IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 24 January 2024, accepted 5 March 2024, date of publication 14 March 2024, date of current version 20 March 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3376253

RESEARCH ARTICLE

Efficient Motion Planning With Minimax Objectives: Synergizing Interval Prediction and Tree-Based Planning

CONG PHAT VO[1], PHILJOON JUNG[2], TAE-HYUN KIM[2], AND JEONG HWAN JEON[1,3], (Member, IEEE)

[1]Department of Electrical Engineering, Ulsan National Institute of Science and Technology (UNIST), Ulsan 44919, Republic of Korea
[2]Pangyo Research and Development Center, Hanwha Systems Company Ltd., Seongnam 13524, Republic of Korea
[3]Graduate School of Artificial Intelligence, Ulsan National Institute of Science and Technology (UNIST), Ulsan 44919, Republic of Korea

Corresponding author: Jeong hwan Jeon (jhjeon@unist.ac.kr)

This work was supported by Korea Research Institute for defense Technology planning and advancement (KRIT) grant funded by the Korea government (DAPA (Defense Acquisition Program Administration)) (No. KRIT-CT-21-009, Development of Realtime Automatic Mission Execution and Correction Technology based on Battlefield Information, 2021).

ABSTRACT This paper presents an efficient motion planning framework for a perturbed linear system using a minimax objective function while ensuring the safety of the system. Specifically, the proposed approach is naturally deployed to handle model uncertainties by a recursive least squares-based set-membership mechanism. Next, a minimax-based objective optimization problem is formed to handle the goal flexibility. The robust model predictive control algorithm is then designed to solve this robust optimization objective. Furthermore, a refined strategy is able to approximate robust objectives by synergizing interval prediction and tree-based planning to achieve the best surrogate performance. It is extended to incorporate a hierarchical control architecture in a specific context. This extension serves to enhance path efficiency and, in turn, alleviates the constraints associated with modeling assumptions. The primary difficulty involves integrating and adjusting theoretical assurances at each level, a task accomplished through a comprehensive examination of suboptimality from end to end. The proposed framework is versatile across a variety of models, incorporating a solid, data-informed approach for selecting models. This integration permits a more flexible approach to modeling assumptions. Moreover, we consistently maintain the practicability of our method throughout its application, a fact that is evidenced by its successful deployment in complex simulated settings.

INDEX TERMS Motion planning, model predictive control, tree-based planning, interval prediction.

I. INTRODUCTION

Deploying robotic vehicles that operate fully autonomously in inhabited areas, especially in apartment complexes, is a great potential step toward building a smart city [1], [2]. During operation, the presence of dynamic obstacles, such as pedestrians, other robots, and vehicles, within the environment, adds to the complexity as it constantly evolves over time, making the task of discovering efficient, collision-free paths increasingly demanding [3], [4], [5]. Thus, one of the most fundamental criteria for these robots is motion safety against a high level of unpredictability and uncertainty. The rapid growth of this research topic has resulted from high real-world demand. In this context, tackling issues involving uncertain systems has always been a major challenge, with a considerable volume of literature dedicated to addressing each undertaking. Previous research predominantly focuses on stabilizing these systems with respect to a predetermined reference state or trajectory, such as sliding-mode control [6], [7], [8], $\mathcal{H}_\infty$ control [9], system-level synthesis control [10], [11] and Model Predictive Control (MPC) [12], [13], [14], [15]. Within the MPC framework, one of the currently popular strategies has the unique ability to both track and predict trajectories.

The associate editor coordinating the review of this manuscript and approving it for publication was Zheng Chen.

© 2024 The Authors. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.
For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

Its features have been developed to tackle model uncertainty, although they often lack robust guarantees. An upgraded variant called the tube-MPC algorithm emerges as an attempt to provide theoretical assurances of robust constraint satisfaction [16], [17], [18]. The objective is to maintain the system state within a secure region centered around the origin, typically formulated as a convex problem. Introducing greater flexibility, an adaptive MPC [19], [20], [21] is devised, incorporating the fulfillment of the Persistence of Excitation (PE) condition [22]. This approach aims to ensure the controlled system exhibits properties of robust stability and recursive feasibility. Nevertheless, many practical tasks do not simply involve stability issues but also pay attention to safety, particularly those involving obstacle avoidance. Such traditional obstacle avoidance techniques as the Dynamic Window Approach (DWA) [23], [24], [25] and the Timed Elastic Band (TEB) [26], [27] are primarily designed for static obstacle avoidance. Consequently, maneuvering around dynamic obstacles poses a unique challenge, requiring the necessary information about obstacles through object detection solutions to adapt effectively.

Fortunately, the MPC algorithms have continued to be a highly effective control technique, maintaining their efficacy by accommodating complex dynamical systems and robust control objectives. Several methodologies have been investigated, including the Differential Evolution Algorithm (DEA) [28], [29], Sand Cat Swarm Optimization (SCSO) [30], Grey Wolf Optimizer (GWO) [31] and Non-dominated Sorting Genetic Algorithm III (NSGAIII) [32] for seeking objective optimums in complex landscapes. However, these methods are more suited for problems where a global or multi-objective optimum exists; they might not inherently focus on the worst-case scenarios. Herein, the duties are described as risk management in uncertain or adversarial conditions. Thus, minimax is optimal for scenarios where safeguarding against the worst-case outcome is crucial [33], [34], [35], [36], [37]. It is notably suitable to offer greater flexibility in defining goals. Reachability analysis of the minimax-based problems allows selecting the best possible dynamics while adhering to controllability constraints. It yields the ability of a cumulative regret minimization within the Linear Quadratic (LQ) problem to achieve approximately a certain level [38]. Randomized candidate dynamics selection [39], [40] and noise injection [41] also achieve similar results. Moreover, the minimax control strategy can be extended to robust versions by Markov Decision Processes in finite time [42], [43]. Yet, the minimax objective function has contained several intrinsic limitations, such as the set of possible dynamics ambiguity that has to satisfy specific properties [44]. This is not directly applicable when making sequential decisions with continuous states. Meanwhile, random exploration is not feasible in critical cases where safety must be guaranteed. The proposed approach not only considers this issue but also solves restrictions on the cost function form that yield difficulty in calculating to reach the optimal value explicitly.

This research paper focuses on overcoming the challenges of planning and controlling the perturbed mobility system while maximizing a bounded cost function, especially in situations where errors must be avoided. In detail, following an offline estimation phase based on previous work [20], an interval predictor is employed to generate oversets and find a near-optimal control through an improved tree-based planning procedure [45]. Furthermore, the evaluation of suboptimality between a minimax objective-based control strategy and optimal performance involves holding forth to generic cost solutions. It deals with complex functions common to practical problems involving combinatorial optimization and branching decisions. The main contributions are outlined as follows:

- The recursive least-squares-based set-membership estimation mechanism provides a sequence of bounding sets of the unknown model parameters as a premise to formulate the robust objective function.
- We exploit the ability of the interval predictors to design the improved tree-based planning approach that is able to approximate robust objectives with a tractable surrogate via defining a generic cost.
- The proposed framework is theoretically analyzed to indicate an upper bound for the approximation bias and suboptimality. It guarantees the best surrogate performance achieved during planning.

The remaining sections of the paper are structured as the following. The dynamic system and objectives are described in Section II. Section III presents the complete integration of the estimation, prediction, and control. Section IV establishes the simulation scenarios and provides a discussion of the results. Finally, Section V summarizes the conclusions drawn from the study.

II. PROBLEM FORMULATION

This work investigates a perturbed linear system featuring unknown parameters in the time-domain representation through the following formula:

$$\begin{cases} \dot{x}_t(t) = A(\xi)x_t(t) + Bu(t) + d(t), \\ y_t(t) = Cx_t(t) + \gamma(t) \end{cases} \tag{1}$$

where $x_t(t) \in \mathbb{R}^p$, $u(t) \in \mathbb{R}^q$, $d(t) \in \mathbb{R}^p$, $y(t) \in \mathbb{R}^n$ and $\gamma(t) \in \mathbb{R}^n$ denote the system state, the control input, the disturbances, the measurement output and noise, respectively; the state matrix $A(\xi) \in \mathbb{R}^{p\times p}$ depending on the uncertain parameter ξ; the control matrix $B \in \mathbb{R}^{p\times q}$ and $C \in \mathbb{R}^{n\times p}$ are known.

The objective of this work is to develop a robust planning and control framework for the system (1) that delivers great decisions in critical cases (robust) while maintaining flexibility in performing tasks (adaptive). Exploiting the results of the interval prediction technique supported by the online uncertain parameter estimation, the obstacle avoidance tasks are formulated with the robust objective and a tractable surrogate by defining a generic cost. To start with, we define

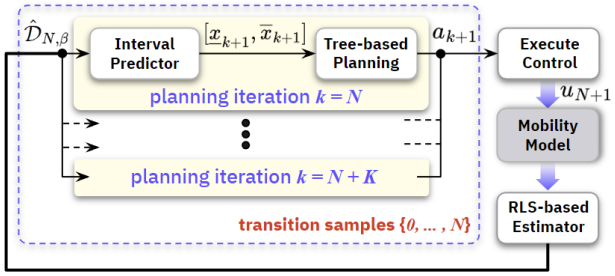

FIGURE 1. Illustration of the proposed decision-making framework and its relationship with other system components.

that the system (1) is subject to the constraint sets of the state $x_t(t) \in \mathcal{S} \subset \mathbb{R}^p$, the control $u(t) \in \mathcal{A} \subset \mathbb{R}^q$ and the uncertainty set $\mathcal{D} \subset \mathbb{R}^d$ with a priori knowledge. Similar to [39], to procure an appropriate set of all admissible values for $\xi \in \mathcal{D}$, the form of $A(\xi)$ is supposed as the following.

Assumption 1: Suppose that the matrices $A \in \mathbb{R}^{p\times p}$, $\rho_i \in \mathbb{R}^{d\times p\times p}$ are known such that $\xi_i \in \mathcal{D}, \forall i \in d$, and the dynamic matrix structure can be formed as follows:

$$A(\xi) = A + \sum_{i=1}^{d} \xi_i \rho_i. \tag{2}$$

Assumption 2: In the system (1), there exist admissible perturbation bounds $\underline{d}, \overline{d} \in \Gamma^p$, $\underline{\gamma}, \overline{\gamma} \in \Gamma^n$ such that $d \in [\underline{d}(t), \overline{d}(t)], \gamma \in [\underline{\gamma}(t), \overline{\gamma}(t)], \forall t \geq 0$, both containing the origin in their interior. In addition, $\underline{x}_0, \overline{x}_0 \in \Gamma^p$ such that the initial conditions $x_0 \in [\underline{x}_0, \overline{x}_0] \subset \mathcal{S}$. The initial uncertainty set is defined by $\mathcal{D}_0 = \left\{\xi \in \mathbb{R}^d, |\xi - \xi_0| \leq \overline{\xi}\right\}$, where an initial ξ_0 and $\overline{\xi}$ are positive values.

Realistically, most practical works require high robustness and motion safety, thus the knowledge of physical models and well-designed controllers are still given the top priority. It implies that the proposed strategy can still ensure that expectations are met but also allows for uncertain terms around the nominal model, $\mathcal{E}(x_t(t), u(t), \xi) = [\rho_1 x_t^n(t), \dots, \rho_N x_t^n(t)] \in \mathbb{R}^{p\times N}$, which are known signals that linearly depend on ξ, at step $n \in \mathcal{N}$ with $\mathcal{N} = \{1, 2, \dots, N+1\}$ and a given number of transition samples $N \in \mathbb{N}$. Herein, we suppose that a linear regression system $Y = \mathcal{E}(t)\xi^* + \vartheta(t)$ based on the hypothesis in [46] instead of the measurement output, in which the lumped uncertainty $\vartheta(t)$ is the sum of disturbance $d(t)$ and the measurement noise $\gamma(t)$, where ξ^* is the true values of ξ. According to Assumption 2, $\vartheta(t)$ can be bounded by the specified positive value $\overline{\vartheta} = \max\left\{||\underline{d}||, ||\overline{d}||\right\} + \max\left\{||\underline{\gamma}||, ||\overline{\gamma}||\right\}$.

III. THE PROPOSED CONTROL APPROACH

In this section, the proposed strategy is delineated as the following. Initially, an online approach is utilized to concurrently determine the unknown parameters and the uncertainty set. Subsequently, the interval prediction block is formulated, leveraging inclusion features to prefigure information with regard to the presently observed states, which is illustrated in Figure 1.

A. STATE ESTIMATION AND PREDICTION

In order to premise the robust planning control framework design, the estimation $\hat{\xi}$ of unknown system parameters ξ can be calculated by [20] and [47] as follows:

$$\dot{\hat{\xi}}(t) = \hat{\xi}(t) + \dot{\kappa}(t)\dot{\breve{\sigma}}(t)\mathcal{E}(t)\dot{\tilde{\vartheta}}(t) \tag{3}$$

where $\hat{\xi}(t) \in \mathbb{R}^N, \breve{\sigma}(0) = 1$ and the equalities at time instant t are indicated as the following, respectively

$$\dot{\tilde{\vartheta}}(t) = Y(t) - \mathcal{E}(t)\dot{\hat{\xi}}(t) \tag{4a}$$

$$\dot{\breve{\sigma}}(t) = \frac{\breve{\sigma}}{1 + \dot{\kappa}(t)\breve{\sigma}\mathcal{E}(t)^2} \tag{4b}$$

$$\dot{\kappa}(t) = \begin{cases} 0, & \text{if } \left|\dot{\tilde{\vartheta}}\right| \leq \overline{\vartheta}, \\ \frac{\left|\dot{\tilde{\vartheta}}(t)\right| - \xi}{\overline{\vartheta}\mathcal{E}(t)^2\dot{\breve{\sigma}}(t)} & \text{otherwise} \end{cases} \tag{4c}$$

Subsequently, our purpose is to design the appropriate set of all admissible values $\hat{\mathcal{D}}(t)$ such that carries the faithful parameters with an eminent probability $\mathbb{P}\left(\xi \in \hat{\mathcal{D}}\right)$ at least $1 - \beta, \forall\beta \in [0, 1)$. According to (3), the set $\hat{\mathcal{D}}(t)$ can be given as follows:

$$\hat{\mathcal{D}}(t) = \mathcal{D} \bigcap_{\iota\in[\varrho,t]} \left\{\xi \in \mathbb{R}^N : |\xi - \hat{\xi}(\iota)| \leq \widetilde{\delta}(\iota)\right\} \tag{5}$$

where $\widetilde{\delta}(\iota) = \breve{\sigma}^{\frac{1}{2}}\varpi(\iota)$ is a scalar with $\varpi(0) = \overline{\xi}$ and the expression $\varpi(t)$:

$$\dot{\varpi}(t)^2 = \varpi^2(t) + \dot{\kappa}\left(\overline{\vartheta}^2 - \frac{\widetilde{\delta}^2}{1 + \dot{\kappa}\mathcal{E}^2\breve{\sigma}}\right) \tag{6}$$

Remark 1: According to (3) and (5), it is obvious that the property $\hat{\mathcal{D}}(t) \subseteq \mathcal{D}(t), t \geq 0$ is satisfied with its size shrinking. Owing to $\xi^* \in \mathcal{D}(0)$, we can give the conclusion that $\xi^* \in \mathcal{D}(t), \forall t \leq 0$. Meanwhile, the estimation error $||\xi^* - \hat{\xi}(t)|| \leq \widetilde{\delta}(t), \forall\xi^* \in \mathcal{D}_0$ is guaranteed to be bounded and nonincreasing when Assumption 2 is fulfilled.

Furthermore, to leverage the benefits of the promising results in [20], the interval prediction approach is described to obtain $[\underline{x}(t), \overline{x}(t)]$ with the inclusion property:

$$\underline{x}(t) \leq x_t(t) \leq \overline{x}(t), \forall t \geq t_N. \tag{7}$$

The above property (7) yields $\underline{A} \leq A(\xi) \leq \overline{A}$ from the set $\hat{\mathcal{D}}_{N,\beta}$; and the information on the current state x_N and the admissible disturbance bounds $[\underline{d}(t), \overline{d}(t)]$ provide the information on this predictor with N samples. From this hypothesis, the polytopic structure is deployed to produce more stable predictions based on matrix interval arithmetic to derive the predictor. It is observed that $\hat{\mathcal{D}}(t)$ can be enclosed on ξ obtained from (8) into a polytope for $A(\xi)$, which can be expressed as follows:

$$A(\xi) = A_N + \sum_{i=1}^{2^d} \mathcal{W}_i \mathcal{X}_{N,i} \tag{8}$$

where $A_N = A(\hat{\xi}(t))$, $\mathcal{W}_i \geq 0$, $\sum_{i=1}^{2^d} \mathcal{W}_i = 1$, $\mathcal{X}_{N,i} = A(\hat{\xi} + \widetilde{\kappa}_i\widetilde{\delta}) - A_N$ for $\widetilde{\kappa}_i \in \{-1, 1\}^d$ with $i \in [1, 2^d]$. Besides, A_N is Metzler to inherit beneficial features of nonnegative systems.

Assumption 3: Suppose that an orthogonal matrix $H \in \mathbb{R}^{p\times p}$ has existence with the non-diagonal elements of $H^\top A_N H$ are all non-negative.

Remark 2: According to [46], the capability of (8) and the inclusion property (7) are verified and Assumption 3 is supposed for the system (1), then the interval predictor is given as follows:

$$\begin{cases} \dot{\underline{x}}(t) = A_N\underline{x}(t) - \mathcal{X}^+\underline{x}^-(t) & -\mathcal{X}^-\overline{x}^+(t) \\ & +Bu(t) + \underline{d}(t) - \overline{d}(t), \\ \dot{\overline{x}}(t) = A_N\overline{x}(t) - \mathcal{X}^+\overline{x}^-(t) & -\mathcal{X}^-\underline{x}^+(t) \\ & +Bu(t) + \overline{d}(t) - \underline{d}(t) \end{cases} \tag{9}$$

where the polynomials $\mathcal{X}^+ = \sum_{i=1}^{2^N} \mathcal{X}_i^+$, $\mathcal{X}^- = \sum_{i=1}^{2^N} \mathcal{X}_i^-$.

B. ROBUST PLANNING AND CONTROL

After briefly presenting the antecedent strategies of state estimation and prediction, the proposition of a robustly stabilizing system (1) near the origin is introduced in [20], taking into account parametric uncertainty and restricted disturbances. It also ensures that $[\underline{x}_0, \overline{x}_0]$ falls within $\mathcal{S}$, establishing defined boundaries for the state. Then, the robust MPC approach reaches significant effectiveness by utilizing a control mechanism that stabilizes within the predictive intervals. However, the conservativeness of the robust MPC algorithms can be challenged due to the goal flexibility in the practical application. Therefore, the minimax control objective can be a promising candidate for those duties. Then, the process of planning the path is incorporated to handle a sequence of states or decisions that the system up to time to reach a desired goal. Hereby, the tree-based planning algorithm based on [45] is improved to face a sequential decision problem in a confidence set with continuous states via the generic costs J. In addressing the optimal and robust control objective, the hierarchical control mechanism is applied to execute a first approximation and discretization of the continuous decision space $(\mathbb{R}^q)^{\mathbb{N}}$. This approach allows for the selection of a high-level action a at each time step, where each action $a \in \mathcal{A}$ aligns with the choice of a low-level controller, respectively. Although suboptimality is introduced as a consequence of discretization, this may be alleviated by constructing a manifold set of basic controllers.

Inspired by this aim, the supremum of the robust objective $\mathcal{J}$ is taken to achieve the maximum value of the worst possible results under $\hat{\mathcal{D}}_{N,\beta}$, the objective is defined as the following.

$$\mathcal{J}(\mathbf{u}) = \inf_{\substack{\xi\in\hat{\mathcal{D}}_{N,\beta} \\ d\in[\underline{d},\overline{d}]^{\mathbb{N}}}} \left[\sum_{n=N+1}^{\infty} \eta^n J(x_t^n(\mathbf{u}, d))\right] \tag{10}$$

where J refers to an arbitrary bounded cost function, $\eta \in (0, 1]$ represents a discount factor, $x_t^n(\mathbf{u}, d)$ denotes the state reached at step n under the planned control sequence $\mathbf{u} \in (\mathbb{R}^q)^{\mathbb{N}}$ and disturbances d within the specified admissible bounds $[\underline{d}, \overline{d}]$.

Thanks to the capability of the interval prediction approach in (9), the idea is to seek the lower bound value of the robust objective $\mathcal{J}$ in (10). Thus, a surrogate objective is proposed to approximate the robust objective by finding the minimum value of $J(x_t)$, called the pessimistic cost $\underline{J}_n$. Note that this replacement renders the cost a non-Markov property, as it involves evaluating the worst possible across the entire past states.

Theorem 1: The surrogate objective plays the role of a lower bound of the robust objective (10), which is defined as follows:

$$\hat{\mathcal{J}}(\mathbf{u}) \overset{def}{=} \sum_{n=N+1}^{\infty} \eta^n \underline{J}_n(\mathbf{u}) \leq \mathcal{J}(\mathbf{u}) \tag{11}$$

where $\underline{J}_n(\mathbf{u}) = \min J(x_t)$, $\forall x_t \in [\underline{x}_n(\mathbf{u}), \overline{x}_n(\mathbf{u})]$.

Proof: See Appendix 1. ■

Remark 3: Conservative approximations involve intentionally overestimating risks in models for added safety. Control sequences are designed based on a surrogate objective $\hat{\mathcal{J}}$ to prevent undesirable events, such as collisions. Ensuring compliance with this feature provides confidence in the effectiveness of the control sequence in the real system, improving overall safety and reliability.

As the proposal above, the observations with respect to time N including both predictor dynamics and objective costs, are deterministic. This means that the manageable objective function includes all the randomness associated with disturbances and measurements. Therefore, an optimistic planning (OP) algorithm, as described in [45], is modified for suitability without the requirement for Markovity or state enumeration. It can be optimized as a sequence of actions by considering the corresponding costs. Specifically, it generates the look-ahead tree $\mathcal{T}$ by expanding progressively the leaf a_k, i.e., the value of the action sequence a at each iteration $k \in \mathcal{K}$ with $\mathcal{K} = \{1, 2, \ldots, K\}$ and the number of planning iterations $K \in \mathbb{N}$. These sequences a_k can be bounded by finding an argument of the maximum, which is formed as follows:

$$a_k = \arg\max_{a\in\mathcal{U}_k} \sum_{n=0}^{l_a-1} \underline{J}_n(a) + \frac{\eta^{l_a}}{1-\eta} \tag{12}$$

where $\mathcal{U}_k$ denotes the set of the leaves of tree $\mathcal{T}_k$ and l_a denotes the length of the sequence action a.

Lemma 1: Leveraging the suboptimality inherent in the OP algorithm (12) from [45] for the surrogate objective (11) after K iterations, we have

$$\hat{\mathcal{J}}(a^\star) - \hat{\mathcal{J}}(a_K) \leq \mathcal{P}\left(K^{-\frac{\log 1/\eta}{\log \nu}}\right) \tag{13}$$

where a problem-dependent variable ν is defined to gauge the extent to which paths closely approximate optimality, $\limsup_{l_a\to\infty} \left|\left\{a \in A^{l_a} : \hat{\mathcal{J}}(a^\star) - \hat{\mathcal{J}}(a) \leq \frac{\eta^{l_a+a+1}}{1-\eta}\right\}\right|^{1/l_a}$ *and*

$\mathcal{P}(\bullet)$ is a positive function, which can be explained as $h(s) = \mathcal{P}(g(s))$, there exists $s_0, \varsigma > 0$ such that $h(s) \leq \varsigma g(s)$, $\forall s > s_0$.

Hereby, the optimal surrogate value $\hat{\mathcal{J}}(a^\star)$ can be gradually obtained at a polynomial rate when bringing into play enough computational planning budget K. Nevertheless, between $\hat{\mathcal{J}}$ and the true robust objective $\mathcal{J}$, there always exists a bias. First, it originates from the approximation error of the truly accessible set of an enclosing interval (7). Second, the predictive error of the interval predictor (9) is affected by the difference of the time-invariance $A(\xi) \in \hat{\mathcal{D}}_{N,\beta}$ and a time-varying $A(\xi(t)) \in \hat{\mathcal{D}}_{N,\beta}$ of the dynamics uncertainty. Last, it is a little loose when verifying the lower bound of the surrogate objective (11). Yet, this problem can be solved under additional assumptions.

Assumption 4: Suppose that a Lipschitz regularity for the cost J is satisfied and there exists $\rho > 0$, $P > 0$, $Q_0 \in \mathbb{R}^{p\times p}$ and $N_0 \in \mathbb{N}$ such that $\begin{bmatrix} A(\xi_N)^\top P + PA(\xi_N) + Q_0 & P \\ {}^\top P & -\rho I_r \end{bmatrix} < 0, \quad \forall N > N_0$.

It can be seen that Assumption 4 pertains to matrices A_N, which are formed indirectly from the algorithm. Then, examining the validity becomes complicated. A more robust yet simpler-to-verify criterion is that a set of the polytope (8) at a certain iteration, where this stability property holds uniformly. This suggests that the features are appropriately activated when the estimation reaches the vicinity of the true dynamics $A(\xi)$. Additionally, this approach allows for the imposition of extra constraints on the estimation error, which is dependent on the input.

Theorem 2: Under an additional PE condition, there exists positive values $\underline{\rho}, \overline{\rho}$ such that $\underline{\rho}^2 \leq \lambda_{\min}(\mathcal{E}_n^\top \Gamma^{-1} \mathcal{E}_n) \leq \overline{\rho}^2, \forall n \geq n_0$ in which the vector of eigenvalues $\lambda(\bullet)$ of a matrix $(\bullet) \in \mathbb{R}^{n\times n}$, Γ are the weights of the regularised regression problem. It can easily be deployed into the actual system when Assumption 4 is satisfied. Furthermore, to ensure asymptotic near-optimality with $\mathbb{P}(\leq) 1 - \beta$, the proposed framework can be bounded with K iterations for planning when $N \to \infty$ and $K \to \infty$, which can be expressed as:

$$\overline{\mathcal{J}}(a^\star) - \hat{\mathcal{J}}(a_K) \leq \widetilde{\mathcal{D}} + \mathcal{P}\left(\frac{\log\left(N^{d/2}/\beta\right)}{N}\right) + \mathcal{P}\left(K^{-\frac{\log 1/\eta}{\log \nu}}\right) \tag{14}$$

where $\overline{\mathcal{J}}(a)$ denotes the optimal expectation that is achieved when executing an action $a \in \mathcal{A}$ while $a^\star$ represents an optimal action and $\widetilde{\mathcal{D}}$ represents a constant that signifies an inherent suboptimality endured due to its robustness against instantaneous disturbances $d(t)$.

Proof: See Appendix 2. ∎

To enhance the capability of the proposed approach, the membership in polynomial time is validated to confidently disqualify the (A, ρ)–modeling in Assumption 1 by means of linear programming [48]. It permits the exploration of a diverse range of potential features, reducing uncertainty by ignoring and retaining only models aligned with current observations. Subsequently, all retained hypotheses must be taken into account continuously during the planning and still ensure safety. One significant advantage of employing a linear model set lies in its ability to establish stringent conditions for quantifying how well the modeling assumptions align with the observed data. In particular, the polytopic set (8) is determined with $\mathbb{P}(A(\xi)) \leq 1 - \beta$ given $N - 1$ observations. Then this confidence set can be propagated to the next N observation due to the linearity.

Assumption 5: Suppose there exists a finite set of candidates $\mathcal{F}^m, \forall m \in \mathcal{M}$ with $\mathcal{M} = \{1, 2, \ldots, M\}, M \in \mathbb{N}$ that gathers the dynamics $\mathcal{F}$: $\dot{x}_t(t) = \mathcal{F}^m(x_t(t), u(t)), \forall t \geq 0$.

The designed planning algorithm is adjusted to effectively harmonize these simultaneous hypotheses by maximizing a robust objective that considers discrete ambiguity, which is expressed as

$$\mathcal{J} = \sup_{a\in\mathcal{A}^{\mathbb{N}}} \min_{m\in\mathcal{M}} \sum_{n=N+1}^{\infty} \eta^n J_n^m, \tag{15}$$

where J_n^m denotes the cost incurred by executing the action sequence a up to step n under the dynamics $\mathcal{F}^m$. While the mentioned objective could be optimized using the similar approach above, it would lead to a coarse and imprecise approximation. Thus, the finite uncertainty structure outlined in Assumption 5 is utilized to asymptotically regain the true $\mathcal{J}$. This reformation to the OP algorithm involves replacing the robust upper-bound (12) on sequence values a, that belongs to the look-ahead tree $\mathcal{T}$ as follows:

$$\mathcal{T}_a(k) \overset{def}{=} \min_{m\in\mathcal{M}} \sum_{n=0}^{l_a-1} \eta^n J_n^m + \frac{\eta^l_a}{1-\eta}. \tag{16}$$

In addition, from the definition in (16), the robust upper-bound value and the robust lower-bound value can be extended to inner nodes

$$\mathcal{T}_a(k) \overset{def}{=} \begin{cases} \min\limits_{m\in\mathcal{M}} \sum\limits_{n=0}^{l_a-1} \eta^n J_n^m + \dfrac{\eta^{l_a}}{1-\eta} & \text{if } a \text{ is a leaf;} \\ \max\limits_{b\in\mathcal{A}} \mathcal{T}_{ab}(k) & \text{else.} \end{cases}, \tag{17}$$

$$\mathcal{L}_a(k) \overset{def}{=} \begin{cases} \min\limits_{m\in\mathcal{M}} \sum\limits_{n=0}^{l_a-1} \eta^n J_n^m & \text{if } a \text{ is a leaf;} \\ \max\limits_{b\in\mathcal{A}} \mathcal{L}_{ab}(n) & \text{else.} \end{cases} \tag{18}$$

Remark 4: The exact recovery of a solution to a robust objective (15) with discrete ambiguity is achievable asymptotically as the planning iteration K approaches infinity. It is important to note the distinction from the results concerning the robust objective (10) with continuous ambiguity $A(\xi) \in \hat{\mathcal{D}}_{N,\beta}$. In that context, the OP algorithm only attains the surrogate approximation $\hat{\mathcal{J}}$, as elaborated in Theory 2. Notably, the regret is contingent on the number K of node

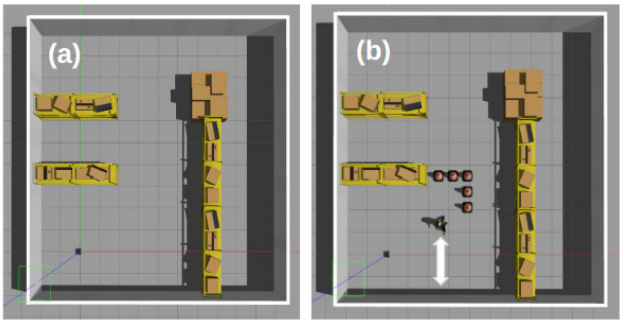

FIGURE 2. Virtual worlds in the Gazebo environment for performance evaluations of the mobile robot control, which placed (a) static obstacles in the first scenario and (b) a pedestrian in the second scenario.

TABLE 1. Parameters of DWA.

Parameters	mobile robot	vehicle
Number of samples in the origin	(8,0,8)	(10,0,10)
Coefficient of the closest point	32	40
Coefficient of avoiding obstacles	0.02	0.25
Coefficient of reaching local goal	20	24

expansions in this case, but each expansion now demands M times more simulations. Ultimately, the new upper-bound (16) is integrated into the surrogate objective function (11).

Theorem 3: On sequence values of actions a, the robust values inherit similar properties as the optimal values. It means that the improved tree-based planning strategy (16) is exhibited with the identical bound property in Lemma 1 with regard to the extended objective (15), it can be bounded as

$$\mathcal{L}_a(k) \leq \mathcal{L}_a(K) \leq \mathcal{J}_a \leq \mathcal{T}_a(K) \leq \mathcal{T}_a(k) \tag{19}$$

where $\mathcal{J}_a \overset{def}{=} \max_{\mathbf{u}\in a,\mathcal{A}^\infty} \min_{m\in\mathcal{M}} \sum_{n=l_a+1}^{\infty} \eta^n J_n^m$.

Proof: See Appendix 3. ■

IV. SIMULATION VALIDATION

This section investigates the effectiveness of the suggested strategy via a series of trials on both the mobile robot in a 2D coordinate system and the autonomous vehicle. These captured simulation environments are expressed in Figure 2 and 5, respectively. For more objectivity, several control parameters are set for all trials. The parameters η and β are set to 0.9, and the planning iteration is denoted as $K = 100$. Disturbances are uniformly sampled $[-0.1, 0.1]^r$ and measurements follow a Gaussian distribution with covariance $\Gamma_s = 0.1^2 I_s$. In comparison, the DWA and TEB methods are popular techniques used in similar situations by quickly generating feasible trajectories that avoid collisions. The parameters for the DWA and TEB strategies are shown in Table 1 and Table 2, respectively.

For the mobile robot, the state x_t includes its position (x, y) and velocity (v_x, v_y), which are actuated by a control signal $u = (u_x, u_y) \in [-1, 1]^2$ under the presence of unknownly anisotropic friction coefficients (ξ_x, ξ_y). The objective is to navigate to a goal state x_g while avoiding collisions with obstacles. The cost function is used to describe this task, that is $\mathcal{J}(x_t) = \frac{\beta(x_t)}{1+\|x_t - x_g\|_2}$, where $\beta(x_t)$ is 0 if a collision occurs and 1 otherwise. The control action set is $\mathcal{A} = \{(-1,-1), (-1,1), (1,-1), (1,1)\}$ corresponding to the up, down, left, and right directions. Due to the specific features of our scenario, a robust baseline cannot be applied. We compare our proposed approach with a conventional control strategy that plans using the estimated dynamics $\hat{\xi}_N$. Both approaches have the same partial knowledge of the dynamical models and cost, remarking on the advantages of our robust formulation that is independent of algorithmic design. Table 3 displays the results of 100 simulations for a single episode. It is evident that the conventional approaches perform worse in time to reach the goal than the proposed method, although they still successfully ensure safety without any collisions in the medium with static obstacles, which is expressed in Figure 3. In the presence of pedestrians in the environment, the conventional approaches collide with

TABLE 2. Parameters of TEB.

Parameters	mobile robot	vehicle
Minimum number of samples	3	5
Automatic resizing hysteresis	0.1	0.15
Coefficient of the safety margin	0.1	0.1
Coefficient of avoiding obstacles	0.05	0.55
Coefficient of non-holonomic kinematics	1000	1000

TABLE 3. Failures rate and travel time of the approach performances of the mobile robot.

Scenarios	Static Obstacle		Pedestrian	
Approaches	Failures	Time (s)	Failures	Time (s)
DWA	0/100	39.6	12/100	49.3
TEB	0/100	44.8	8/100	56.2
Proposed	**0/100**	**39.1**	**1/100**	**43.5**

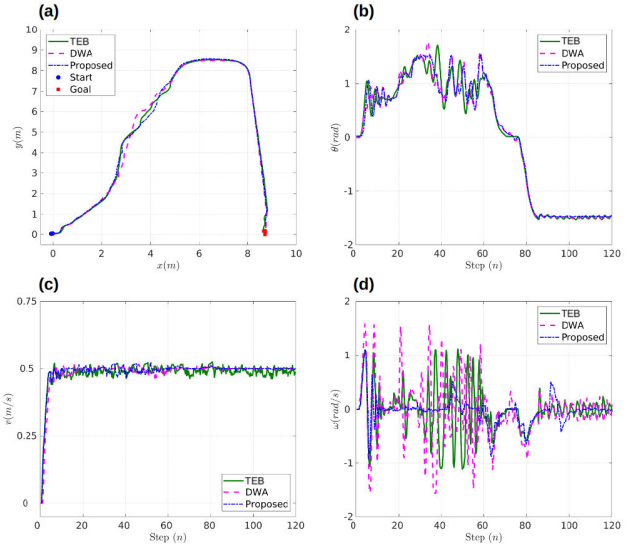

FIGURE 3. Performance test of the mobile robot in the first scenario: (a) Odometry, (b) Yaw angle, (c) Forward velocity, (d) Rotational velocity.

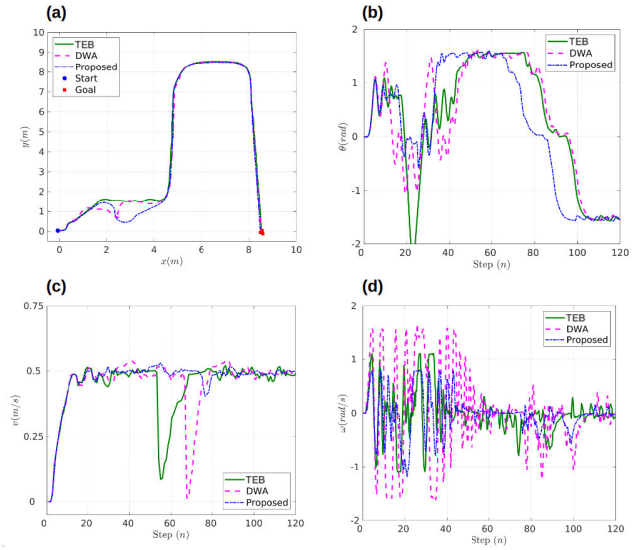

FIGURE 4. Performance test of the mobile robot in the second scenario: (a) Odometry, (b) Yaw angle, (c) Forward velocity, (d) Rotational velocity.

obstacles in 12% and 8% for the DWA and TEB strategies, respectively. Meanwhile, the proposed strategy achieves the goal with little failure, even with the shortest travel time in all trials. Such features of these driving trajectories are observed probably because the proposed approach avoided obstacles through various actions such as stopping and turning to avoid obstacles depending on the situation, which is shown in Figure 4. On the other hand, the others tended to stop first to avoid obstacles and wait for them to pass before driving.

For the autonomous vehicle, the state of the ego-vehicle $x_{t,0} \in \mathbb{R}^4$ is working among $\mathcal{H}$ other vehicles with its states $x_{t,h} \in \mathbb{R}^4$, $\forall h \in \{0, \mathcal{H}\}$, including its position (x_h, y_h), velocity v_h and heading ψ_h. It leads to a joint traffic state $x_t = [x_{t,0}, \ldots, x_{t,\mathcal{H}}]^\top \in \mathbb{R}^{4\mathcal{H}+4}$. These vehicles adhere to predefined behaviors that are characterized by a set of parameters $\dot{x}_h = \mathcal{F}_h(x_t, \xi_h)$ in the presence of unknown parameters $\xi_h \in \mathbb{R}^5$. The objective is to navigate to a predefined goal and to simultaneously track the desired velocity while ensuring safety, i.e., avoiding collisions with vehicles. The cost function is simplified to describe this task in the range 0 to 1, that is, $\mathcal{J}(x_t) = c_1\|x_t - x_g\|_2 + c_2\frac{v}{v_{max}} + c_3\frac{l_c}{l_s - 1} + b_{penalty}$, where $c_1, c_2, c_3, b_{penalty}$ are parameter coefficients; l_c, l_s are the current lane order and the sum of lanes, respectively; and v_{max} denotes the maximum velocity setting in suitable scenarios, respectively. In terms of the scaling of $\mathcal{J}$ in $[0, 1]$, it receives the maximum value of 1 when crashing into another vehicle, while it receives the minimum value of zero when the ego-vehicle reaches full velocity without changing lanes. The control action $\mathcal{A}$ consists of changing the target lane (lateral actions) and velocity (longitudinal actions), i.e., left lane, right land, keep, faster, slower. It is noted that these actions might be disabled separately by setting parameters for specific scenarios. Hereby, the navigating goal of the vehicle is automatically performed by a steering controller

TABLE 4. Failures rate and travel time of the approach performances of the self-driving decision for the autonomous vehicle.

Scenarios	Highway		T-junction	
Approaches	Failures	Time (s)	Failures	Time (s)
DWA	9/100	**9.2**	7/100	16.3
TEB	6/100	12.8	4/100	19.2
Proposed	**0/100**	9.5	**0/100**	**15.7**

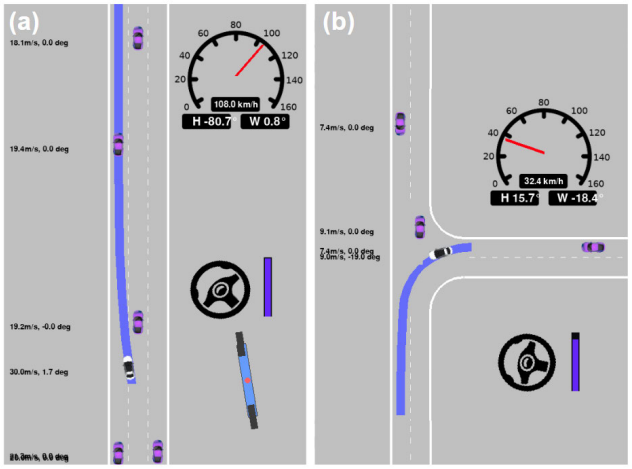

FIGURE 5. Captured simulation in the self-develop environment for performance evaluations of the self-driving decision for the autonomous vehicle. (a) highway scenario and (b) T-junction scenario.

to reach the lane in a highway scenario with $c_1 = 0, c_2 = 0.4, c_3 = 0.1, b_{penalty} = -1, v_{max} = 30m/s$, while it is ignored in a T-junction scenario where only the velocity is considered with $c_1 = 0.4, c_2 = 0, c_3 = 0, b_{penalty} = -1$, $v_{max} = 10m/s$.

Thanks to the advantage of the structure imposed in (2), the proposed strategy results in a greatly shrinking uncertainty space $\hat{\xi}$ when the whole state matrix A is estimated. The series of repeated trials are performed and the results are gathered in Table 4. It is expressed that the rate of failure of other approaches is 9% and 6% in the highway scenario and a bit reduced to 7% and 4% in the T-junction for the DWA and TEB strategies, respectively. Meanwhile, the proposed approach completely guarantees safety with zero failure cases in both scenarios. Although the travel time is slightly longer compared with the DWA in the highway scenario, it is still acceptable when safety is a priority in situations where robust durability is necessary. Our method demonstrates superior performance in the most challenging scenarios and consistently prevents collisions, albeit at the expense of slightly increased travel time. Additionally, we examine how suboptimality, expressed as $\overline{\mathcal{J}}(x_t^N) - \sum_{n>N} \eta^{n-N} J(x_t^n)$, changes relative to the parameter N. This process involves comparing the empirical returns from a state x_t^N to the optimal value $\overline{\mathcal{J}}(x_t^N)$ that the controller would achieve by acting optimally with completely known dynamics. While the assumptions of Theorem 2 are not entirely met, such as due to a non-smooth cost, the average suboptimality of the

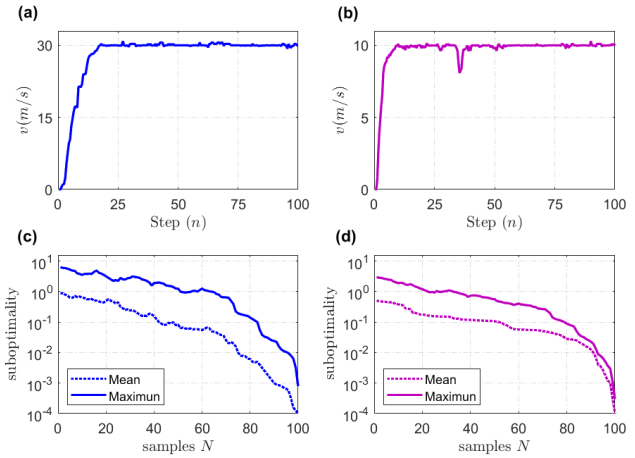

FIGURE 6. Performance of the proposed strategy in the self-develop environment: achieved velocity and the mean and maximum suboptimality (a), (c) in the highway scenario, and (b), (d) in the T-junction scenario, respectively.

designed framework still decreases polynomially concerning the number of samples, which is expressed in Figure 6.

V. CONCLUSION

The suggested framework introduces an integrated robust-adaptive planning control system designed to ensure motion safety for perturbed linear systems using the minimax objective function. This method incorporates a set-membership mechanism based on recursive least squares, interval prediction, and tree-based planning to ensure predicted performance while establishing a suboptimality bound via generic costs. With the partially known dynamical model, the flexibility of the proposed framework is further extended through a multi-model approach, and its efficacy is investigated through simulation studies. Additionally, it is common practice to estimate model parameters from observational data, and confidence intervals are frequently available but not fully utilized in the decision-making process. On the other hand, the proposed approach enables the assessment and optimization of worst possible outcomes for manifold strategies.

However, there are certain disadvantages that need to be improved, e.g., the reliability of these guarantees is contingent upon the accuracy of the underlying assumptions or the set of possible dynamics ambiguity that has to satisfy specific properties. Specifically, Assumption 1 requires thorough validation. If not, there is a risk that decisions based on incorrect foundations could have severe repercussions in such critical environments. More generally, safety analyses can never protect against unmodeled events, such as a package falling down or continually changing lanes.

A promising research direction will focus on the study of the more appropriate pure exploration setting, which aims at relating the policy suboptimality to the number of samples used. It also overcomes the current limitations of relying on the benefits of partially known dynamical models. This initiative aims to efficiently exploit the utilization of observational data and confidence intervals, thus refining decision-making processes. Another path of interest to us could be to leverage offline Reinforcement Learning methods that could enable to safely improve pre-trained policies using real driving data, especially around nominal states that can be confidently estimated, thereby elevating the adaptability and accuracy of the control system. A fine-tuning training process in real conditions would involve experiencing real failures again, though hopefully in reduced numbers, which could be realistic under human interventions.

APPENDIX

A. PROOF OF THEOREM 1

Proof: The predictor designed in Section 2 satisfies the inclusion property (7). Consequently, for any dynamics $A(\xi) \in \hat{\mathcal{D}}_{N,\beta}$ and disturbances $d \in [\underline{d}, \overline{d}]$, if the corresponding state at time t_n is bounded by $x_t \in [\underline{x}_n, \overline{x}n]$, it implies that $J(x_t^n) \geq \min x_t \in [\underline{x}_n(\mathbf{u}), \overline{x}_n(\mathbf{u})] J(x_t) = \underline{J}_n(\mathbf{u})$ for a sequence of controls $\mathbf{u}$.

Thus, by taking the min over $\hat{\mathcal{D}}_{N,\beta}$ and $[\underline{d}, \overline{d}]$, we also have for any sequence of controls $\mathbf{u}$:

$$\begin{aligned} \mathcal{J}(\mathbf{u}) &= \min_{\substack{A(\xi)\in\hat{\mathcal{D}}_{N,\beta} \\ \underline{d}\leq d\leq\overline{d}}} \sum_{n=N+1}^{\infty} \eta^n J(x_t^n) \\ &\geq \sum_{n=N+1}^{\infty} \eta^n \underline{J}_n(\mathbf{u}) \\ &= \hat{\mathcal{J}}(\mathbf{u}) \end{aligned} \tag{20}$$

And $\mathcal{J}(\mathbf{u}) \leq \mathbb{E}_d \left[\sum_{n=N+1}^{\infty} \eta^n J(x_t^n)\right]$ by definition. The proof is completed. ■

B. PROOF OF THEOREM 2

Proof: We have

$$\|\xi - \hat{\xi}_{N,\phi}\|^2_{\mathcal{G}_{N,\phi}} \geq \lambda_{\min}(\mathcal{G}_{N,\phi}) \|\xi - \hat{\xi}_{N,\phi}\|_2^2 \tag{21}$$

where the Gramian matrix $\mathcal{G}_{N,\phi} = \sum_{n=1}^{N} \mathcal{E}_n^\top \Gamma^{-1} \mathcal{E}_n + \phi I_d \in \mathbb{R}^{d\times d}$ with penalty parameter $\phi \in \mathbb{R}_*^+$ and the identity matrix I_d and Remark 1 gives that the model estimation error can be bounded

$$||\xi - \hat{\xi}_{N,\phi}|| \leq \widetilde{\delta}_N(\beta) \tag{22}$$

Moreover, it is propagated through the state prediction $A(\xi)$. Let the j^{th} column of a matrix M be represented as M_j and its coefficient at position i, j denoted as $M_{i,j}$, then $((A(\xi) - A(\hat{\xi}_{N,\phi}))^\top (A(\xi) - A(\hat{\xi}_{N,\phi})))_{i,j} = (\xi - \hat{\xi}_{N,\phi})^\top \rho_i^\top \rho_j (\xi - \hat{\xi}_{N,\phi}) \leq \lambda_{\max}(\rho_i^\top \rho_j) \|\xi - \hat{\xi}_{N,\phi}\|_2^2$. Thus, the boundedness of the estimation error is given

$$\|A(\xi) - A(\hat{\xi}_{N,\phi})\|_F^2 \leq \mathcal{P}\left(\frac{\widetilde{\delta}_N(\beta)^2}{\lambda_{\min}(\mathcal{G}_{N,\phi})}\right) \tag{23}$$

Next, for the boundedness of the prediction error $e = \overline{x} - \underline{x}$, a candidate Lyapunov function is selected as follow:

$$\overline{\mathcal{J}} = e^\top P e \tag{24}$$

which $\overline{\mathcal{J}}$ is non-negative definite provided that $P > 0$ and $X = \left[e\ \overline{d} - \underline{d}\ \underline{x}^+ + \overline{x}^-\right]^\top$, for any $Q \in \mathbb{R}^{p\times p}$, $\rho, \alpha \in \mathbb{R}$ and $|\underline{x}^+ + \overline{x}^-| \leq 2|e|$. Then taking its derivative

$$\begin{aligned}\dot{\overline{\mathcal{J}}} &= X^\top \begin{bmatrix} A_N^\top P + PA_N + Q & P & P|\mathcal{X}| \\ P & -\rho I_r & 0 \\ |\mathcal{X}|^\top P & 0 & -\alpha I_p \end{bmatrix} X \\ &\quad - e^\top Qe + \alpha|\underline{x}^+ + \overline{x}^-|^2 + \rho|\overline{d} - \underline{d}|^2 \\ &\leq X^\top \Upsilon X - e^\top Qe + \rho\|\overline{d} - \underline{d}\|_2^2 \end{aligned} \tag{25}$$

where $\Upsilon = \left[\begin{array}{cc|c} A_N^\top P + PA_N + Q + 4\alpha I_p & P & P|\mathcal{X}| \\ P & -\rho I_r & 0 \\ \hline |\mathcal{X}|^\top P & 0 & -\alpha I_p \end{array}\right]$

As a consequence, we would have $\dot{\overline{\mathcal{J}}} \leq -\mu\overline{\mathcal{J}} + \rho\|\overline{d} - \underline{d}\|_2^2$ with $\mu = \frac{\lambda_{\min}(Q)}{\lambda_{\max}(P)}$ if we had $\Upsilon \leq 0$, $Q > 0$, $\rho > 0$. Since $\overline{\mathcal{J}}(t_N) = 0, \forall t > t_N$, this further implies that

$$\overline{\mathcal{J}}(t) \leq \frac{\rho}{\mu}\Phi^2(t) \tag{26}$$

where $\Phi(t) = \sup_{\tau\in[0,t]} \|\overline{d}(\tau) - \underline{d}(\tau)\|_2$.

Next, the expression Υ is non-positive if and only if $J + R < 0$, where $J = \begin{bmatrix} A_N^\top P + PA_N + Q + 4\alpha I_p & P \\ P & -\rho I_r \end{bmatrix}$, $R = \alpha^{-1}\left[|\mathcal{X}|^\top P\ 0\right]^\top \left[|\mathcal{X}|^\top P\ 0\right]$ and $\alpha > 0$.

To investigate this condition, $Q = \frac{1}{2}Q_0 - 4\alpha I_p$ is set. Suppose that Assumption 4 is satisfied and P is fixed, one obtains

$$\begin{aligned}\lambda_{\max}(R) &\leq \alpha^{-1}\lambda_{\max}(P)^2\lambda_{\max}(|\mathcal{X}|^\top|\mathcal{X}|) \\ &\leq \alpha^{-1}\lambda_{\max}(P)^2\|\mathcal{X}\|_F^2 \end{aligned} \tag{27}$$

Next, we can obtain $R \leq \begin{bmatrix} \frac{1}{2}Q_0 & 0 \\ 0 & 0 \end{bmatrix}$ by taking $\alpha = \frac{2\lambda_{\max}(P)^2\|\mathcal{X}\|_F^2}{\lambda_{\min}(Q_0)}$. Consequently,

$$J + R \leq \begin{bmatrix} A_N^\top P + PA_N + Q_0 & P \\ P & -\rho I_r \end{bmatrix} < 0 \tag{28}$$

Therefore, we obtain $\Upsilon \leq 0$ with such a selection of Q and α. Based on (26), $\lambda_{\min}(Q_0) = 2\lambda_{\min}(Q) + 8\alpha$, we attain the inequality as following:

$$\|e(t)\|_2^2 \leq \frac{2\rho\lambda_{\max}(P)}{\lambda_{\min}(P)(\lambda_{\min}(Q_0) - 8\alpha)}\Phi^2(t) \tag{29}$$

Ultimately, the prediction error can be finally bounded as

$$\|e(t)\|_2 \leq \left(\sqrt{\frac{2\rho\lambda_{\max}(P)}{\lambda_{\min}(P)\lambda_{\min}(Q_0)}} + \mathcal{P}\left(\frac{\widetilde{\delta}_N(\beta)^2}{\lambda_{\min}(\mathcal{G}_{N,\phi})}\right)\right)\Phi(t) \tag{30}$$

Meantime, we have $x_t \in [\underline{x}_t(\mathbf{u}), \overline{x}_n(\mathbf{u})]$ by the property (7) that yields $J(x_t^n) \leq \max J(x_t)$. In addition, according to (3) and (8), we get $\lambda_{\min}(\mathcal{G}_{N,\phi}) \geq (N - n_0)\underline{\rho}^2 + \sum_{n<n_0}\mathcal{E}_n^\top\Gamma^{-1}\mathcal{E}_n$ and $\widetilde{\delta}_N(\beta) = \sqrt{2\log\left(\frac{\det(\mathcal{G}_{N,\phi})^{1/2}}{\beta\det(\phi I_d)^{1/2}}\right)} + \sqrt{\phi d}R$, one obtains $\frac{\widetilde{\delta}_N(\beta)^2}{\lambda_{\min}(\mathcal{G}_{N,\phi})}$ equivalent to $\frac{\log(N^{d/2}/\beta)}{N}$. Supposing that J is L-lipschitz, then

$$\begin{aligned}\overline{\mathcal{J}}(\mathbf{u}) - \hat{\mathcal{J}}(\mathbf{u}) &\leq \sum_{n=N+1}^{\infty}\eta^n \underset{x_t\in[\underline{x}_n(\mathbf{u}),\overline{x}_n(\mathbf{u})]}{(\max\text{ - }\min)} J(x) \\ &\leq \sum_{n=N+1}^{\infty}\eta^n L\left\|\underline{x}_n(\mathbf{u}) - \overline{x}_n(\mathbf{u})\right\|_2 \\ &\leq \widetilde{\mathcal{D}} + \mathcal{P}\left(\frac{\log(N^{d/2}/\beta)}{N}\right) \end{aligned} \tag{31}$$

with $\widetilde{\mathcal{D}} = L\sqrt{\frac{2\rho\lambda_{\max}(P)}{\lambda_{\min}(P)\lambda_{\min}(Q_0)}}\sum_{n>N}\eta^n\Phi(t_n)$.

Finally, the consequence of Lemma 1 is utilized to account for planning with a finite K-iteration and relate $\hat{\mathcal{J}}(a^\star)$ to $\hat{\mathcal{J}}(a_K)$. The proof is completed. ■

C. PROOF OF THEOREM 3

Proof: Initially, the values $\mathcal{L}_a^m(k)$ and $\mathcal{T}_a^m(k)$ signify the lowest and highest admissible cost achieved by the possible continuations of a sequence a, respectively. Consequently, $\mathcal{L}_a^m(k)$ and $\mathcal{T}_a^m(k)$ denote non-decreasing and non-increasing functions with k, respectively, whereas J_a^m and $\mathcal{J}_a$ remain independent of k.

Additionally, supposing that $0 \leq J \leq 1$, then we have $\eta^\omega + \eta^{\omega+1} + \cdots \approx \frac{\eta^\omega}{1-\eta}$ from a node of depth ω. Consequently, for any sequence of costs $(J_n)_{n\in\mathbb{N}}$ came by path-tracking in $a\mathcal{A}^\infty$ with any dynamical models: $\mathcal{L}_a^m(k) = \sum_{n=0}^{\omega-1}\eta^n J_n^m \leq \sum_{n=0}^{\infty}\eta^n J_n^m \leq \sum_{n=0}^{\omega-1}\eta^n J_n^m + \frac{\eta^\omega}{1-\eta} = \mathcal{T}_a^m(k), \forall k \geq 0, a \in \mho_k$ of depth ω. We obtain,

$$\min_{m\in\mathcal{M}}\mathcal{L}_a^m(k) \leq \min_{m\in\mathcal{M}}\sum_{n=0}^{\infty}\eta^n J_n \leq \min_{m\in\mathcal{M}}\mathcal{T}_a^m(k) \tag{32}$$

It is noticeable that both sides of (32) are independent of the specific path followed in $a\mathcal{A}^\infty$. This independence extends to the robust path as well, i.e., $\min_{m\in\mathcal{M}}\mathcal{L}_i^m(k) \leq \max_{a\in a\mathcal{A}^\infty}\min_{m\in\mathcal{M}}\sum_{t=0}^{\infty}\eta^n J_n^m \leq \min_{m\in\mathcal{M}}\mathcal{T}_i^m(k)$, then

$$\mathcal{L}_a(k) \leq \mathcal{J}_a \leq \mathcal{T}_a(k) \tag{33}$$

In accordance with Theorem 1, the boundedness $\frac{\eta^{\omega_K}}{1-\eta}$ of the simple regret ε_K is first introduced in [45], where ω_K denotes the depth of $\mathcal{T}_K$. Indeed, this feature is contingent on the returned action being part of the deepest explored branch. Its evidence is similarly established in (19). Accordingly, this directly implies that the returned action is denoted as $a = i_0$, where i represents some node of maximal depth ω_K expanded at round $k \leq K$. Then, this choice law confirms $\mathcal{T}_a(k) = \mathcal{T}_i(k) = \max_{x_t\in\mathcal{A}}\mathcal{T}_{x_t}(k)$ and $\mathcal{J} - \mathcal{J}_a = \mathcal{J}_{a^\star} - \mathcal{J}_a \leq \mathcal{T}_{a^\star}(k) - \mathcal{J}_a(k) \leq \mathcal{T}_a(k) - \mathcal{L}_a(k) = \mathcal{T}_i(k) - \mathcal{L}_i(k) = \frac{\eta^{\omega_K}}{1-\eta}$.

Subsequently, the depth ω_K of $\mathcal{T}_K$ can be bounded with the planning K iteration. To the end, the expanded nodes are consistently illustrated at the subtree $\mathcal{T}_\infty$, encompassing all nodes of an optimal depth ω that are $\frac{\eta^\omega}{1-\eta}$. Consequently, the max-backups of (16) up to the root bring $\mathcal{T}_i(k) = \mathcal{T}_\emptyset(k)$ when $\mathcal{T}_i(k) \geq \mathcal{T}_j(k), \forall j \in \mho_k$ where a node i of depth ω is expanded at round k.

Applying the problem-dependent variable ν to nodes in $\mathcal{T}_\infty$, it is assumed that ω_0 and c such that the quantity n_ω, representing the number of nodes with a depth of $\omega \geq \omega_0$ in $\mathcal{T}_\infty$, is constrained by $c\nu^\omega$. Consequently,

$$K = \sum_{\omega=0}^{\omega_K} n_\omega = n_0 + \sum_{\omega=\omega_0+1}^{\omega_K} n_\omega \leq n_0 + c \sum_{\omega=\omega_0+1}^{\omega_K} \nu^\omega \quad (34)$$

We consider two scenarios: $\nu = 1$ and $\nu > 1$. Firstly, the regret can be bounded as $\varepsilon_K \leq \mathcal{P}\left(\eta^{Kc}\right)$ due to $K \leq n_0 + c(\omega_K - \omega_0)$ when $\nu = 1$. In the case of $\nu > 1$, the boundedness value $\varepsilon_K \leq \frac{\eta^{\omega_K}}{1-\eta} = \frac{1}{1-\eta}\left(\frac{(K-n_0)(\nu-1)}{c\nu^{\omega_0+1}}\right)^{\frac{\log\eta}{\log\nu}} \leq \mathcal{P}\left(K^{-\frac{\log 1/\eta}{\log\nu}}\right)$ due to $K \leq n_0 + c\nu^{\omega_0+1}\frac{\nu^{\omega_K-\omega_0}-1}{\nu-1}$ and $\omega_K \geq \omega_0 + \log_\nu \frac{(K-n_0)(\nu-1)}{c\nu^{\omega_0+1}}$. The proof is completed. ■

REFERENCES

[1] M. Hossain, "Autonomous delivery robots: A literature review," *IEEE Eng. Manag. Rev.*, vol. 51, no. 4, pp. 77–89, Dec. 2023.
[2] V. Mahor, S. Bijrothiya, R. Mishra, R. Rawat, and A. Soni, "The smart city based on ai and infrastructure: A new mobility concepts and realities," in *Autonomous Vehicles: Using Machine Intelligence*, vol. 1, R. Rawat, A. M. Sowjanya, S. I. Patel, V. Jaiswal, I. Khan, and A. Balaram, Eds. Hoboken, NJ, USA: Wiley, 2022, ch. 1.
[3] K. Berntorp, "Path planning and integrated collision avoidance for autonomous vehicles," in *Proc. Amer. Control Conf. (ACC)*, Seattle, WA, USA, May 2017, pp. 4023–4028.
[4] C. Ntakolia, S. Moustakidis, and A. Siouras, "Autonomous path planning with obstacle avoidance for smart assistive systems," *Expert Syst. Appl.*, vol. 213, Mar. 2023, Art. no. 119049.
[5] Z. Jian, Z. Yan, X. Lei, Z. Lu, B. Lan, X. Wang, and B. Liang, "Dynamic control barrier function-based model predictive control to safety-critical obstacle-avoidance of mobile robot," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2023, pp. 3679–3685.
[6] N. P. Nguyen, H. Oh, and J. Moon, "Continuous nonsingular terminal sliding-mode control with integral-type sliding surface for disturbed systems: Application to attitude control for quadrotor UAVs under external disturbances," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 58, no. 6, pp. 5635–5660, Dec. 2022.
[7] B. Moudoud, H. Aissaoui, and M. Diany, "Adaptive integral-type terminal sliding mode control: Application to trajectory tracking for mobile robot," *Int. J. Adapt. Control Signal Process.*, vol. 37, no. 3, pp. 603–616, Mar. 2023.
[8] Y. Zheng, J. Zheng, K. Shao, H. Zhao, H. Xie, and H. Wang, "Adaptive trajectory tracking control for nonholonomic wheeled mobile robots: A barrier function sliding mode approach," *IEEE/CAA J. Autom. Sinica*, early access, pp. 1–15, 2024.
[9] T. Roy and R. K. Barai, "Control-oriented LFT modelling and H-control of differentially driven wheeled mobile robot," in *Robust Control-Oriented Linear Fractional Transform Modelling*. Cham, Switzerland: Springer, 2023, pp. 111–133.
[10] S. Dean, H. Mania, N. Matni, B. Recht, and S. Tu, "Regret bounds for robust adaptive control of the linear quadratic regulator," in *Adv. Neural Inf. Process. Syst.*, S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, Eds. Red Hook, NY, USA: Curran Associates, Inc., 2018, pp. 4188–4197.
[11] W. Xiao, G. Wang, J. Tian, and L. Yuan, "A novel adaptive robust control for trajectory tracking of mobile robot with uncertainties," *J. Vibrat. Control*, early access, Mar. 2023, Art. no. 107754632311618.
[12] M. Ammour, R. Orjuela, and M. Basset, "A MPC combined decision making and trajectory planning for autonomous vehicle collision avoidance," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 12, pp. 24805–24817, Dec. 2022.
[13] H. Yang, Y. He, Y. Xu, and H. Zhao, "Collision avoidance for autonomous vehicles based on MPC with adaptive APF," *IEEE Trans. Intell. Vehicles*, vol. 9, no. 1, pp. 1559–1570, Jan. 2024.
[14] Y. Li, J. Fan, Y. Liu, and X. Wang, "Path planning and path tracking for autonomous vehicle based on MPC with adaptive dual-horizon-parameters," *Int. J. Automot. Technol.*, vol. 23, no. 5, pp. 1239–1253, Oct. 2022.
[15] X. Zhang, W. Zhang, Y. Zhao, H. Wang, F. Lin, and Y. Cai, "Personalized motion planning and tracking control for autonomous vehicles obstacle avoidance," *IEEE Trans. Veh. Technol.*, vol. 71, no. 5, pp. 4733–4747, May 2022.
[16] Y. Dai and D. Wang, "A tube model predictive control method for autonomous lateral vehicle control based on sliding mode control," *Sensors*, vol. 23, no. 8, p. 3844, Apr. 2023.
[17] A. Wischnewski, T. Herrmann, F. Werner, and B. Lohmann, "A tube-MPC approach to autonomous multi-vehicle racing on high-speed ovals," *IEEE Trans. Intell. Vehicles*, vol. 8, no. 1, pp. 368–378, Jan. 2023.
[18] Y. Mi, K. Shao, Y. Liu, X. Wang, and F. Xu, "Integration of motion planning and control for high-performance automated vehicles using tube-based nonlinear MPC," *IEEE Trans. Intell. Vehicles*, early access, pp. 1–16, 2023.
[19] G. Chang and Q. Suqin, "An adaptive MPC trajectory tracking algorithm for autonomous vehicles," in *Proc. 17th Int. Conf. Comput. Intell. Secur. (CIS)*, Chengdu, China, Nov. 2021, pp. 197–201.
[20] C. P. Vo, J. Lee, and J. H. Jeon, "Robust adaptive path tracking control scheme for safe autonomous driving via predicted interval algorithm," *IEEE Access*, vol. 10, pp. 124333–124344, 2022.
[21] Y. Liang, Y. Li, A. Khajepour, Y. Huang, Y. Qin, and L. Zheng, "A novel combined decision and control scheme for autonomous vehicle in structured road based on adaptive model predictive control," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 9, pp. 16083–16097, Sep. 2022.
[22] B. A. Hernandez Vicente and P. A. Trodden, "Stabilizing predictive control with persistence of excitation for constrained linear systems," *Syst. Control Lett.*, vol. 126, pp. 58–66, Apr. 2019.
[23] D. Fox, W. Burgard, and S. Thrun, "The dynamic window approach to collision avoidance," *IEEE Robot. Autom. Mag.*, vol. 4, no. 1, pp. 23–33, Mar. 1997.
[24] L.-S. Liu, J.-F. Lin, J.-X. Yao, D.-W. He, J.-S. Zheng, J. Huang, and P. Shi, "Path planning for smart car based on Dijkstra algorithm and dynamic window approach," *Wireless Commun. Mobile Comput.*, vol. 2021, pp. 1–12, Feb. 2021.
[25] T. Hossain, H. Habibullah, R. Islam, and R. V. Padilla, "Local path planning for autonomous mobile robots by integrating modified dynamic-window approach and improved follow the gap method," *J. Field Robot.*, vol. 39, no. 4, pp. 371–386, Jun. 2022.
[26] C. Rösmann, F. Hoffmann, and T. Bertram, "Timed-elastic-bands for time-optimal point-to-point nonlinear model predictive control," in *Proc. Eur. Control Conf. (ECC)*, Jul. 2015, pp. 3352–3357.
[27] J. Wu, X. Ma, T. Peng, and H. Wang, "An improved timed elastic band (TEB) algorithm of autonomous ground vehicle (AGV) in complex environment," *Sensors*, vol. 21, no. 24, p. 8312, Dec. 2021.
[28] K. Fotova Čiković, I. Martinč ević, and J. Lozić, "Application of data envelopment analysis (DEA) in the selection of sustainable suppliers: A review and bibliometric analysis," *Sustainability*, vol. 14, no. 11, p. 6672, May 2022.
[29] G.-G. Wang, D. Gao, and W. Pedrycz, "Solving multiobjective fuzzy job-shop scheduling problem by a hybrid adaptive differential evolution algorithm," *IEEE Trans. Ind. Informat.*, vol. 18, no. 12, pp. 8519–8528, Dec. 2022.
[30] A. Seyyedabbasi and F. Kiani, "Sand cat swarm optimization: A nature-inspired algorithm to solve global optimization problems," *Eng. Comput.*, vol. 39, no. 4, pp. 2627–2651, Aug. 2023.
[31] F. Kiani, A. Seyyedabbasi, and P. Mahouti, "Optimal characterization of a microwave transistor using grey wolf algorithms," *Anal. Integr. Circuits Signal Process.*, vol. 109, no. 3, pp. 599–609, Dec. 2021.
[32] I. Khettabi, L. Benyoucef, and M. A. Boutiche, "Sustainable multi-objective process planning in reconfigurable manufacturing environment: Adapted new dynamic NSGA-II vs new NSGA-III," *Int. J. Prod. Res.*, vol. 60, no. 20, pp. 6329–6349, Oct. 2022.
[33] G. N. Iyengar, "Robust dynamic programming," *Math. Oper. Res.*, vol. 30, no. 2, pp. 257–280, 2005.
[34] A. Nilim and L. El Ghaoui, "Robust control of Markov decision processes with uncertain transition matrices," *Operations Res.*, vol. 53, no. 5, pp. 780–798, Oct. 2005.
[35] W. Wiesemann, D. Kuhn, and B. Rustem, "Robust Markov decision processes," *Math. Oper. Res.*, vol. 38, no. 1, pp. 153–183, 2013.

[36] H. Sun, L. Yang, Y. Chen, and X. Zhang, "Controlling tractor-semitrailer vehicles in automated highway systems: Adaptive robust and Lyapunov minimax approach," *Asian J. Control*, vol. 23, no. 6, pp. 2642–2656, Nov. 2021.
[37] M. Saraoglu, H. Jiang, M. Schirmer, I. Mutlu, and K. Janschek, "A minimax-based decision-making approach for safe maneuver planning in automated driving," in *Proc. Amer. Control Conf. (ACC)*, San Diego, CA, USA, May 2023, pp. 4683–4690.
[38] M. Kazem Shirani Faradonbeh, A. Tewari, and G. Michailidis, "Optimism-based adaptive regulation of linear-quadratic systems," 2017, *arXiv:1711.07230*.
[39] Y. Ouyang, M. Gagrani, and R. Jain, "Learning-based control of unknown linear systems with Thompson sampling," 2017, *arXiv:1709.04047*.
[40] M. Abeille and A. Lazaric, "Improved regret bounds for Thompson sampling in linear quadratic control problems," in *Proc. 35th Int. Conf. Mach. Learn.*, vol. 80, Jul. 2018, pp. 1–9.
[41] S. Dean, H. Mania, N. Matni, B. Recht, and S. Tu, "On the sample complexity of the linear quadratic regulator," 2017, *arXiv:1710.01688*.
[42] M. Turchetta, F. Berkenkamp, and A. Krause, "Safe exploration in finite Markov decision processes with Gaussian processes," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 1–9.
[43] D. Bertsimas, D. B. Brown, and C. Caramanis, "Theory and applications of robust optimization," *SIAM Rev.*, vol. 53, no. 3, pp. 464–501, 2011.
[44] B. L. Gorissen, İ. Yanıkoğlu, and D. den Hertog, "A practical guide to robust optimization," *Omega*, vol. 53, pp. 124–137, Jun. 2015.
[45] J.-F. Hren and R. Munos, "Optimistic planning of deterministic systems," in *Lecture Notes in Computer Science*, S. Girgin, M. Loth, R. Munos, P. Preux, and D. Ryabko, Eds. Cham, Switzerland: Springer, 2008, pp. 151–164.
[46] J.-J. E. Slotine and W. Li, *Applied Nonlinear Control*. Upper Saddle River, NJ, USA: Prentice-Hall, 1991.
[47] G. Tan, C. Wen, and Y. Chai Soh, "Identification for systems with bounded noise," *IEEE Trans. Autom. Control*, vol. 42, no. 7, pp. 996–1001, Jul. 1997.
[48] V. Delos and D. Teissandier, "Minkowski sum of polytopes defined by their vertices," *J. Appl. Math. Phys.*, vol. 3, no. 1, pp. 62–67, 2015.

CONG PHAT VO received the B.E. degree in electrical and electronic engineering and the S.M. degree in mechatronics engineering from Ho Chi Minh City University of Technology and Education, Vietnam, in 2013 and 2016, respectively, and the Ph.D. degree in mechanical and automotive engineering from the University of Ulsan, Ulsan, South Korea, in 2021.

He is currently a Postdoctoral Researcher in electrical engineering with Ulsan National Institute of Science and Technology (UNIST), Ulsan. His current research interests include optimal control, optimization, reinforcement learning, and motion planning.

PHILJOON JUNG received the B.S. degree in electronic and electrical engineering from Hongik University, Seoul, South Korea, in 2018, and the M.S. degree in electrical engineering from Ulsan National Institute of Science and Technology, Ulsan, South Korea, in 2020.

From 2020 to 2023, he was a Research Member with the Autonomous Robot Division, Twinny Company Ltd. In 2023, he joined Hanwha Systems Ltd., Seongnam, South Korea, where he is currently a Researcher. His research interests include computer vision, robot perception algorithms, and deep learning.

TAE-HYUN KIM received the B.S. degree in computer software engineering from the Kumoh Institute of Technology, Gyeongsangbuk-do, South Korea, in 2020.

From 2019 to 2022, he was with Navis Automotive Systems Company Ltd. He is currently a Researcher with Hanwha Systems. His research interests include algorithms and embedded SW.

JEONG HWAN JEON (Member, IEEE) received the B.S. degree in mechanical and aerospace engineering from Seoul National University, Seoul, South Korea, in 2007, and the S.M. and Ph.D. degrees in aeronautics and astronautics from Massachusetts Institute of Technology (MIT), Cambridge, MA, USA, in 2009 and 2015, respectively.

He was with nuTonomy (an Aptiv company, since its 2017 acquisition) as a Senior/Principal Research Scientist. He is currently an Assistant Professor of electrical engineering with Ulsan National Institute of Science and Technology (UNIST), Ulsan, South Korea. His current research interests include algorithmic, computational, data-based, and control-theoretic approaches to the decision-making, planning, and control architectures for autonomous systems and future mobility, including self-driving cars.