computers

MDPI

Article

Modeling Autonomous Vehicle Responses to Novel Observations Using Hierarchical Cognitive Representations Inspired Active Inference †

Sheida Nozari [1,*], Ali Krayani [1], Pablo Marin [2], Lucio Marcenaro [1], David Martin Gomez [2] and Carlo Regazzoni [1]

[1] Department of Electrical, Electronic, Telecommunications Engineering and Naval Architecture, University of Genoa, 16126 Genoa, Italy; ali.krayani@edu.unige.it (A.K.); lucio.marcenaro@unige.it (L.M.); carlo.regazzoni@unige.it (C.R.)
[2] Department of Systems Engineering and Automation, University Carlos III of Madrid, 28903 Madrid, Spain; pamarinp@ing.uc3m.es (P.M.); dmgomez@ing.uc3m.es (D.M.G.)
* Correspondence: sheida.nozari@edu.unige.it
† This paper is an extended version of our paper published in the 6th International Conference on System-Integrated Intelligence (SysInt 2022), 7–9 September 2022, Genova, Italy.

Abstract: Equipping autonomous agents for dynamic interaction and navigation is a significant challenge in intelligent transportation systems. This study aims to address this by implementing a brain-inspired model for decision making in autonomous vehicles. We employ active inference, a Bayesian approach that models decision-making processes similar to the human brain, focusing on the agent's preferences and the principle of free energy. This approach is combined with imitation learning to enhance the vehicle's ability to adapt to new observations and make human-like decisions. The research involved developing a multi-modal self-awareness architecture for autonomous driving systems and testing this model in driving scenarios, including abnormal observations. The results demonstrated the model's effectiveness in enabling the vehicle to make safe decisions, particularly in unobserved or dynamic environments. The study concludes that the integration of active inference with imitation learning significantly improves the performance of autonomous vehicles, offering a promising direction for future developments in intelligent transportation systems.

Keywords: active inference; Bayesian learning; imitation learning; action-oriented model; world model; autonomous driving

check for updates

Citation: Nozari, S.; Krayani, A.; Marin, P.; Marcenaro, L.; Gomez, D.M.; Regazzoni, C. Modeling Autonomous Vehicle Responses to Novel Observations Using Hierarchical Cognitive Representations Inspired Active Inference. *Computers* **2024**, *13*, 161. https://doi.org/10.3390/computers13070161

Academic Editor: Stefan Bosse

Received: 24 November 2023
Revised: 29 January 2024
Accepted: 9 February 2024
Published: 28 June 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The rapidly developing field of autonomous driving systems (ADS) has the potential to revolutionize transportation in smart cities [1,2]. One of the main challenges that ADS face is their ability to adapt to novel observations in a constantly changing environment. This challenge is due to the inherent complexity and uncertainty of the real world, which may lead to unexpected and unpredictable situations.

Driving scenarios involve multiple unpredictable factors, such as diverse driver behaviors and environmental fluctuations. These challenges require a shift from traditional rule-based learning models [3] to adaptive and cognitive entities, enabling autonomous vehicles (AVs) to navigate through complex and unpredictable terrains effectively.

Cognitive learning is a promising approach to tackle the challenge of adapting to novel situations in a dynamic environment [4]. This approach aligns with the principles of Bayesian brain learning [5], which suggests that the human brain operates as a Bayesian inference system, constantly updating its probabilistic models to perform a wide range of cognitive tasks such as perception, planning, and learning. This allows autonomous agents to update their beliefs about the external world in response to novel observations.

In the context of ADS, integrating these concepts leads to the emergence of vehicles with a cognitive hierarchy. One of the open questions in this area is to what extent autonomous vehicles can move beyond mere rule-following or imitating prior sub-optimal experiences. The hierarchical principles go beyond rule-based behavior, allowing autonomous vehicles to perceive their surroundings through probabilistic perception, infer causal structures, predict future states, and act with agency. This provides an adaptive learning agent that continuously enhances its cognitive models through experiences.

At the heart of the cognitive hierarchy of AVs is the critical role played by generative models (GMs) [6]. These models allow AVs to comprehend the underlying dynamics of the environment, which, in turn, empowers them to anticipate future behavior and engage in proactive decision making [7]. By relying on GMs, AVs go beyond simple perception and use predictive inferences to represent the agent's beliefs about the world. Furthermore, the autonomous agent must be able to plan a sequence of actions that will help it gain information and reduce uncertainty about the surroundings.

Active inference [8], grounded in the Bayesian brain learning paradigm, is a computational framework bridging the gap between perception and action. It suggests that an intelligent agent, such as an autonomous vehicle (AV), should not just passively observe its environment but actively engage in exploratory actions to refine its internal probabilistic models. This is achieved through a continuous cycle of observation, belief updating, and action selection. The process begins with perception, where the AV uses an array of sensors to engage in the multisensory and multimodal observation of its surroundings, accumulating sensory evidence about the external world. This sensory evidence is then integrated into a probabilistic model, commonly represented as a belief distribution. The belief distribution encapsulates the vehicle's understanding of the environment, encompassing not only the current state but also a spectrum of potential states, inherently acknowledging uncertainty.

In active inference, messages are propagated between different layers of a probabilistic model, allowing for the exchange of information. These messages are probabilistic updates that convey crucial insights about the environment's dynamics and help to refine the AV's internal probabilistic representation. In addition, active inference introduces the concept of action selection guided by inference. Rather than responding automatically, the AV engages in a deliberate action-planning process. The agent chooses actions that maximize the expected sensory evidence while also accounting for their internal beliefs and uncertainty, contributing to the minimization of free energy [9].

Furthermore, active inference is a powerful technique that enables autonomous agents to develop a sense of self-awareness [10]. By continuously comparing their sensory observations with their internal beliefs, these agents strive to minimize free energy and gain an understanding of their own knowledge and the limits of their perception. This self-awareness allows them to recognize when to seek additional information through exploratory actions and when they can rely on their existing knowledge to make decisions. Simply put, active inference helps autonomous agents become cognitively self-aware while also compelling them to minimize the difference between their internal models and external reality. This allows them to adapt and learn from their environment, make better decisions, and navigate uncertain and complex situations.

Motivated by the previous discussion, we propose a cognitive hierarchical framework for modeling AV responses to abnormalities, such as novel observations, during a lane-changing scenario, based on active inference. The AV, equipped with self-awareness, should learn to self-drive in a dynamic environment while interacting with other participants. The proposed framework consists of two essential computational modules: a perception module and an adaptive learning module. The perception module analyzes the sensory signals and creates a model based on the observed interaction between the participants in a dynamic environment. This allows the AV to perceive the external world as a bundle of exteroceptive and proprioceptive sensations from multiple sensory modalities and to integrate information from different sensory inputs and match them appropriately. In this work, the AV integrates proprioceptive stimuli (i.e., the AV's positions) with exteroceptive stimuli (i.e., the relative distance

between the AV and another object) and describes the integration process using Bayesian inference. The adaptive learning module consists of a world model and an active model. The world model is essential to the cognitive processes of perception, inference, prediction, and decision making in active inference systems. It bridges the agent's internal representation and interactions with the external environment, enabling the agent to adapt its behavior to uncertain scenarios. The active model plans the AV's actions as inferred from the world model in terms of minimizing the cost function due to uncertainty (i.e., unforeseen situations), and its components are linked by the active inference process.

The main contributions of this paper can be summarized as follows:

- We present a comprehensive hierarchical cognitive framework for autonomous driving, addressing the challenge of responding to novel observations in dynamic environments. This framework marks a fundamental shift from rule-based learning models to cognitive entities capable of allowing AVs to navigate unforeseen terrains.
- The proposed framework is firmly grounded in the principles of Bayesian learning, enabling ADS to adapt its probabilistic models continually. This adaptation is essential for the continuous improvement of the cognitive model through experiences. Consequently, an AV can consistently update its beliefs regarding the surroundings.
- We expand upon a global dictionary to incrementally develop a dynamic world model during the learning process. This world model efficiently structures newly acquired environmental knowledge, enhancing AV perception and decision making.
- Through active inference, the proposed approach equips the AV with a sense of self-awareness by continually comparing sensory observations with internal beliefs and aiming to minimize free energy. This self-awareness enables them to make informed decisions about seeking additional information through exploratory actions and when to rely on existing knowledge.
- The dynamic interaction between the ego AV and its environment, as facilitated by active inference, forms the basis for adaptive learning. This adaptability augments the AV's decision-making capabilities, positioning it as a cognitive entity capable of navigating confidently and effectively in uncertain and complex environments.

2. Related Works

One of the most pivotal advancements in intelligent transportation is the development of autonomous driving (AD). AVs are defined as agents capable of navigating from one location to another without human control. These vehicles perceive their surroundings using a variety of sensors, processing this information to operate independently of human intervention [11]. AD involves addressing challenges in perception and motion planning, particularly in environments with dynamic objects. The complex interactions among multiple agents pose significant challenges, primarily due to the unpredictability of their future states. Most model-based AD strategies require the manual creation of driving policy models [12,13], or they incorporate safety assessments to assist human drivers [14,15].

In recent years, there has been a substantial increase in the demand for AVs that can imitate human behavior. Advances in ADS have opened up a wide range of potential applications where an agent is required to make intelligent decisions and execute realistic motor actions in diverse scenarios. A key aspect of future developments in AVs hinges on the agent's ability to perform as an expert in similar situations. Research indicates that utilizing expert knowledge is more effective and efficient than starting from scratch [16–18]. One practical method for transferring this expertise is through providing optimal demonstrations of the desired behavior for the learning agent (L) to replicate [19].

Imitation Learning (IL) involves acquiring skills or behaviors by observing an expert perform a specific task. This approach is vital to the development of machine intelligence, drawing inspiration and foundational concepts from cognitive science. IL has long been considered a crucial component in the evolution of intelligent agents [17].

IL is similar to standard supervised learning, but instead of pairing features with labels, it pairs states with actions. In IL, a state represents the agent's current situation and

the condition of any target object involved. The IL process typically begins with collecting example demonstrations from an expert agent (E), which are then translated into state-action pairs. However, simply learning a direct state-to-action relationship isn't enough to ensure the desired behavior. Challenges such as errors in demonstration collection or a lack of comprehensive demonstrations can arise [20]. Additionally, the learner's task might slightly differ from the demonstrated one due to environmental changes, obstacles, or targets. Therefore, IL often includes a step where the learner applies the learned actions and adjusts its approach based on task performance.

The existing works of the IL approach for driving can handle simple driving tasks such as lane following [21,22]. However, if the agent is dealing with a new environment or a more complicated task (such as lane changing), the human driver must take control, or the system ultimately fails [23,24]. More specifically, a typical IL procedure is direct learning, where the main goal is to learn a mapping from states to actions that mimic the demonstrator explicitly [25,26]. Direct learning methods are categorized into classification methods when the learner's actions can be classified into discrete classes [27,28] and regression methods which are used to learn actions in a continuous space [29]. Direct learning often fails to reproduce the proper behavior due to issues such as insufficient demonstrations or the need to perform different tasks in changing environments. Additionally, indirect learning can complement direct approaches by refining the policies based on sub-optimal expert demonstrations [30].

The primary limitations of IL include the policy's inability to surpass the expert's suboptimal performance and its susceptibility to distributional shifts [31]. Consequently, IL often incorporates an additional step where the learning agent refines the estimated policy according to its current context. This self-improvement process can be guided by measurable rewards or learning from specific instances.

Many of these approaches fall under the umbrella of reinforcement learning (RL) methods. RL enables the encoding of desired behaviors, such as reaching a target and avoiding collisions, and does not solely rely on perfect expert demonstrations. Additionally, RL focuses on maximizing the expected return over an entire trajectory, unlike IL, which treats each observation independently [32]. This conceptual difference often positions RL as superior to IL. However, without prior knowledge from an expert, the RL learning agent may struggle to identify desired behaviors in environments with sparse rewards [33]. Furthermore, even when RL successfully maximizes rewards, the resulting policy may not align with the behaviors anticipated by the reward designer. The trial-and-error nature of RL also necessitates task-specific reward functions, which can be complex and challenging to define in many scenarios.

Learning approaches like IL and RL can be complex without adequate representation or model learning from the environment. To address these challenges, autonomous systems often adopt an incremental learning approach. This method enables the agent to acquire new knowledge while retaining previously learned information [34,35]. As a result, the learning agent becomes capable of processing and understanding new situations that it encounters over time.

3. Proposed Framework

The hierarchical cognitive schematic introduced in Figure 1 comprises several modules that form a cognitive cycle, preparing an AV to perceive its environment and interact with its surroundings. When the system faces a new situation (i.e., novel sensorial input), an AV interprets the external world by formulating and testing hypotheses about its evolution. It generates predictions based on prior knowledge acquired from past experiences, performs actions based on its beliefs, observes the outcomes, and refines the beliefs accordingly. The different modules in the architecture can be likened to different areas of the biological brain, each one handling particular functionalities. Some parts are responsible for sensory perception, while others are dedicated to planning and decision making. All parts are interconnected and operate together. The cognitive model is characterized by inferences

across different modules that enable it to predict the perceptual outcomes of actions. Moreover, the model must utilize these representations to minimize prediction errors and predict how sensory signals change for specific actions. The following sections present a detailed description of the different modules involved in the architecture.

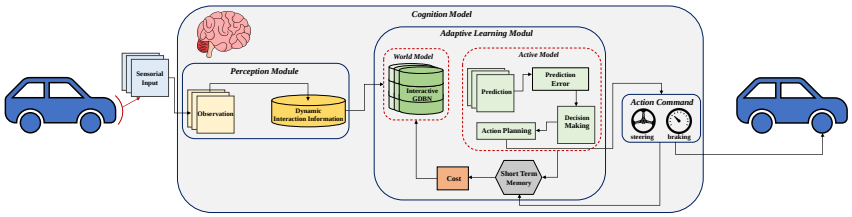


Figure 1. The proposed cognitive architecture.

3.1. Perception Module

To effectively operate in its environment, an autonomous vehicle (AV) requires a perception module that can learn how it interacts with other dynamic entities. This module takes in multi-sensorial information to identify causalities among the data perceived by the agent. Using various sensors to gather information is vital in constructing a model capable of predicting the agent's dynamics for motion planning. The ability to perceive multimodal stimuli is crucial, as it provides multimodal information under different conditions to augment the scene library of the ADS.

To accurately predict how information will be perceived, the module integrates exteroceptive and proprioceptive perceptions to formulate a contextual viewpoint. This viewpoint includes both internal and external perceptions of the agent. The primary aim is to use this information to predict subsequent internal or external states. To achieve this, the movements of both the AV and other participants are simulated at each instant through interacting rules dependent on their positions and motions, generating coupled trajectory data. Analyzing this multisensory data helps encode the dynamic interaction of the associated agents as probabilities into a coupled generative dynamic Bayesian network (C-GDBN). The resulting dynamic interaction model is self-aware and capable of identifying abnormalities and incrementally learning new interacting behaviors derived from an initial one, influencing the agent's decision making.

3.2. Adaptive Learning Module

In addition to the perception module, we have developed an adaptive learning module that enhances the AV's ability to respond to its surroundings. This module continuously analyzes the agent's interactions with the environment and adapts the AV's responses accordingly by acquiring incremental knowledge of the evolving context. This approach ensures the AV can proactively anticipate changes, make better decisions, and respond adeptly. Integrating the adaptive learning module represents a significant step forward in promoting an adaptive interaction between the AV and its surroundings. The module comprises two components: the world model and the active model.

3.2.1. World Model

The world model (WM) acts like a simulator in the brain, providing insights into how the brain learns to execute sensorimotor behaviors [36]. In the proposed architecture, the WM is formulated using generative models, leveraging interactive experiences derived from multimodal sensory information. Initially, the WM is established via the situation model (SM), serving as a fundamental input module (see Figure 2). The SM, represented as a Coupled GDBN (C-GDB), models the motions and dynamic interactions between two entities in the environment, enabling the estimation of the vehicles' intentions through probabilistic reasoning. This constructed C-GDBN demonstrates the gathered sub-optimal information concerning the interaction of an expert AV (E) with one vehicle (V1) in the same lane, where E changes lanes to overtake V1 without a collision (comprehensive details

on structuring the SM can be found in our previous work [37]). To initialize the WM by using the provided SM, we transfer the knowledge of E to a first-person perspective, where an intelligent vehicle (L) learns by interacting with its environment via observing the E's behavior to integrate the gained knowledge into its understanding of the surroundings.

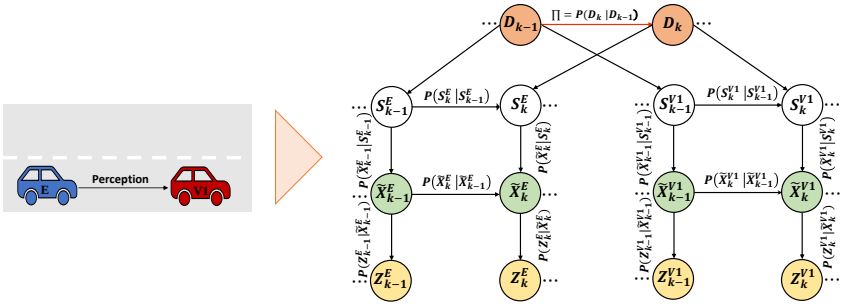


Figure 2. Coupled-GDBN representing the dynamic interactions between E and V1.

The first-person model (FP-M) establishes a dynamic model that shifts L from a third-person viewpoint to a first-person experience. This allows L to perceive driving tasks as E does, enhancing its imitation accuracy. Such a perspective empowers L to react promptly during interactions with V1. FP-M's structure is derived by translating the hierarchical levels of SM into the FP context (as illustrated in Figure 3). The top level of hierarchy in FP-M denotes pre-established configurations. $(\tilde{\mathbf{D}})$ from the dynamic behavior of how E and V1 interact in the environment. Each configuration represents a joint discrete state as:

$$\tilde{\mathbf{D}}_{\boldsymbol{k}} = \mathbf{f}(\tilde{\mathbf{D}}_{\boldsymbol{k-1}}) + \mathrm{w}_k, \tag{1}$$

where $\tilde{\mathbf{D}}_{\boldsymbol{k}}$ is a latent discrete state evolving from the previous state $\tilde{\mathbf{D}}_{\boldsymbol{k-1}}$ via a non-linear state evolution function $\mathbf{f}(\cdot)$ representing the transition dynamic model and via a Gaussian process noise $\mathrm{w}_k \sim \mathcal{N}(0, \mathbf{Q})$. The discrete state variables $\tilde{\mathbf{D}}_{\boldsymbol{k}} = [\mathrm{S}_k^{\mathrm{E}}, \mathrm{S}_k^{\mathrm{V1}}]$ represent jointly the discrete states of E and V1 where $\mathrm{S}_k^{\mathrm{E}} \in \boldsymbol{\mathcal{S}}^{\mathbf{E}}$, $\mathrm{S}_k^{\mathrm{V1}} \in \boldsymbol{\mathcal{S}}^{\mathbf{V1}}$, $\tilde{\mathbf{D}}_{\boldsymbol{k}} \in \boldsymbol{\mathcal{D}}$, and $\boldsymbol{\mathcal{S}}^{\mathbf{E}}$ and $\boldsymbol{\mathcal{S}}^{\mathbf{V1}}$ are learned according to the approach discussed in [38], while $\boldsymbol{\mathcal{D}} = \{\tilde{\mathbf{D}}_1, \tilde{\mathbf{D}}_2, \ldots, \tilde{\mathbf{D}}_m\}$ is the set that represents the dictionary consisting of all the possible joint discrete states (i.e., configurations) and m is the total number of configurations. Therefore, by tracking the evolution of these configurations over time, it is possible to determine the transition matrix that quantifies the likelihood of transitioning from one configuration to the next, as defined by the following:

$$\Pi = \begin{bmatrix} \mathrm{P}(\tilde{\mathbf{D}}_1|\tilde{\mathbf{D}}_1), & \ldots, & \mathrm{P}(\tilde{\mathbf{D}}_1|\tilde{\mathbf{D}}_m) \\ \vdots & \ddots & \vdots \\ \mathrm{P}(\tilde{\mathbf{D}}_m|\tilde{\mathbf{D}}_1), & \ldots, & \mathrm{P}(\tilde{\mathbf{D}}_m|\tilde{\mathbf{D}}_m) \end{bmatrix} \tag{2}$$

where $\Pi \in \mathbb{R}^{m,m}$, $\mathrm{P}(\tilde{\mathbf{D}}_i|\tilde{\mathbf{D}}_j)$ represents the transition probability from configuration i to configuration j and $\sum_{k=1}^{m} \mathrm{P}(\mathbf{D}_i|\tilde{\mathbf{D}}_k) = 1 \; \forall i$.

The hidden continuous states $(\tilde{X})$ in the FP-M represent the dynamic interaction in terms of generalized relative distance consisting of relative distance and relative velocity, which is defined as the following:

$$\tilde{\boldsymbol{X}}_k = [\tilde{\mathrm{X}}_k^{\mathrm{E}} - \tilde{\mathrm{X}}_k^{\mathrm{V1}}] = [(\mathrm{x}_k^{\mathrm{E}} - \mathrm{x}_k^{\mathrm{V1}}), (\dot{\mathrm{x}}_k^{\mathrm{E}} - \dot{\mathrm{x}}_k^{\mathrm{V1}})]. \tag{3}$$

The initialization is based on SM where the continuous latent state $\tilde{\mathbf{X}}_k = [\tilde{\mathrm{X}}_k^{\mathrm{E}}, \tilde{\mathrm{X}}_k^{\mathrm{V1}}] \in \mathbb{R}^{n_x}$ represents a joint belief state where $\tilde{\mathrm{X}}_k^{\mathrm{E}}$ and $\tilde{\mathrm{X}}_k^{\mathrm{V1}}$ denote the hidden generalized states (GSs) of E and V1, respectively. The GSs consist of the vehicles' position and velocity where $\tilde{\mathrm{X}}_k^i = [\mathrm{x}_k^i, \mathrm{y}_k^i, \dot{\mathrm{x}}_k^i, \dot{\mathrm{y}}_k^i]$ and $i \in \{\mathrm{E}, \mathrm{V1}\}$. The continuous variables $\tilde{\mathbf{X}}_k$ evolve from the previous state $\tilde{\mathbf{X}}_{k-1}$ via the linear state function $\mathbf{g}(\cdot)$ and via a Gaussian noise w_k, as follows:

$$\tilde{\mathbf{X}}_k = \mathbf{g}(\tilde{\mathbf{X}}_{k-1}, \mathbf{D}_k) = \mathbf{F}\tilde{\mathbf{X}}_{k-1} + \mathbf{B}\mathrm{U}_{\mathbf{D}_k} + \mathrm{w}_k, \tag{4}$$

where $\mathbf{F} \in \mathbb{R}^{n_x, n_x}$ is the state evolution matrix and $\mathrm{U}_{\tilde{\mathbf{D}}_k} = \dot{\mu}_{\tilde{\mathrm{D}}_k}$ is the control unit vector.

Likewise, the observations in the FP-M depict the measured relative distance between the two vehicles defined as $\mathbf{Z}_k = [\mathrm{Z}_k^{\mathrm{E}} - \mathrm{Z}_k^{\mathrm{V1}}]$, where $\mathbf{Z_k} \in \mathbb{R}^{n_z}$ is the generalized observation, which is generated from the latent continuous states via a linear function $\mathbf{h}(\cdot)$ corrupted by Gaussian noise $\nu_k \sim \mathcal{N}(0, \mathbf{R})$ as the following:

$$\mathbf{Z}_k = \mathbf{h}(\tilde{\mathbf{X}}_k) + \nu_k = \mathbf{H}\tilde{\mathbf{X}}_k + \nu_k. \tag{5}$$

Since the observation transformation is linear, there exists the observation matrix $\mathbf{H} \in \mathbb{R}^{n_z, n_z}$ mapping hidden continuous states to observations.

Consequently, within the FP framework, L can reproduce anticipated interactive maneuvers, serving as a benchmark to assess its interaction with V1. The concealed continuous states within the FP-M depict the dynamic interplay, characterized by a generalized relative distance that encompasses both relative distance and relative velocity.

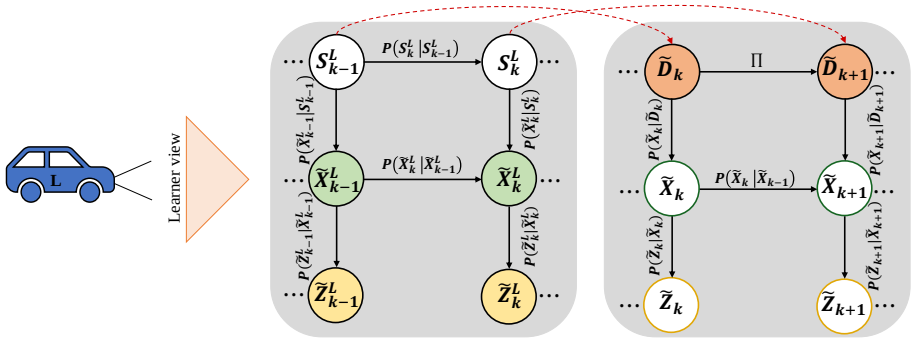


Figure 3. First-person model consists of a proprioceptive model (right side) and the learned joint configurations (left side) from the learning agent view.

3.2.2. Active Model

Active first-person model (AFP-M) links the WM to the decision-making framework which is associated with L behavior. This connection is achieved by augmenting the FP-M with active states representing the L's movements in the environment. Consequently, the AFP-M represents a generative model $\mathrm{P}(\tilde{\mathbf{Z}}, \tilde{\mathbf{X}}, \tilde{\mathbf{D}}, \boldsymbol{a})$ as illustrated graphically in Figure 4, which is conceptualized based on the principles of a partially observed Markov decision process (POMDP). The AFP-M encompasses joint probability distributions over observations, hidden environmental states at multiple levels, and actions executed by L, factorized as follows:

$$\mathrm{P}(\tilde{\mathbf{Z}}, \tilde{\mathbf{X}}, \tilde{\mathbf{D}}, \boldsymbol{a}) = \mathrm{P}(\tilde{\mathbf{D}}_\mathbf{0})\mathrm{P}(\tilde{\mathbf{X}}_\mathbf{0}) \prod_{k=2}^{K} \mathrm{P}(\tilde{\mathbf{Z}}_k|\tilde{\mathbf{X}}_k) \\ \mathrm{P}(\tilde{\mathbf{X}}_k|\tilde{\mathbf{X}}_{k-1}, \tilde{\mathbf{D}}_k)\mathrm{P}(\tilde{\mathbf{D}}_k|\tilde{\mathbf{D}}_{k-1}, a_{k-1})\mathrm{P}(a_{k-1}|\tilde{\mathbf{D}}_{k-1}). \tag{6}$$

In the context of a POMDP:

- L often relies on observations, formulated as $\mathrm{P}(\tilde{\mathbf{Z}}_k|\tilde{\mathbf{X}}_k)$, to deduce actual environmental states that are not directly perceived.
- L forms beliefs about the hidden environmental states, represented as $(\tilde{\mathbf{D}}_k, \tilde{\mathbf{X}}_k)$. These beliefs evolve according to $\mathrm{P}(\tilde{\mathbf{X}}_k|\tilde{\mathbf{X}}_{k-1}, \tilde{\mathbf{D}}_k)$ and $\mathrm{P}(\tilde{\mathbf{D}}_k|\tilde{\mathbf{D}}_{k-1}, a_{k-1})$.
- L engages with its surroundings by choosing actions that minimize the abnormalities and prediction errors.

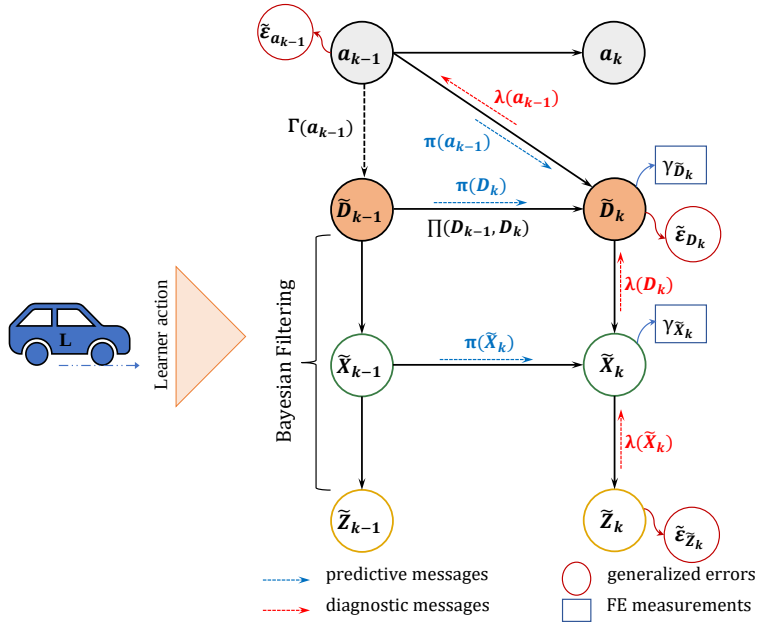


Figure 4. Graphical representation of the active first-person model.

Joint Prediction and Perception

In the initial stage of the process (at $k = 1$), L employs prior probability distributions, denoted as $\mathrm{P}(\tilde{\mathbf{D}}_0)$ and $\mathrm{P}(\tilde{\mathbf{X}}_0)$, to predict environmental states. This prediction is realized through the expressions $\tilde{\mathbf{D}}_0 \sim \mathrm{P}(\tilde{\mathbf{D}}_0)$ and $\tilde{\mathbf{X}}_0 \sim \mathrm{P}(\tilde{\mathbf{X}}_0)$. The methodological framework for the prediction is grounded in a sophisticated hybrid Bayesian filter, specifically the modified Markov jump particle filter (M-MJPF) [39], which integrates the functionalities of both particle filter (PF) and Kalman filter (KF). As the process progresses beyond the initial stage (for $k > 1$), L leverages the previously accumulated knowledge about the evolution of configurations. This knowledge is encapsulated in the probability distribution $\mathrm{P}(\tilde{\mathbf{D}}_k|\tilde{\mathbf{D}}_{k-1})$, which is encoded in the transition matrix as outlined in (2). The PF mechanism propagates N particles, each assigned equal weight and derived from the importance density distribution $\pi(\tilde{\mathbf{D}}_k) = \mathrm{P}(\tilde{\mathbf{D}}_k|\tilde{\mathbf{D}}_{k-1}, a_{k-1})$. This process results in the formation of a particle set, represented as $\{\tilde{\mathbf{D}}_k^{(i)}, w_k^{(i)}\}_{i=1}^N$. Concurrently, a series of Kalman filters (KFs) is utilized for each particle in the set, facilitating the prediction of the corresponding continuous GSs, denoted as $\left\{\tilde{\mathbf{X}}_k^{(i)}\right\}_{i=1}^N$. The prediction of these GSs is directed by a higher level, as indicated in Equation (4), which can be articulated in probabilistic terms as $\mathrm{P}(\tilde{\mathbf{X}}_k^{(i)}|\tilde{\mathbf{X}}_{k-1}^{(i)}, \tilde{\mathbf{D}}_k^{(i)})$. The posterior distribution associated with these predicted GSs is characterized by the following description:

$$\pi(\tilde{\mathbf{X}}_k^{(i)}) = \mathrm{P}(\tilde{\mathbf{X}}_k^{(i)}, \tilde{\mathbf{D}}_k^{(i)}|\tilde{\mathbf{Z}}_{k-1}) = \int \mathrm{P}(\tilde{\mathbf{X}}_k^{(i)}|\tilde{\mathbf{X}}_{k-1}^{(i)}, \tilde{\mathbf{D}}_k^{(i)})\lambda(\tilde{\mathbf{X}}_{k-1}^{(i)})\mathrm{d}\tilde{\mathbf{X}}_{k-1}^{(i)}, \tag{7}$$

where $\lambda(\tilde{\mathbf{X}}_{k-1}^{(i)}) = \mathrm{P}(\tilde{\mathbf{Z}}_{k-1}|\tilde{\mathbf{X}}_{k-1}^{(i)})$ represents the diagnostic message that has been previously propagated, following the observation of $\tilde{\mathbf{Z}}_{k-1}$ at time $k-1$. This mechanism plays a crucial role in the algorithm's process: upon receiving a new observation $\tilde{\mathbf{Z}}_k$, a series of diagnostic messages are propagated in a bottom-up manner to update the L's belief about the hidden states. Consequently, the updated belief in the GSs is expressed as $\mathrm{P}(\tilde{\mathbf{X}}_k^{(i)}, \tilde{\mathbf{D}}_k^{(i)}|\tilde{\mathbf{Z}}_k) = \pi(\tilde{\mathbf{X}}_k^{(i)}) \times \lambda(\tilde{\mathbf{X}}_k^{(i)})$. In parallel, the belief in the discrete hidden states is

refined by adjusting the weights of the particles, as denoted by $w_k^{(i)} = w_k^{(i)} \times \lambda(\tilde{\mathbf{D}}_k)$, where $\lambda(\tilde{\mathbf{D}}_k)$ is defined as a discrete probability distribution.

$$\lambda(\tilde{\mathbf{D}}_k) = \left[\frac{\frac{1}{\lambda(\tilde{\mathbf{D}}_k^{(1)})}}{\frac{1}{\sum_{i=1}^{m} \lambda(\tilde{\mathbf{D}}_k^{(i)})}}, \frac{\frac{1}{\lambda(\tilde{\mathbf{D}}_k^{(2)})}}{\frac{1}{\sum_{i=1}^{m} \lambda(\tilde{\mathbf{D}}_k^{(i)})}}, \ldots, \frac{\frac{1}{\lambda(\tilde{\mathbf{D}}_k^{(m)})}}{\frac{1}{\sum_{i=1}^{m} \lambda(\tilde{\mathbf{D}}_k^{(i)})}} \right], \tag{8}$$

such that

$$\lambda(\tilde{\mathbf{D}}_k^{(i)}) = \lambda(\tilde{\mathbf{X}}_k^{(i)}) \mathrm{P}(\tilde{\mathbf{X}}_k^{(i)} | \tilde{\mathbf{D}}_k^{(i)}) = \mathcal{D}_B \left(\lambda(\tilde{\mathbf{X}}_k^{(i)}), \mathrm{P}(\tilde{\mathbf{X}}_k^{(i)} | \tilde{\mathbf{D}}_k^{(i)}) \right) = \\ - \ln \int \sqrt{\lambda(\tilde{\mathbf{X}}_k^{(i)}), \mathrm{P}(\tilde{\mathbf{X}}_k^{(i)} | \tilde{\mathbf{D}}_k^{(i)})} \mathrm{d}\tilde{\mathbf{X}}_k^{(i)}, \tag{9}$$

where $\mathcal{D}_B$ denotes the Bhattacharyya distance, a measure used to quantify the similarity between two probability distributions. The probability distribution $\mathrm{P}(\tilde{\mathbf{X}}_k | \tilde{\mathbf{D}}_k)$ is assumed to follow a Gaussian distribution $(\mathcal{N})$. This Gaussian distribution is characterized by a mean vector and a covariance matrix as $\mathcal{N}(\boldsymbol{\mu}_{\tilde{\mathbf{D}}_k}, \boldsymbol{\Sigma}_{\tilde{\mathbf{D}}_k})$.

Action Selection

The decision-making process of L hinges on its ability to decide between exploration and exploitation, depending on its interaction with the external environment. This discernment is predicated on the detection of observation anomalies. Specifically, L assesses its current state by analyzing the nature of its interactions. In scenarios where the observations align with familiar or normal patterns, L solely observes V1. Conversely, in instances characterized by novel or abnormal observations, L encounters a more complex situation involving multiple agents (e.g., two dynamic agents V1 and V2). This latter scenario represents a deviation from the experiences encapsulated in the expert demonstrations (see Figures 5 and 6). Consequently, based on this assessment, L opts for action a_k, which is informed by its interpretation of environmental awareness and the perceived need for exploration or exploitation, according to the following:

$$a_k = \begin{cases} \underset{a_k \in \mathcal{A}}{\operatorname{argmax}} \, \mathrm{P}(a_k | \tilde{\mathbf{D}}_k^{(\beta)}), & \text{if observation is normal then exploit.,} \\ \mathbf{q}\left(a_{k-1}, \boldsymbol{d}_k\right), & \text{if observation is abnormal then explore..} \end{cases} \tag{10}$$

In (10), under normal observation, L will imitate E's action selected from the active inference table $\boldsymbol{\Gamma}$, defined as the following:

$$\boldsymbol{\Gamma} = \begin{bmatrix} \mathrm{P}(a_1 | \tilde{\mathbf{D}}_1), & \mathrm{P}(a_2 | \tilde{\mathbf{D}}_1), & \ldots, & \mathrm{P}(a_m | \tilde{\mathbf{D}}_1) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{P}(a_1 | \tilde{\mathbf{D}}_m), & \mathrm{P}(a_2 | \tilde{\mathbf{D}}_m), & \ldots, & \mathrm{P}(a_m | \tilde{\mathbf{D}}_m) \end{bmatrix} \tag{11}$$

where $\sum_{i=1}^{m} \mathrm{P}(a_i | \tilde{\mathbf{D}}_k) = 1 \; \forall k$, $\mathrm{P}(a_i | \tilde{\mathbf{D}}_j) = \frac{1}{m}$ is the probability of selecting action $a_i \in \mathcal{A}$ conditioned to be in configuration $\tilde{\mathbf{D}}_j \in \boldsymbol{\mathcal{D}}$, and $\mathcal{A} = \{\dot{\boldsymbol{\mu}}_{\tilde{\mathbf{D}}_1}, \dot{\boldsymbol{\mu}}_{\tilde{\mathbf{D}}_2}, \ldots, \dot{\boldsymbol{\mu}}_{\tilde{\mathbf{D}}_m}\}$ is the set of available actions. In addition, β denotes the index of the particle with the maximum weight given by the following:

$$\beta = \underset{i}{\operatorname{argmax}} \, \{w_k^{(i)}\}_{i=1}^{N}. \tag{12}$$

In (10), if L encounters a situation that is abnormal and hasn't been seen by E before, then L will look for new ways to act. It does this by calculating the Euclidean distance (d_k), which is the shortest distance, between itself and V1 when they are in the same lane. Based on the measured distance, L adjusts its speed to ensure it doesn't exceed the speed of V1, helping to prevent a collision by slowing down or braking.

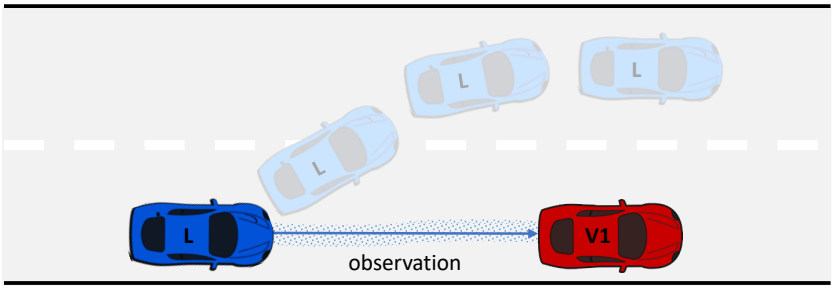


Figure 5. Learning agent receives a normal observation. It is allowed to overtake the vehicle directly ahead in its home lane.

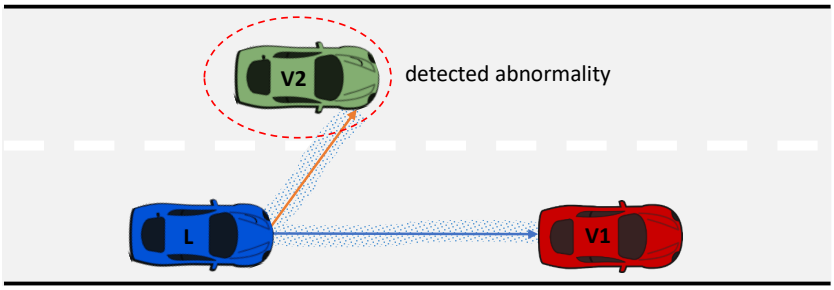


Figure 6. The learning agent detects an abnormal situation due to traffic in the adjacent lane, preventing it from overtaking the vehicle ahead.

Free Energy Measurements and GEs

The predictive messages, $\pi(\tilde{\mathbf{D}}_{\boldsymbol{k}})$ and $\pi(\tilde{\mathbf{X}}_{\boldsymbol{k}}^{(i)})$, propagated top-down through the hierarchy. At the same time, the AFP-M receives sensory responses in the form of diagnostic messages, $\lambda(\tilde{\mathbf{X}}_{\boldsymbol{k}}^{(i)})$ and $\lambda(\tilde{\mathbf{D}}_{\boldsymbol{k}})$, that move from the bottom level up the hierarchy. Calculating multi-level free energy (FE) helps to understand how well the current observations match what the model predicts.

At the discrete level, FE is measured as the distinction between two types of messages, $\pi(\tilde{\mathbf{D}}\boldsymbol{k})$ and $\lambda(\tilde{\mathbf{D}}\boldsymbol{k})$, as they enter the node $\tilde{D}k$. These messages are in the form of discrete probability distributions. Therefore, we propose using Kullback–Leibler Divergence ($\mathcal{D}KL$) [40] as a method to measure the probability distance and calculate the difference between these distributions.

$$\mathrm{Y}_{\tilde{\mathrm{D}}_k} = \mathcal{D}_{\mathcal{KL}}\big(\pi(\tilde{\mathbf{D}}_{\boldsymbol{k}}), \lambda(\tilde{\mathbf{D}}_{\boldsymbol{k}})\big) + \mathcal{D}_{\mathcal{KL}}\big(\lambda(\tilde{\mathbf{D}}_{\boldsymbol{k}}), \pi(\tilde{\mathbf{D}}_{\boldsymbol{k}})\big), \tag{13}$$

At the continuous level, FE is conceptualized as the distance between different probabilistic messages arriving at node $\tilde{X}k$. This involves the Bhattacharyya distance between the messages $\pi(\tilde{\mathbf{X}}\boldsymbol{k}^{(i)})$ and $\lambda(\tilde{\mathbf{X}}_{\boldsymbol{k}}^{(i)})$, originating from the observation level, that is defined as follows:

$$\mathrm{Y}_{\tilde{\mathrm{X}}_k^{(i)}} = -ln\left(\mathcal{BC}(\pi(\tilde{\mathbf{X}}_{\boldsymbol{k}}^{(i)}), \lambda(\tilde{\mathbf{X}}_{\boldsymbol{k}}^{(i)}))\right), \tag{14}$$

where $\mathcal{BC}$ is the Bhattacharyya coefficient.

Furthermore, generalized errors (GEs) facilitate understanding how to suppress such abnormalities in the future. The GE associated with (13) and conditioned upon transitioning from $\tilde{\mathbf{D}}_{\boldsymbol{k-1}}$ is defined as follows:

$$\tilde{\boldsymbol{\mathcal{E}}}_{\tilde{\mathrm{D}}_k} = [\tilde{\mathbf{D}}_{\boldsymbol{k}}, \mathrm{P}(\dot{\boldsymbol{\mathcal{E}}}_{\tilde{\mathrm{D}}_k})] = [\tilde{\mathbf{D}}_{\boldsymbol{k}}, \lambda(\tilde{\mathbf{D}}_{\boldsymbol{k}}) - \pi(\tilde{\mathbf{D}}_{\boldsymbol{k}})], \tag{15}$$

where $\dot{\boldsymbol{\mathcal{E}}}\tilde{D}k$ represents an aleatory variable characterized by a discrete probability density function (pdf), denoted as $\mathrm{P}(\dot{\boldsymbol{\mathcal{E}}}\tilde{D}k)$. The errors identified at the discrete level are then conveyed to the observation level. This process is essential for computing the generalized

error at this level, represented as $(\tilde{\boldsymbol{\mathcal{E}}}\tilde{Z}k)$, which explains the emergence of a new interaction within the surroundings.

Incremental Active Learning

By integrating the principles of adaptive learning and active inference, our objective is to minimize the occurrence of abnormalities. This goal can be achieved either by constructing a robust and reliable WM or by actively adapting to the dynamics of the environment. Such an approach ensures a comprehensive understanding of and interaction with environmental variables, thereby enhancing the system's predictive accuracy and responsiveness. The information acquired in the previous phases will be utilized to modify the beliefs of L and to incrementally expand its knowledge regarding the environment. This process involves updating the active inference table ($\boldsymbol{\Gamma}$) and expanding the transition matrix ($\boldsymbol{\Pi}$). These updates also will take into account the parameter of abnormality observation for considering the similarity between the two configurations.

In situations involving abnormalities, L incrementally encodes the novel experiences in WM by updating both the active inference matrix and the transition matrix. It's important to note that during such abnormal situations, L may encounter scenarios that involve configurations not previously experienced. These configurations are characterized by new relative distances between L and other dynamic objects in its environment, differing from those configurations previously known by the entity E. The discovery and understanding of these new configurations enable L to learn and adapt, thereby enhancing its ability to respond to similar situations in the future.

Consequently, a set $\boldsymbol{\mathcal{C}}$ consisting of the relative distance-action pair can be performed during the abnormal period T (i.e., exploration) as $\boldsymbol{\mathcal{C}} = \{\tilde{\mathbf{Z}}_k, a_k\}_t^T$ can be defined as $\boldsymbol{\mathcal{D}}^{'} = \{\mathbf{D}_{m+1}, \mathbf{D}_{m+2}, \ldots, \mathbf{D}_{m+n}\} = \{\mathbf{D}_1^{'}, \mathbf{D}_2^{'}, \ldots, \mathbf{D}_n^{'}\}$, where n is the total number of the newly acquired configurations and $\mathbf{D}_l^{'} \sim \mathcal{N}(\boldsymbol{\mu}_{\mathbf{D}_l^{'}}, \boldsymbol{\Sigma}_{\mathbf{D}_l^{'}})$ such that $\mathbf{D}_l^{'} \in \boldsymbol{\mathcal{D}}^{'}$. Therefore, the newly experienced action-configuration pairs characterized by $\mathrm{P}(a_k^{'}|\mathbf{D}_1^{'})$ are encoded in $\boldsymbol{\Gamma}^{'}$ according to the following:

$$\boldsymbol{\Gamma}^{'} = \begin{bmatrix} \mathrm{P}(a_1^{'}|\mathbf{D}_1^{'}), & \mathrm{P}(a_2^{'}|\mathbf{D}_1^{'}), & \ldots, & \mathrm{P}(a_n^{'}|\mathbf{D}_1^{'}) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{P}(a_1^{'}|\mathbf{D}_n^{'}), & \mathrm{P}(a_2^{'}|\mathbf{D}_n^{'}), & \ldots, & \mathrm{P}(a_n^{'}|\mathbf{D}_n^{'}) \end{bmatrix}, \tag{16}$$

Similarly, by analyzing the dynamic evolution of these new configurations, it becomes possible to estimate their transition probabilities $\mathrm{P}(\tilde{\mathbf{D}}_t|\tilde{\mathbf{D}}_{t-1})$ encoded in $\boldsymbol{\Pi}^{'}$, which is defined as follows:

$$\boldsymbol{\Pi}^{'} = \begin{bmatrix} \mathrm{P}(\mathbf{D}_1^{'}|\mathbf{D}_1^{'}), & \ldots, & \mathrm{P}(\mathbf{D}_1^{'}|\mathbf{D}_n^{'}) \\ \vdots & \ddots & \vdots \\ \mathrm{P}(\mathbf{D}_n^{'}|\mathbf{D}_1^{'}), & \ldots, & \mathrm{P}(\mathbf{D}_n^{'}|\mathbf{D}_n^{'}) \end{bmatrix}, \tag{17}$$

where $\sum_{k=1}^{m} \mathrm{P}(\mathbf{D}_i^{'}|\mathbf{D}_k^{'}) = 1 \; \forall i$. Consequently, the updated global transition matrix $\boldsymbol{\Pi}^{''} \in \mathbb{R}^{(m+n),(m+n)}$ is expressed as follows:

$$\boldsymbol{\Pi}^{''} = \begin{bmatrix} \boldsymbol{\Pi} & \mathbf{0}_{m,n} \\ \mathbf{0}_{n,m} & \boldsymbol{\Pi}^{'} \end{bmatrix}, \tag{18}$$

where $\boldsymbol{\Pi}$ is the original transition matrix and $\boldsymbol{\Pi}^{'}$ is the newly acquired one.

Action Update

L evaluates performed action at time $k-1$ using the FE calculated at time k, as defined in (13) and (14). In abnormal conditions, L learns future behaviors by gathering information about its surrounding environment.

During the online learning procedure, L modifies/updates the current active inference table and transition matrix, which is based on diagnostic messages, represented by $\lambda(\tilde{\mathbf{D}}\boldsymbol{k})$ and $\lambda(a_{k-1})$. Additionally, the transition matrix is refined using the GE defined in (15) as below:

$$\pi^*(\tilde{\mathbf{D}}_{\boldsymbol{k}}) = \pi(\tilde{\mathbf{D}}_{\boldsymbol{k}}) + \mathrm{P}(\dot{\boldsymbol{\mathcal{E}}}_{\tilde{\mathrm{D}}_k}). \tag{19}$$

The active inference table $\boldsymbol{\Gamma}$ can be adjusted according to the following:

$$\pi^*(a_k) = \pi(a_k) + \mathrm{P}(\dot{\boldsymbol{\mathcal{E}}}_{a_k}), \tag{20}$$

where, $\pi(a_k) = \mathrm{P}(\cdot|\tilde{\mathbf{D}}\boldsymbol{k})$ represents a specific row within $\boldsymbol{\Gamma}$. Furthermore, $\mathrm{P}(\dot{\boldsymbol{\mathcal{E}}}a_k)$ denotes the pdf of the GE associated with the active states, which can be calculated as follows:

$$\boldsymbol{\mathcal{E}}_{a_{k-1}} = [a_{k-1}, \mathrm{P}(\dot{\boldsymbol{\mathcal{E}}}_{a_{k-1}})] = [a_{k-1}, \lambda(a_{k-1}) - \pi(a_{k-1})], \tag{21}$$

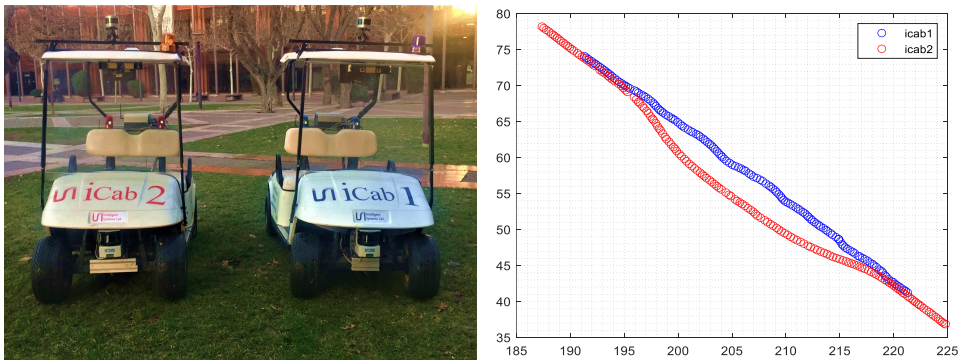
where $\lambda(a_{k-1}) = \lambda(\tilde{\mathbf{D}}_{\boldsymbol{k}}) \times \mathrm{P}(\tilde{\mathbf{D}}_{\boldsymbol{k}}|a_{k-1})$.

4. Results

In this section, we evaluate the proposed framework across different settings. First, we introduce the experimental dataset. Then, we describe the learning process, encompassing both the offline and online phases.

4.1. Experimental Dataset

The dataset used in this study was gathered from real experiments conducted on a university campus involving various maneuvering scenarios with two Intelligent Campus Automobile (iCab) vehicles [41]. During these experiments, expert demonstrations were recorded. These involved two AVs, iCab1 and iCab2, interacting to execute a specific driving maneuver: iCab2, acting as the expert vehicle (E), overtakes iCab1, which represents a dynamic object (V1), from the left side (see Figure 7). Each AV, denoted as (i), was equipped with both exteroceptive and proprioceptive sensors. These sensors collected data on odometry trajectories and control parameters to study the interactions between the AVs. The sensory data provided four-dimensional information, including the AVs' positions in (x, y) coordinates and their velocities $(\dot{x}, \dot{y})$.



(**a**) (**b**)

Figure 7. (**a**) Autonomous vehicles: iCab1 and iCab2. (**b**) An example of an overtaking scenario.

4.2. Offline Learning Phase

The expert demonstrations from the interactions between the two autonomous vehicles, iCab1 and iCab2, are utilized to learn the situation model (SM) offline. The learned SM comprises 24 joint clusters that encode the dynamic interaction between the AVs and the corresponding transition matrix, as depicted in Figure 8. Following this, the FP-M is initialized with these 24 learned configurations, which include position data and control parameters, as detailed in Section 3.2.1.

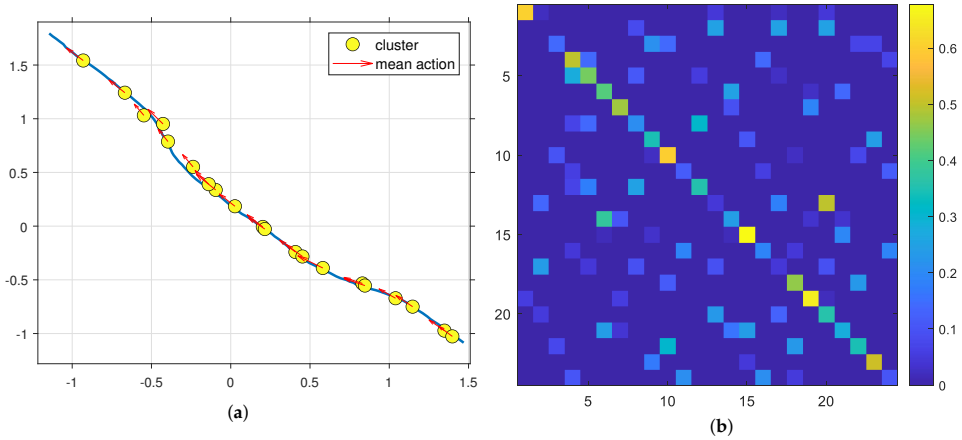


Figure 8. (**a**) Generated clusters and the associated mean actions to them, and (**b**) the generated transition matrix based on AVs' movements.

4.3. Online Learning Phase

The offline-acquired FP-M is enhanced with the action node to shape the AFP-M, as discussed in Section 3.2.2. This model enables L to operate in first-person during the online phase. As L undertakes a specific driving task, it assesses the situation based on its beliefs about the environmental state's evolution and actual observations. In normal situations, where observations align with predictions, L opts to perform expert-like maneuvers by imitating the expert's actions, known as exploitation. Conversely, in the face of abnormal conditions, such as encountering an unexpected vehicle, L begins to explore new actions based on these novel observations. This approach helps L avoid future abnormal situations and successfully achieve its goals. These goals are twofold: avoiding collisions with other vehicles in L's home lane when overtaking is not feasible due to traffic in the adjacent lane, and, when observations are normal, L safely overtaking the home-lane vehicle by following the expert's demonstration.

4.3.1. Action-Oriented Model

Figure 9 illustrates a trajectory performed by L in a normal situation, successfully overtaking the home-lane vehicle (V1) by following the prediction.

Conversely, Figure 10 shows how L alternates between exploration and exploitation during a driving task when faced with an abnormality, such as encountering a vehicle (V2) in the adjacent lane. In this scenario, L deviates from the predicted path to avoid a collision with V2 and slows down to maintain a safe distance from V1. This exploratory behavior enables L to learn new actions, like braking, in response to the proximity of the two vehicles. As a result, L expands its knowledge, adjusts its beliefs, and later applies these novel collected experiences in similar situations. Figure 10 further demonstrates that after learning new observations and actions related to reducing speed, L switches to exploitative behavior, using the newly acquired data to maintain a safe distance from V1 as long as it observes traffic in the adjacent lane.

Furthermore, Figure 11 depicts how, after navigating the abnormality and with available space in the adjacent lane (as V2 is moving faster), L once again has the opportunity to overtake V1. These results demonstrate the learned action-oriented model's effective adaptation to the dynamic variability in the environment.

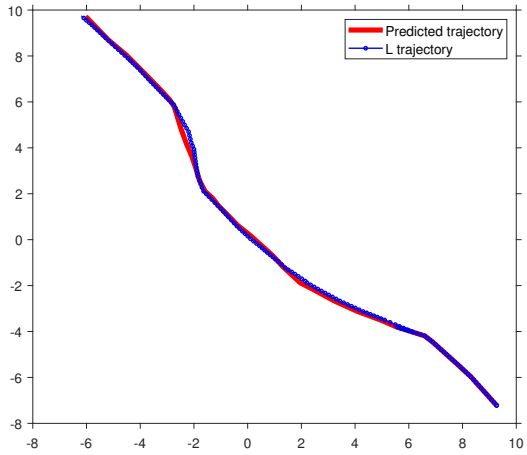


Figure 9. Thelearner observes a normal interaction, allowing it to follow its predictions.

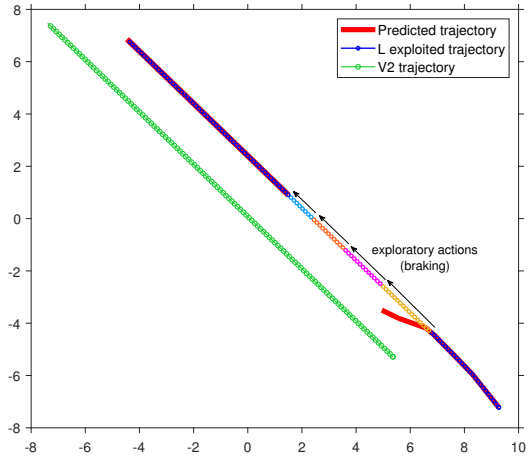


Figure 10. Confronted with an abnormality, the learner explores new actions in response to these unexpected observations.

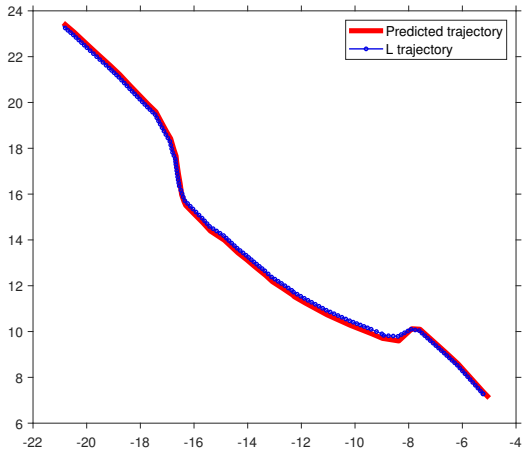


Figure 11. After overcoming the abnormality, the learner resumes imitating expert demonstrations.

4.3.2. Free Energy Measurement

In this section, we explore how updating and expanding the agent's beliefs about its surroundings minimizes the FE measurement. This minimization occurs through hierarchical processing, where prior expectations generate top-down predictions about likely observations. Any mismatches between these predictions and actual observations

are then passed up to higher levels as prediction errors. We examine the efficiency of two action-oriented models in terms of cumulative FE measurement:

- Model A, developed in a normal situation during the online learning phase, where L can overtake V1.
- Model B, formulated in an abnormal situation during the online learning phase, where L is temporarily unable to overtake V1 due to traffic in the adjacent lane.

Figures 12 and 13 offer a visual comparison of FE trends across training episodes. Figure 12 presents the results for Model A, where it was trained over 1000 episodes. In contrast, Figure 13 illustrates the performance of Model B over 2000 training episodes. This increase in training episodes for Model B is due to the more complex experimental scenario it addresses. The results demonstrate the capabilities of the proposed framework and its adaptability to new environments. Here, the autonomous vehicle (AV) continuously updates and improves its decision-making processes by refining its beliefs about the surroundings.

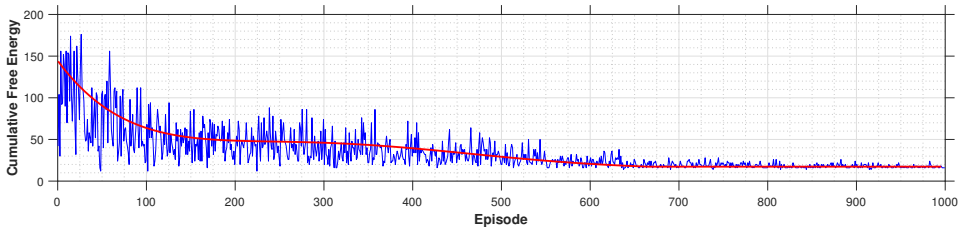


Figure 12. Cumulative free energy related to Model A. The red curve shows the measurement trend.

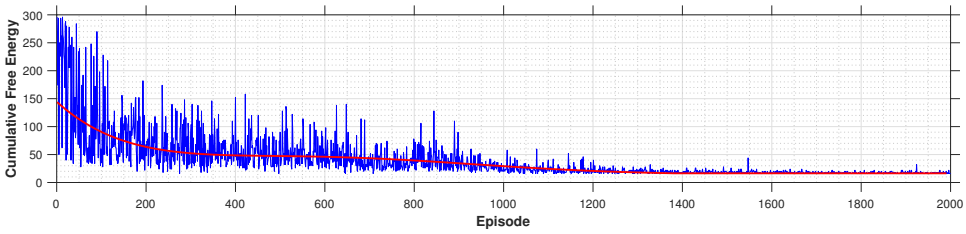


Figure 13. Cumulative free energy related to Model B. The red curve shows the measurement trend.

5. Conclusions

In this study, we present a comprehensive framework for autonomous vehicle navigation that integrates principles of active inference and incremental learning. Our approach addresses both normal and abnormal driving scenarios, providing a dynamic and adaptive solution for real-world autonomous driving challenges. The proposed framework effectively combines offline learning from expert demonstrations with online adaptive learning. This dual approach allows the autonomous agent to not only replicate expert maneuvers in normal conditions but also to develop novel strategies in response to unobserved environmental changes. Additionally, we introduced an action-oriented model that enables the agent to alternate between exploration and exploitation strategies. This adaptability is crucial in dynamic environments where the agent must constantly balance the need for safety with efficient navigation. In the proposed scenarios involving abnormalities, the learning agent demonstrated an ability to incrementally encode novel experiences and update its beliefs and transition matrix accordingly. This capability ensures that the intelligent agent continuously enhances its understanding of the environment and improves its decision-making process over time. Moreover, the results highlight the effectiveness of the framework in minimizing free energy, indicating a suitable alignment between the agent's predictions and environmental realities. This alignment is key to the agent's successful navigation and decision-making abilities. In conclusion, this research contributes significantly to the field of autonomous navigation by presenting an ability to learn from both pre-existing knowledge and ongoing environmental feedback, which is a promising solution for real-world autonomous navigation challenges.

While the current framework shows promising results, future work could focus on scaling the model for more complex environments, such as urban settings crowded with dynamic elements like pedestrians or varied traffic patterns.

Author Contributions: Conceptualization, S.N. and A.K.; methodology, S.N.; software, S.N.; validation, S.N.; formal analysis, S.N., A.K. and C.R.; investigation, S.N.; resources, S.N. and A.K.; data curation, S.N. and P.M.; writing—original draft preparation, S.N.; writing—review and editing, S.N. and A.K.; visualization, S.N. and A.K.; supervision, D.M.G. and C.R.; project administration, L.M.; funding acquisition, L.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AD	autonomous driving
ADS	autonomous driving systems
AFP-M	active first-person model
AV	autonomous vehicle
E	expert agent
FE	free energy
FP-M	first-person model
GDBN	generative dynamic Bayesian network
GE	generalized error
GM	generative model
GS	generalized state
IL	imitation learning
KF	Kalman filter
L	learning agent
MJPF	Markov jump particle filter
PF	particle filter
POMDP	partially observed Markov decision process
RL	reinforcement learning
SM	situation model
WM	world model

References

1. Bezai, N.E.; Medjdoub, B.; Al-Habaibeh, A.; Chalal, M.L.; Fadli, F. Future cities and autonomous vehicles: Analysis of the barriers to full adoption. *Energy Built Environ.* **2021**, *2*, 65–81. [CrossRef]
2. Parekh, D.; Poddar, N.; Rajpurkar, A.; Chahal, M.; Kumar, N.; Joshi, G.P.; Cho, W. A review on autonomous vehicles: Progress, methods and challenges. *Electronics* **2022**, *11*, 2162. [CrossRef]
3. Fürnkranz, J.; Gamberger, D.; Lavrač, N. *Foundations of Rule Learning*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2012.
4. Goldstein, M.H.; Waterfall, H.R.; Lotem, A.; Halpern, J.Y.; Schwade, J.A.; Onnis, L.; Edelman, S. General cognitive principles for learning structure in time and space. *Trends Cogn. Sci.* **2010**, *14*, 249–258. [CrossRef]
5. Knill, D.C.; Pouget, A. The Bayesian brain: The role of uncertainty in neural coding and computation. *Trends Neurosci.* **2004**, *27*, 712–719. [CrossRef]
6. Gao, X.; Zhang, Z.Y.; Duan, L.M. A quantum machine learning algorithm based on generative models. *Sci. Adv.* **2018**, *4*, eaat9004. [CrossRef]
7. Krayani, A.; Khan, K.; Marcenaro, L.; Marchese, M.; Regazzoni, C. A Goal-Directed Trajectory Planning Using Active Inference in UAV-Assisted Wireless Networks. *Sensors* **2023**, *23*, 6873. [CrossRef]
8. Friston, K.; FitzGerald, T.; Rigoli, F.; Schwartenbeck, P.; Pezzulo, G. Active inference: A process theory. *Neural Comput.* **2017**, *29*, 1–49. [CrossRef]
9. Friston, K. The free-energy principle: A unified brain theory? *Nat. Rev. Neurosci.* **2010**, *11*, 127–138. [CrossRef]

10. Regazzoni, C.S.; Marcenaro, L.; Campo, D.; Rinner, B. Multisensorial generative and descriptive self-awareness models for autonomous systems. *Proc. IEEE* **2020**, *108*, 987–1010. [CrossRef]
11. Ondruš, J.; Kolla, E.; Vertal', P.; Šarić, Ž. How do autonomous cars work? *Transp. Res. Procedia* **2020**, *44*, 226–233. [CrossRef]
12. Paden, B.; Čáp, M.; Yong, S.Z.; Yershov, D.; Frazzoli, E. A survey of motion planning and control techniques for self-driving urban vehicles. *IEEE Trans. Intell. Veh.* **2016**, *1*, 33–55. [CrossRef]
13. González, D.; Pérez, J.; Milanés, V.; Nashashibi, F. A review of motion planning techniques for automated vehicles. *IEEE Trans. Intell. Transp. Syst.* **2015**, *17*, 1135–1145. [CrossRef]
14. Pakdamanian, E.; Sheng, S.; Baee, S.; Heo, S.; Kraus, S.; Feng, L. Deeptake: Prediction of driver takeover behavior using multimodal data. In Proceedings of the CHI '21: CHI Conference on Human Factors in Computing Systems, Yokohama, Japan, 8–13 May 2021; pp. 1–14.
15. Wang, Y.; Liu, Z.; Zuo, Z.; Li, Z.; Wang, L.; Luo, X. Trajectory planning and safety assessment of autonomous vehicles based on motion prediction and model predictive control. *IEEE Trans. Veh. Technol.* **2019**, *68*, 8546–8556. [CrossRef]
16. Atkeson, C.G.; Schaal, S. Robot learning from demonstration. *Proc. ICML* **1997**, *97*, 12–20.
17. Schaal, S. Is imitation learning the route to humanoid robots? *Trends Cogn. Sci.* **1999**, *3*, 233–242. [CrossRef] [PubMed]
18. Billard, A.; Calinon, S.; Dillmann, R.; Schaal, S. Robot programming by demonstration. In *Springer Handbook of Robotics*; Springer: Berlin/Heidelberg, Germany, 2008; pp. 1371–1394.
19. Raza, S.; Haider, S.; Williams, M.A. Teaching coordinated strategies to soccer robots via imitation. In Proceedings of the 2012 IEEE International Conference on Robotics and Biomimetics (ROBIO), Guangzhou, China, 11–14 December 2012; pp. 1434–1439.
20. Hussein, A.; Gaber, M.M.; Elyan, E.; Jayne, C. Imitation learning: A survey of learning methods. *Acm Comput. Surv. (Csur)* **2017**, *50*, 1–35. [CrossRef]
21. Onishi, T.; Motoyoshi, T.; Suga, Y.; Mori, H.; Ogata, T. End-to-end learning method for self-driving cars with trajectory recovery using a path-following function. In Proceedings of the 2019 International Joint Conference on Neural Networks (IJCNN), Budapest, Hungary, 14–19 July 2019; pp. 1–8.
22. Chen, Z.; Huang, X. End-to-end learning for lane keeping of self-driving cars. In Proceedings of the 2017 IEEE Intelligent Vehicles Symposium (IV), Los Angeles, CA, USA, 11–14 June 2017; pp. 1856–1860.
23. Bojarski, M.; Del Testa, D.; Dworakowski, D.; Firner, B.; Flepp, B.; Goyal, P.; Jackel, L.D.; Monfort, M.; Muller, U.; Zhang, J.; et al. End to end learning for self-driving cars. *arXiv* **2016**, arXiv:1604.07316.
24. Sauer, A.; Savinov, N.; Geiger, A. Conditional affordance learning for driving in urban environments. *Proc. Conf. Robot. Learn.* **2018**, *87*, 237–252.
25. Vogt, D.; Ben Amor, H.; Berger, E.; Jung, B. Learning two-person interaction models for responsive synthetic humanoids. *J. Virtual Real. Broadcast.* **2014**, *11*. [CrossRef]
26. Droniou, A.; Ivaldi, S.; Sigaud, O. Learning a repertoire of actions with deep neural networks. In Proceedings of the 4th International Conference on Development and Learning and on Epigenetic Robotics, Genoa, Italy, 13–16 October 2014; pp. 229–234.
27. Liu, M.; Buntine, W.; Haffari, G. Learning how to actively learn: A deep imitation learning approach. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, Melbourne, Australia, 15–20 July 2018; pp. 1874–1883.
28. Argall, B.D.; Chernova, S.; Veloso, M.; Browning, B. A survey of robot learning from demonstration. *Robot. Auton. Syst.* **2009**, *57*, 469–483. [CrossRef]
29. Ross, S.; Bagnell, D. Efficient reductions for imitation learning. *Proc. Conf. Robot. Learn.* **2010**, *9*, 661–668.
30. Gangwani, T.; Peng, J. State-only imitation with transition dynamics mismatch. *arXiv* **2020**, arXiv:2002.11879.
31. Ogishima, R.; Karino, I.; Kuniyoshi, Y. Combining imitation and reinforcement learning with free energy principle. In Proceedings of the Ninth International Conference on Learning Representations (ICLR 2021), Virtual, 3–7 May 2021.
32. Kuefler, A.; Morton, J.; Wheeler, T.; Kochenderfer, M. Imitating driver behavior with generative adversarial networks. In Proceedings of the 2017 IEEE Intelligent Vehicles Symposium (IV), Los Angeles, CA, USA, 11–14 June 2017; pp. 204–211.
33. Schroecker, Y.; Vecerik, M.; Scholz, J. Generative predecessor models for sample-efficient imitation learning. *arXiv* **2019**, arXiv:1904.01139.
34. Yang, Q.; Gu, Y.; Wu, D. Survey of incremental learning. In Proceedings of the 2019 Chinese Control and Decision Conference (CCDC), Nanchang, China, 3–5 June 2019; pp. 399–404.
35. Chalup, S.K. Incremental learning in biological and machine learning systems. *Int. J. Neural Syst.* **2002**, *12*, 447–465. [CrossRef]
36. Kim, S.; Laschi, C.; Trimmer, B. Soft robotics: A bioinspired evolution in robotics. *Trends Biotechnol.* **2013**, *31*, 287–294. [CrossRef] [PubMed]
37. Nozari, S.; Krayani, A.; Marin, P.; Marcenaro, L.; Martin, D.; Regazzoni, C. Adapting Exploratory Behaviour in Active Inference for Autonomous Driving. In Proceedings of the ICASSP 2023–2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Rhodes Island, Greece, 4–10 June 2023; pp. 1–5.
38. Nozari, S.; Krayani, A.; Marin-Plaza, P.; Marcenaro, L.; Gomez, D.M.; Regazzoni, C. Active Inference Integrated with Imitation Learning for Autonomous Driving. *IEEE Access* **2022**, *10*, 49738–49756. [CrossRef]
39. Krayani, A.; Alam, A.S.; Marcenaro, L.; Nallanathan, A.; Regazzoni, C. Automatic Jamming Signal Classification in Cognitive UAV Radios. *IEEE Trans. Veh. Technol.* **2022**, *71*, 12972–12988. [CrossRef]

40. Pardo, L. *Statistical Inference Based on Divergence Measures*; CRC Press: Boca Raton, FL, USA, 2018.
41. Marín-Plaza, P.; Beltrán, J.; Hussein, A.; Musleh, B.; Martín, D.; de la Escalera, A.; Armingol, J.M. Stereo Vision-based Local Occupancy Grid Map for Autonomous Navigation in ROS. In Proceedings of the 11th International Joint Conference, VISIGRAPP 2016, Rome, Italy, 27–29 February 2016; pp. 701–706. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.