $\tilde{S}$ 
10:  return  $\tilde{G}$ 
```

to a DBN, and update our estimate of the signal with the most recent observation and previous estimate (summarized by Algorithm 1). We consider each step of the algorithm below, starting with the representation of the Dec-MDP.

We first cast the problem completed by the agents in the language of a factored n -agent Dec-MDP. Because the neutral observer is particularly concerned with the signal, we factor the state space into two independent regions, which we call signal space \mathcal{S}_G and agent space \mathcal{S}_N .

Definition 1. A factored n -agent Dec-MDP is *signalized* if the state space \mathcal{S} can be factored into $\mathcal{S}_G \times \mathcal{S}_N$ where $\mathcal{S}_G = \mathcal{S}_0$, $\mathcal{S}_N = \mathcal{S}_1 \times \dots \times \mathcal{S}_n$, and for every factored state $s \in \mathcal{S}$, joint action $\vec{a} \in \mathcal{A}$, and successor state $s' \in \mathcal{S}$, the transition function \mathcal{T} can be written as a tuple of transition probabilities $(\mathcal{T}_G, \mathcal{T}_N)$ such that

$$\mathcal{T}(s'|s, \vec{a}) = \mathcal{T}_G(s'_G|s_G) \mathcal{T}_N(s'_N|s_N, \vec{a}).$$

\mathcal{S}_G refers to the **signal space**, and \mathcal{S}_N refers to the **agent space**. We call $s_G \in \mathcal{S}_G$ a **signal**.

Given a signalized factored n -agent Dec-MDP, we now turn to a formal description of the agent-aware state estimation problem. Although a set of joint observations Ω and observation function \mathcal{O} has already been defined, these specify the probability of each agent making an observation of a state, not the probability of an independent observer watching them solve the problem. Therefore, along with the original problem solved by the agents and the policies used by each agent, the problem of the neutral observer includes a new set of observations and a new observation function.

Definition 2. An *agent-aware state estimation problem* (AASEP) is defined by the tuple $\langle \mathcal{M}, \Pi, \mathcal{O}, \mathcal{Z} \rangle$, where

- \mathcal{M} is a signalized factored n -agent Dec-MDP,
- $\Pi = \{\pi_1, \dots, \pi_n\}$ is a set of stochastic policy trees $\pi_i : \Omega^t \times \mathcal{A}_i \rightarrow [0, 1]$ such that $\pi_i(\vec{h}, a_i) = \Pr(a_i|\vec{h})$ defines the probability that agent i takes action a_i with history \vec{h} of t observations o_1, o_2, \dots, o_t .

- $\mathcal{O} = \mathcal{O}_0 \times \mathcal{O}_1 \times \dots \times \mathcal{O}_n$ is a finite set of joint observations: \mathcal{O}_0 are observations of a global state s_0 and $\mathcal{O}_{i>0}$ are observations of a local state s_i made by the neutral observer for each agent i .
- $\mathcal{Z} = \{\mathcal{Z}_0, \mathcal{Z}_1, \dots, \mathcal{Z}_n\}$ is a set of observation functions $\mathcal{Z}_i : \mathcal{S}_i \times \mathcal{O}_i \rightarrow [0, 1]$ such that $\mathcal{Z}_i(s_i, o_i) = \Pr(o_i|s_i)$ defines the probability that the observer receives observation o_i from state factor s_i for each agent i .

The objective of an AASEP is to estimate the current signal $s \in \mathcal{S}_G$ from a history \mathbf{H} (composed of observations from \mathcal{O}) of neutral observations of agents behaving with respect to the signal and their environment over time. Note that in Definition 2 stochastic policy trees are used to model agent behavior. When modeling real-world agents outside of our control, there might be no guarantees on their behavior. The policies of each agent may not even be optimal or representative of the reward function of \mathcal{M} , unless \mathcal{M} is carefully constructed. Agents may also have complex interactions with each other that provide little information on the signal, and it thus may be unnecessarily difficult to construct an accurate joint model that is useful in practice.

It is possible to avoid modeling the complexity of these interactions and allow for real-time inference by assuming that the agents behave in a truly decentralized manner, have full local observation, and operate in an environment that can cleanly be separated into independent components; and then working to ensure these assumptions hold on the agents you select for inference. This allows us to specify a new agent-aware state estimation problem where \mathcal{M} is both signalized and transition-independent:

Definition 3. A *transition-independent agent-aware state estimation problem* (TI-AASEP) is an AASEP $\langle \mathcal{M}, \Pi, \mathcal{O}, \mathcal{Z} \rangle$, where \mathcal{M} is transition-independent and locally fully observable, and Π is composed of policies that map from $\mathcal{S}_0 \times \mathcal{S}_i \times \mathcal{A}_i \rightarrow [0, 1]$ such that $\pi_i(s_0, s_i, a_i) = \Pr(a_i|s_0, s_i)$.

Note that any transition-independent n -agent Dec-MDP \mathcal{M} is also signalized, with $\mathcal{S}_G = \mathcal{S}_0$ and $\mathcal{S}_N = \mathcal{S}_1 \times \dots \times \mathcal{S}_n$. Moreover, Π in a TI-AASEP can be represented as a policy tree where $\pi_i(\vec{h}, a_i) = \Pr(a_i|o_i)$. This is sufficient, as \mathcal{M} is locally fully observable. Thus, TI-AASEPs are AASEPs.

With a TI-AASEP representation, many difficulties of the more general form are resolved. Local full observation allows the use of a Markovian policy representation, and the transition independence of \mathcal{M} disallows interactions between agents. This leads to a particularly tractable dynamic Bayesian network—one with discrete variables and a constant-sized conditional probability table at each node. Our observations \mathcal{O}_0 through \mathcal{O}_n form evidence nodes, and the global state space state factor \mathcal{S}_0 is our target for inference. The probabilistic relationships between the evidence nodes and \mathcal{S}_0 are characterized through observation functions, transition functions, and stochastic agent policies.

Figure 2 shows the two-slice DBN representation of a

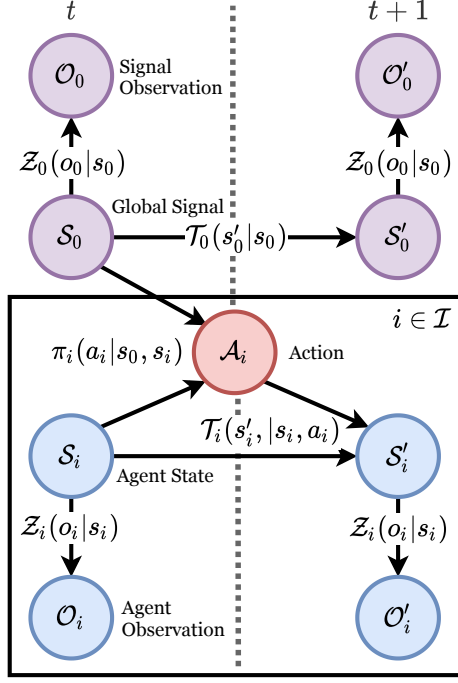


Fig. 2: The generic two-slice dynamic Bayesian network of the transition-independent agent-aware state estimation problem.

transition-independent AASEP. The *purple* nodes represent an ordinary signal estimation HMM, which models a signal s_0 's relationship with a direct observation o_0 . In *blue*, we have the (also standard) tracking task of locating agents in our environment. In *red*, we have the agent model, which relates the two tasks through a variable representing actions that depend on both prior local state and signal. Due to the causal dependence of all O_i on S_0 , we can improve our estimate of S_0 with far more data than if we were restricted to only O_0 , direct observations of the signal.

To efficiently perform inference of the signal using the DBN, we can convert it into a factor graph and perform the sum-product message passing algorithm [22].

Proposition 1. *Inference of a signal s_g over t timesteps using n agents in a TI-AASEP by applying the forward-backwards sum-product message passing algorithm to the DBN has worst-case time complexity $O(tnk^2)$, where each node can take up to k different discrete values.*

Proof (Sketch) 1. *The factor-graph representation of our t -timestep unrolled DBN has $O(tn)$ factors and $O(tn)$ variables, for each node and evidence-node respectively. Each variable performs at most k additions, over at most a k -dimensional distribution, taking $O(k^2)$ time to summarize a constant number of incoming messages. As the factor graph is a tree of finite width, exactly one message needs to be passed per edge for an exact solution. Therefore, the algorithm executes in $O(tnk^2)$ time, with one summarization per variable.*

The worst-case time complexity $O(tnk^2)$ provides a desirable linear dependence on the number of agents. While no proof is given, the worst-case time complexity of exact inference in the general AASEP case has a superlinear dependency on n . This, along with the impracticality of an accurate joint model that models interactions between up to n agents simultaneously, is unlikely to be reliable or real-time except for very small n and t .

In practice, clearly interacting or obstructed agents can be removed from inference. This justifies use of a locally-observable transition-independent Dec-MDP, at the cost of not always using every observation available. Even in the limit of removing every agent, the model simplifies to an HMM over the signal, which has been a standard approach to signal estimation.

Proposition 2. *Inference of a signal s_g over t time steps using zero agents in a TI-AASEP is equivalent to inference of s_g in an HMM, where S_0 is the set of hidden states and O_0 is the set of observations.*

Proof (Sketch) 2. *An HMM is a tuple $\langle S, \mathcal{T}, \mathcal{O}, \mathcal{Z} \rangle$, where S is a state space that transitions based on a transition function \mathcal{T} and emits observations from a set \mathcal{O} based on the observation function \mathcal{Z} . With no agents, a TI-AASEP only contains global state factors (all elements with nonzero subscript disappears), and $\langle S_0, \mathcal{T}_0, \mathcal{O}_0, \mathcal{Z}_0 \rangle$ forms an HMM.*

IV. EFFECTIVE TRAFFIC LIGHT CLASSIFICATION

In this section, we apply our approach to building and solving agent-aware state estimation problems to the task of traffic light classification in Figure 1. This requires defining an n -agent Dec-MDP, creating a TI-AASEP, and building the corresponding dynamic Bayesian network. Representing traffic light classification in this way facilitates efficient and accurate inference in real-time decision making problems.

We begin by constructing a signalized, transition-independent locally fully observable n -agent Dec-MDP $\mathcal{M} = \langle \mathcal{I}, S, \mathcal{A}, \mathcal{T}, \mathcal{R}, \Omega, \mathcal{O} \rangle$, where each observed, unobstructed vehicle at the intersection (excluding our own) is placed in the set of agents \mathcal{I} . The global state in S_0 in this case only contains the signal, S_g that represents the traffic light system {RED, GREEN, YELLOW} for each legal direction of travel (such that for any two intersecting legal directions of travel, at most one is not red). Due to the transition-independence requirement, any agent identified to be interacting with a pedestrian or another obstruction should be removed from \mathcal{I} before inference.

Each agent $i \in \mathcal{I}$ has a corresponding local state S_i , which is the cross product of position discretized as {ATINTERSECTION, TURNINGLEFT, DRIVINGSTRAIGHT, TURNINGRIGHT} and velocity discretized as {NONE (0 m/s), LOW (1-5 m/s), HIGH (>5 m/s)}. We represent actions as a cross product between steering inputs {LEFT, STRAIGHT, RIGHT} and accelerator inputs {MINUS, ZERO, PLUS}. The transition probabilities \mathcal{T}_i are specified via a

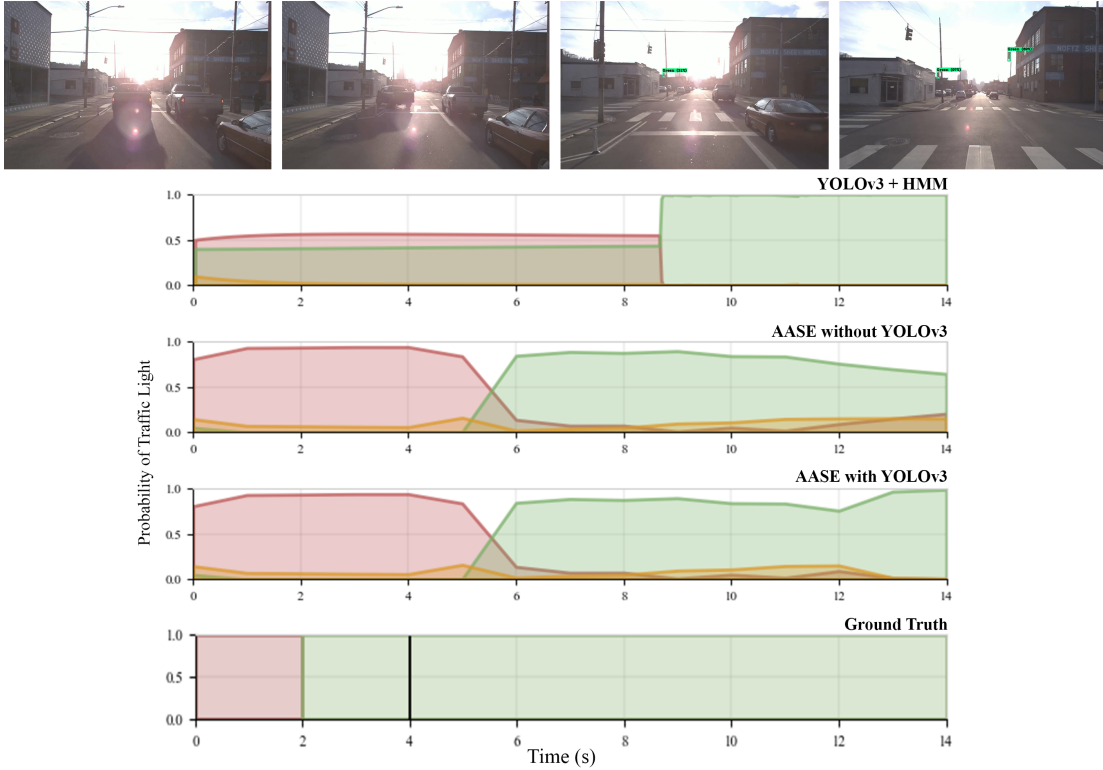


Fig. 3: A select scenario from the Argoverse dataset demonstrating an intersection where the light is obscured by glare. The ground truth transition from *red* to *green* occurs at two seconds, and a black line at four seconds indicates the first visible human reaction. The delay between the human and AASE response is primarily due to velocity and time discretization, both of which can be tightened.

simple approximate physics-based model for position and velocity, taking into account steering inputs and accelerator inputs in a straightforward manner. The transition probabilities \mathcal{T}_0 for the global signal were composed of a high probability self-transition, a low probability transition to the next light configuration in a predefined sequence, and a near zero probability for out-of-sequence light changes.

In order to construct the TI-AASEP, a set of joint observations \mathcal{O} representing vision of the light and other vehicles was required. Each $\mathcal{O}_{i>0}$ equals $\mathcal{S}_{i>0}$, as LiDAR was used to reconstruct estimates of position and velocity directly, which were then discretized. As vehicles can only directly observe the traffic light controlling their direction of travel, $\mathcal{O}_0 = \{\text{GREEN, YELLOW, RED}\}$. With these observations alone, the full global signal cannot be reconstructed, as it includes a separate light for each direction of travel. In addition, we defined a set of observation functions \mathcal{Z} to account for sensor uncertainty and obstructions. We specified \mathcal{Z}_0 to have low but nonzero noise and $\mathcal{Z}_{i>0}$ as noisier. This allows our model to trust direct observations of the light more when available. The final component is Π , the set of policies that map local states to actions. Each agent policy π_i was specified via a simple driver model that assumes drivers are fairly stochastic but largely attempt to obey traffic laws. Using the constructed DBN, we are able to perform inference as specified in Algorithm 1.

V. EXPERIMENTS

We compare our approach against an HMM smoothed vision baseline on a real world collection of rare traffic light occurrences. Our baseline vision model is a *YOLOv3* CNN architecture, which we trained on the *Bosch Small Traffic Lights* dataset [23], [24] that is smoothed with a simple traffic light HMM similar to prior work [16]. To perform an ablation study, we remove the direct CNN traffic light observations from our DBN so that the agent-based component can be evaluated separately. Each model is evaluated on a small collection of challenging four-way intersection scenarios from the *Argoverse* dataset [20]. The *YOLOv3* + HMM model runs on 30 Hz image data, and the AASE models run on a 1 Hz sample of the 10 Hz LiDAR data.

Figure 3 shows the results of our approach on a challenging example where glare from the sun has completely obscured the traffic light signal. The *YOLOv3* + HMM model fails to detect any visual signal until 9 seconds into the scenario. Until then, the HMM simply outputs the prior distribution, indicating total failure. Both AASE approaches have a much higher confidence in the red light and detect the transition to green sooner due to their observations of other cars not obscured by the glare.

VI. DISCUSSION

Our results highlight the advantage that our approach offers in scenarios where direct observation of a global signal

has failed. In Figure 3, YOLO cannot detect the traffic light for the first 9 seconds due to glare from the sun. Thus, if the AV were using YOLO by itself, it would not have driven into the intersection and discovered that the light was green. By contrast, the AASE framework not only has a high confidence of a red light at the start of the scenario but also reaches a high confidence of a green light after 6 seconds, only 2 seconds after the first car begins to move. The delay between the ground truth and the AASE models is mostly due to the delayed reaction of other drivers: we see that no other vehicle moves until 4 seconds (denoted by a black bar on the ground truth graph in Figure 3). Finally, for the ablation study, we remove YOLO observations from the AASE model. Without a direct observation of the light or any vehicles, the probability of a green light decays and the probabilities of yellow and red rise over time as expected. However, once YOLO is added, its confidence remains high even after the vehicles disappear from the intersection due to the direct observations that occur in those time steps.

Overall, this example demonstrates the benefits of using an encompassing framework like AASE. Practically, it takes advantage of information already collected by AVs (such as the position of other vehicles) and allows the use of sensors with nearly independent failure modes. In particular, methods such as LiDAR and radar are robust to glare. Radar is robust in snow and rain, and even cameras are far less likely to have multiple entire vehicles obstructed by glare or rain than the light itself. However, our method is not simply sensor fusion; we do not require the use of additional sensor modalities, nor do we use multiple sensors to directly sense the same object. Rather, we take advantage of the unique benefits of each sensor modality, combine all of the information from various sensors with other AV capabilities, such as predictive driver models, and form a holistic understanding of the state of a given intersection tractably in real-time.

VII. CONCLUSION

We propose a model for traffic light classification that accounts for not only direct observations of a traffic light but also the behavior of other rational actors. Using this approach, we can reason about the state of the traffic light even when the light may be obstructed in traditionally data poor scenarios. An analysis of this approach reveals that it collapses to standard methods that use a CNN and HMM in the presence of no agents, linearly scales in complexity in the number of agents, and results in theoretically correct inference from a collaborative decision making perspective. This offers a strict improvement over current methods, by using multi-agent behavior as a backup when ordinary perception fails. We show that our model successfully classifies a traffic light in real world scenarios when glare from the sun causes a CNN to fail. Finally, because relying on agent observations has different failure modes than direct observations, the scenarios in which state estimation fails can be reduced drastically without adding new sensing modalities.

REFERENCES

- [1] A. Broggi, P. Cerri, M. Felisa, M. C. Laghi, L. Mazzei, and P. P. Porta, "The VisLab intercontinental autonomous challenge: An extensive test for a platoon of intelligent vehicles," *International Journal of Vehicle Autonomous Systems*, 2012.
- [2] J. Svegliato, K. H. Wray, S. J. Witwicki, J. Biswas, and S. Zilberstein, "Belief space metareasoning for exception recovery," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2019.
- [3] C. Basich, J. Svegliato, K. H. Wray, S. Witwicki, J. Biswas, and S. Zilberstein, "Learning to optimize autonomy in competence-aware systems," *arXiv preprint arXiv:2003.07745*, 2020.
- [4] S. Zilberstein, R. Washington, D. S. Bernstein, and A.-I. Mouaddib, "Decision-theoretic control of planetary rovers," in *International Seminar on Advances in Plan-Based Control of Robotic Agents*. London, UK: Springer-Verlag, 2002.
- [5] B. C. Williams, M. D. Ingham, S. H. Chung, and P. H. Elliott, "Model-based programming of intelligent embedded systems and robotic space explorers," *IEEE Special Issue on Modeling and Design of Embedded Software*, 2003.
- [6] J. Casper and R. R. Murphy, "Human-robot interactions during the robot-assisted urban search and rescue response at the World Trade Center," *IEEE Transactions on Systems, Man, and Cybernetics*, 2003.
- [7] M. A. Goodrich, B. S. Morse, D. Gerhardt, J. L. Cooper, M. Quigley, J. A. Adams, and C. Humphrey, "Supporting wilderness search and rescue using a camera-equipped mini UAV," *Journal of Field Robotics*, 2008.
- [8] L. Pineda, T. Takahashi, H.-T. Jung, S. Zilberstein, and R. Grupen, "Continual planning for search and rescue robots," in *15th IEEE/RAS International Conference on Humanoid Robots*, Seoul, Korea, 2015.
- [9] S. Zilberstein, "Building strong semi-autonomous systems," in *29th AAAI Conference on Artificial Intelligence*, 2015.
- [10] L. P. Kaelbling, M. L. Littman, and A. R. Cassandra, "Planning and acting in partially observable stochastic domains," *Artificial intelligence*, 1998.
- [11] C. V. Goldman and S. Zilberstein, "Optimizing information exchange in cooperative multi-agent systems," in *2nd International Joint Conference on Autonomous Agents and Multiagent Systems*, 2003.
- [12] G. D. Forney, "The Viterbi algorithm," *Proceedings of the IEEE*, 1973.
- [13] S. V. Vaseghi and B. P. Milner, "Noise compensation for hidden Markov model speech recognition in adverse environments," *IEEE Transactions on Speech and Audio Processing*, 1997.
- [14] Z. Ghahramani, "An introduction to hidden Markov models and bayesian networks," in *Hidden Markov models: Applications in computer vision*. World Scientific, 2001.
- [15] P. Dymarski, *Hidden Markov Models: Theory and Applications*. BoD Books on Demand, 2011.
- [16] A. E. Gomez, F. A. Alencar, P. V. Prado, F. S. Osorio, and D. F. Wolf, "Traffic lights detection and state estimation using hidden Markov models," in *IEEE Intelligent Vehicles Symposium*. IEEE, 2014.
- [17] F. A. Oliehoek, C. Amato, et al., *A concise introduction to decentralized POMDPs*. Springer, 2016.
- [18] D. S. Bernstein, R. Givan, N. Immerman, and S. Zilberstein, "The complexity of decentralized control of Markov decision processes," *Mathematics of Operations Research*, 2002.
- [19] K. P. Murphy, "Dynamic bayesian networks: representation, inference and learning," *PhD Thesis*, 2002.
- [20] M.-F. Chang, J. Lambert, P. Sangkloy, J. Singh, S. Bak, A. Hartnett, D. Wang, P. Carr, S. Lucey, D. Ramanan, and J. Hays, "Argoverse: 3d tracking and forecasting with rich maps," in *IEEE Conference on Computer Vision and Pattern Recognition*, June 2019.
- [21] C. Amato, G. Chowdhary, A. Geramifard, N. K. Üre, and M. J. Kochenderfer, "Decentralized control of partially observable Markov decision processes," in *52nd IEEE Conference on Decision and Control*. IEEE, 2013.
- [22] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Transactions on Information Theory*, pp. 498–519, 2001.
- [23] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *arXiv:1804.02767*, 2018.
- [24] K. Behrendt, L. Novak, and R. Botros, "A deep learning approach to traffic lights: Detection, tracking, and classification," in *2017 IEEE International Conference on Robotics and Automation*. IEEE, 2017.