

Null-space Active Perception Enhancement for observing Space-bounded Repetitive Dynamic Events

Nikolaos Efstathopoulos

ddk228@edu.hmu.gr

Hellenic Mediterranean University

Vasileios Tsimpos

Hellenic Mediterranean University

Dimitrios Papageorgiou

Aristotle University of Thessaloniki

Research Article

Keywords: Autonomous Observer, Active Perception, Visual servoing, Robot control

Posted Date: October 26th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10903739/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Null-space Active Perception Enhancement for observing Space-bounded Repetitive Dynamic Events

Nikolaos Efstathopoulos^{1*}, Vasileios Tsimpos¹ and
Dimitrios Papageorgiou^{2,3}

^{1*}Department of Electrical and Computer Engineering, Hellenic Mediterranean University, Estavromenos, Heraklion, 71410, Crete, Greece.

²Department of Electrical and Computer Engineering, Aristotle University of Thessaloniki, Thessaloniki, 54124, Greece.

³Institute of Computer Science, Foundation for Research and Technology–Hellas, Heraklion, 70013, Crete, Greece.

*Corresponding author(s). E-mail(s): ddk228@edu.hmu.gr;
Contributing authors: th20477@edu.hmu.gr; dimpapag@ece.auth.gr;

Abstract

Smoothly switching from “*looking at something closely*” to “*looking at the bigger picture*” and vice versa, is a common aspect of everyday human behavior that enables us to perceive and interpret our environment effectively. In most cases, this adaptation is guided by an intuitive estimation of the observed target’s motion envelope or range, which humans continuously adjust according to the target’s dynamics. Drawing our inspiration from this human functionality, in this work, we introduce a novel active perception enhancement controller, that enables autonomous on-line viewpoint adaptation for actively perceiving space-bounded Repetitive Dynamic Events with minimum uncertainty. The control enhancement is based on an approximation of the region of motion of the observed target and exploits the null-space of the main task of the observer for enhancing the quality of observation. The system is experimentally evaluated using a UR10e robot with an in-hand camera as the active observer, while the observed phenomenon is emulated employing both a robotic actor, as a comparison benchmark, and a human actor, for testing the system in a realistic use case (e.g. in the context of Learning by Demonstration). The experimental evaluation

shows the improvement achieved by the usage of the proposed method, in “understanding” the margins of the observed motion and consequently maintaining the whole behavior of the dynamic target within the field-of-view.

Keywords: Autonomous Observer, Active Perception, Visual servoing, Robot control

1 Introduction

In recent decades, the need for autonomous robotic agents in various unstructured environments, such as agricultural or domestic setups, is rapidly increasing. Future robots are expected to operate autonomously while directly facilitating the human in their every-day lives. Autonomy in robotic agents, however, presumes a solid and reliable perception system, even in cases in which the observed phenomenon or event is highly dynamic and uncertain, e.g. a fast motion of a vehicle, or a walking human.

The concept of perception is practical, but also deeply philosophical, constituting one of the central pillars of Gnosticism, Philosophy of Mind, and Phenomenology. This subsequently means that it is impossible for someone to define it in a unique way and study it universally in all its aspects. Assuming the availability of an autonomous active robotic observer, i.e. an agent able of moving in space, in this work we consider perception from the perspective of optimal scene and viewpoint adaptation for observing a space-bounded Repetitive Dynamic Event (RDE). The notion of “space-bounded RDEs” refers to a repetitive (re-occurring) physical phenomenon (either natural or artificial) that follows specific dynamics in each one of its occurrences, with its position being bounded in space, e.g. the repetitive triggering of a mechanical experiment, or the multiple demonstrations of a specific task by a human (in the context of Learning from Demonstrations).

Scene Perception and Event Comprehension Theory (SPECT), a notion introduced and actively being studied in cognitive science theory (Loschky et al. 2026), divides the human perception, seen as a dynamic event, in two different processes: The Front-end process and the Back-end process. Front-end processes occur during each eye fixation (information extraction, attentional selection) and the Back-end processes occur in memory (the current event model, the stored event model, prior knowledge, and executive processes). In robotic perception, following this line of thought, the pioneers of robotic Active Perception (or Active Vision) have approached the problem of Scene and Viewpoint Selection by adopting principles, similar to those described in SPECT (Bajcsy et al. 2016). They express the problem of Active Perception in a coordinate frame involving the following five axes: “Why”, “Where”, “What”, “When” and “How”. The “Why” element is associated mainly with the Back-end process of SPECT, while the “Where”, “What”, “When” and “How” are involved in the Front-end one. Focusing on Scene and Viewpoint Selection, they extensively present those concepts through the “What” and “Where” terms.

In recent decades, many researchers, working on robotic active perception, have been developing algorithms and frameworks that solve the Scene and Viewpoint Selection problem, based on similar principles with the above theories. A central notion,

that frequently gets involved on that, is the “Next Best View” (NBV), which was firstly introduced in the field of computer vision and robotics by C.Connolly (Connolly 1985). NBV-oriented solutions consider an active observer, i.e. a sensor capable to change its position and/or orientation in space, in order to perceive a static or dynamic target, with minimum uncertainty. Since the establishment of the NBV term, this idea has been applied in solving various problems. Among others, such applications are Scene Modeling/Understanding and Viewpoint selection in cluttered environment with ambiguities (Safronov et al. 2021; Eidenberger and Scharinger 2010), as well as in 3D reconstruction applications (Alsadik et al. 2025; Dunn et al. 2009; Rudnev et al. 2025) and exploration/optimal information gathering (He et al. 2025; Rasouli et al. 2020; Renz et al. 2025).

One of the most known and solid approaches to make a sensor active, in order to visually observe a target, is the visual servoing technique, which was introduced by Hill and Park (Hill and Park 1979) in 1979. Until then many variants have been developed (Chaumette and Hutchinson 2006, 2007) applied under different scopes (Spica et al. 2014; Xungao et al. 2018). However, one of the main requirements for observing a target, either a static or a dynamic one, is to maintain the perceivability of the observed system. For instance, Active Vision control schemes, which consider visual sensors, have to ensure that the target is within the Field-of-View (FOV) of the camera or generally within pre-specified bounds. To guarantee this, considering static (non-moving) targets, some methods employ the notion of Prescribed Performance Control (PPC) (Heshmati-Alamdari et al. 2014; Bechlioulis et al. 2019; Triantafyllou et al. 2017), while others attempt to guarantee the observability of the target-features using Deep Q-Networks (DQN) with IBVS schemes (Zhong et al. 2025), satisfy visual feature constraint tasks (Jiang et al. 2023), or through path planning (Zhou et al. 2025). One of the latest prominent works in Active Perception related literature is the so-called Perspective-1-Ellipsoid (P1E) method (Gaudillière et al. 2023), in which an ellipsoid-based theoretical framework is introduced, in the context of camera pose estimation (i.e. best view) for optimally perceiving a static object. Although the above works consider a static target, there are also works that consider moving/dynamic targets, mainly found in Unmanned Aerial Vehicle (UAV) related literature. For instance, in (Chen et al. 2023) a velocity estimator is employed for a tracking and grasping a moving target using an UAV, while in (Chen et al. 2021) a Model Predictive Control scheme is proposed using Unscented Kalman Filtering. However, these works either consider some knowledge of the dynamics of the observed phenomenon, or they utilize specific hardware, such as multiple cameras. Furthermore, current methods do not account for the minimization of the effort of the active sensor (i.e. its motion), i.e. they do not examine the possibility of finding a view point from which the whole behavior of the body may be visible.

In this work we propose a novel framework for autonomous Viewpoint and Scene Selection to actively perceive (possibly chaotic) space-bounded RDEs with minimum uncertainty. The proposed method, drawing inspiration from the ellipse approximation of P1E (Gaudillière et al. 2023) and also from partitioning a visual task (Corke and Hutchinson 2001), statistically estimates the bounds of the spatial characteristics of the motion of the target using an ellipsoid (a level set of a Gaussian distribution).

One novel aspect of the proposed solution is that it does not rely on any knowledge of the dynamics of the observed system or on estimates of its velocity. In particular, the proposed method can potentially serve as an enhancement of any Active Perception method, as it is applied in the Null-space of the main visual servoing task. Such null-space enhancement approaches have been implemented in previous works but under the scope of avoiding joint limits or visual occlusions, for instantaneously keeping the desired object inside the field of view (Ma and Hutchinson 2010), for maximizing manipulability (Hernandez-Barragan et al. 2024), or even for achieving a compliant robot behavior (Li et al. 2024). In contrast, the objective of this work is to investigate the extent to which such an enhancement can compensate for the perception gaps left by the primary visual task, ultimately leading to a more optimal perception strategy. The method is based on an evolving estimate of the envelope of the space-bounded RDE, i.e. the region of motion of the target, through a level set of a Gaussian distribution and it is applied in the null-space of the main visual servoing task, e.g. PPC-based visual servoing with FOV constraints (Triantafyllou et al. 2017; Bechlioulis et al. 2019). The proposed solution significantly differentiates from our previous work (Papageorgiou et al. 2024, 2025; Efstathopoulos et al. 2025; Efstathopoulos and Papageorgiou 2025), in which a model for the observed system was considered, e.g. through Dynamic Movement Primitives (mainly in the context of Learning by Demonstration).

2 Problem Description and Concept Solution

Consider the random motion of a Point of Interest (POI), which could possibly be a recurring phenomenon, as shown in Fig. 1. Let $\mathbf{p}(t) \in \mathbb{R}^3$ be the position of the POI at each time instance $t \in \mathbb{R}_{\geq 0}$ and let us draw the following assumptions:

- **Assumption 1:** The time evolution of $\mathbf{p}(t)$ is confined within an unknown bounded region $\mathcal{S} \subset \mathbb{R}^3$, the so called Region of Motion (ROM), and therefore, by assumption, $\mathbf{p}(t) \in \mathcal{S}$ for all $t \in \mathbb{R}_{\geq 0}$. This is true for each one of the possible re-occurrences of the phenomenon.
- **Assumption 2:** The current position of the POI, i.e. $\mathbf{p}(t)$, is measurable, as long as it is within the Field of View (FOV) of the camera.
- **Assumption 3:** The velocity of the POI, i.e. $\dot{\mathbf{p}}(t)$, is not known, or measurable/predictable; e.g. its motion could potentially be induced by the evolution of a chaotic dynamical system with bounded state.

Furthermore, consider the availability of a moving active sensor, e.g. a robot with an in-hand camera, which is able of perceiving the moving body. Let $\mathbf{p}_c(t) \in \mathbb{R}^3$ and $\mathbf{R}_c(t) \in SO(3)$ be the position and orientation, in a Rotation matrix form, of the sensor respectively at each time instance $t \in \mathbb{R}_{\geq 0}$. For the rest of the document, a camera will be considered as the sensor, for simplicity. Let $\mathcal{S}_c \subset \mathbb{R}^2$ be the projection of $\mathcal{S}$ to the pixel space of the camera, based on the pin-hole projection, which is provided, as a mathematical background, in the next section.

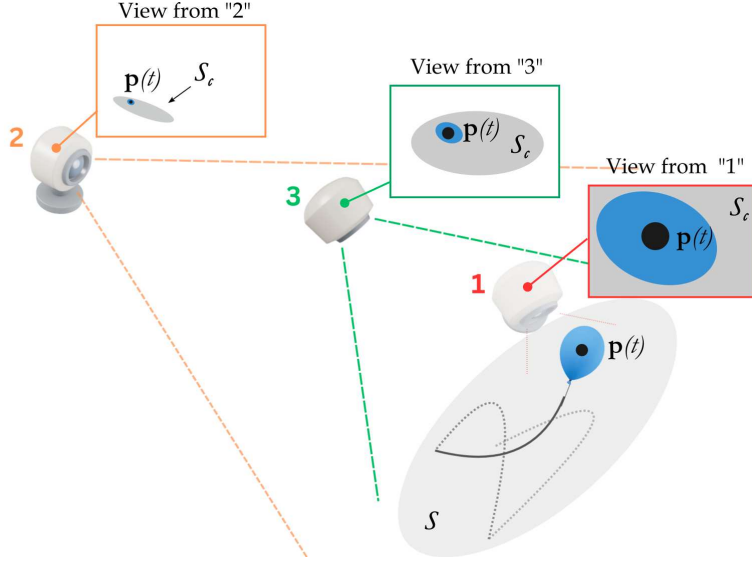


Fig. 1: The problem of observing a moving body.

Remark 1 Notice that due to Assumptions 2 and 3, an analytic visual-servoing (or FOV constraint) control law with a velocity feed-forward term is not considered to be directly applicable in this case.

Our objective is to on-line adjust the pose (i.e., its position and orientation) of the camera for optimally perceiving the whole behavior of the POI, without *a priori* knowing the characteristics of the motion. In other words, our objectives are the following:

- The position of the POI, i.e. $\mathbf{p}(t)$, has to remain (instantaneously) within the FOV of the camera. This is considered to be the primary objective.
- The Region of Motion (ROM) of the POI, i.e. $\mathcal{S}$, has to remain, as much as possible, within the FOV of the camera. This will ensure that the body will be perceivable, even though its velocity is not measurable or predictable, minimizing the required motion/effort of the observer. However, this is considered to be, hierarchically, a secondary objective.
- The projected region of motion of the POI, i.e. $\mathcal{S}_c$, has to occupy a substantial predefined area on the pixel space, for ensuring a high quality of observation and perception. This is also considered to be a secondary objective.

For making the problem more clear, let us discuss the example poses presented in Fig. 1. From the first camera pose, although the POI is instantaneously perceivable and within the FOV of the camera, practically, an abrupt motion of the body can practically lead it, in the close future, outside of the FOV, as due to the absence of the velocity measurement (of the POI) one can not exploit feed-forward terms for tracking the body (see Remark 1). On the other hand, from the second pose,

although the whole ROM, i.e. $\mathcal{S}$, is visible, the area occupied by $\mathcal{S}_c$ on the pixel space is relatively small, thus resulting in a low quality of observation, due to the inherent sensor's limitation in precision. Lastly, from the third pose, it is clearly visible that the above objectives are both satisfied as much as possible, resulting in both enhanced observability, while exploiting the precision capabilities of the camera. Our objective is to design a control law that converges to the latter camera pose, from which the camera is able to perceive the whole region of motion of the POI, with minimal effort, but also exploit the accuracy of the sensor.

2.1 Concept solution

To fulfill the above objectives, we design a control law that is capable of observing both the motion and the whole evolution of the event, with the smallest possible uncertainty. In particular, the proposed control scheme exploits an on-line estimation of ROM, i.e. an estimate of $\partial\mathcal{S}$, based only on the position measurements, to enhance active perception, which is applied in the null-space of the actual observation task. In that context, the basic idea of the proposed control framework is to keep within the FOV both the POI (primary task) and the estimated region of motion (i.e. an estimate of $\mathcal{S}$), as the motion is by assumption bounded in space (see Assumption 1). This, secondary task, which constitutes essentially the contribution of this work, exploits the statistical features of $\mathbf{p}(t)$, in order to keep track of the scene's event and converge to the suitable sensor's pose, aiming towards an optimal active perception. Conceptually, the controller proposed in the next sections combines and connects the aspects of “How”, “What”, and “Where”, as discussed in (Bajcsy et al. 2016).

3 Preliminaries and Mathematical Background

In this section, the mathematical background related to the Image Based Visual Servoing (IBVS) is briefly presented, as well as the Van der Merwe's Scaled Unscented Transform (SUT) algorithm, which is employed in the proposed method to project the characteristics of the motion from the 3D space to the visual plane.

3.1 Preliminaries on Image Based Visual Servoing (IBVS)

We consider the availability of an RGB camera with a depth sensor, that is mounted on a robot, e.g. an eye-in-hand setup, with at least six degrees of freedom that accepts commanded velocities. Let $\mathbf{p}(t) \triangleq [p_x \ p_y \ p_z]^\top \in \mathbb{R}^3$ be an arbitrary observed 3D point with respect to the camera frame. As we consider the availability of a depth sensor, $p_z \in \mathbb{R}_{\geq 0}$ is considered to be directly available. According to the pin-hole principle, any 3D point within the FOV of the camera is mapped to the visual plane (i.e. the pixel space), as follows:

$$\mathbf{s} = \mathbf{f}(\mathbf{p}) \triangleq \begin{bmatrix} \frac{f_x p_x}{p_z} + c_x \\ \frac{f_y p_y}{p_z} + c_y \end{bmatrix} \in \mathbb{R}^2, \quad (1)$$

where $f_x, f_y \in \mathbb{R}_{>0}$ are the focal lengths of the camera in x and y directions respectively, $c_x, c_y \in \mathbb{N}_{>0}$ are the coordinates of the center of the camera sensor in pixels

and $\mathbf{s} \triangleq [s_x \ s_y]^\top \in \mathbb{R}^2$ is the projected point on the pixel space. The time derivative of (1) is the following:

$$\dot{\mathbf{s}} = \frac{\partial \mathbf{f}}{\partial \mathbf{p}} \dot{\mathbf{p}}, \quad (2)$$

where

$$\frac{\partial \mathbf{f}}{\partial \mathbf{p}} = \begin{bmatrix} \frac{f_x}{p_z} & 0 & -\frac{f_x p_x}{p_z^2} \\ 0 & \frac{f_y}{p_z} & -\frac{f_y p_y}{p_z^2} \end{bmatrix} \in \mathbb{R}^{2 \times 3} \quad (3)$$

is the Jacobian that maps the velocity of the point, i.e. $\dot{\mathbf{p}}$, to the rate of the projected point. Considering a static point and a moving camera, the relative velocity of the point with respect to the moving camera is:

$$\dot{\mathbf{p}} = [-\mathbf{I}_3 \ \mathbf{S}(\mathbf{p})] \mathbf{v}_c, \quad (4)$$

where $\mathbf{S}(\cdot) : \mathbb{R}^3 \rightarrow \mathbb{R}^{3 \times 3}$ is the skew symmetric matrix for a given vector, $\mathbf{I}_n \in \mathbb{R}^{n \times n}$ the identity matrix and $\mathbf{v}_c \triangleq [\dot{\mathbf{p}}^\top \ \boldsymbol{\omega}_c^\top]^\top \in \mathbb{R}^6$ the generalized velocity of the camera consisting of the linear, i.e. $\dot{\mathbf{p}}_c \in \mathbb{R}^3$, and angular velocity, i.e. $\boldsymbol{\omega}_c \in \mathbb{R}^3$, respectively. After substituting $\dot{\mathbf{p}}$ from (4) to (2), we get the mapping between the velocity of the camera and the rate of change of the pixel coordinates of the point, which is the following:

$$\dot{\mathbf{s}} = \mathbf{L}(\mathbf{s}, p_z) \mathbf{v}_c \quad (5)$$

with $\mathbf{L}(\mathbf{s}, p_z) \in \mathbb{R}^{2 \times 6}$ being the following Jacobian matrix:

$$\mathbf{L}(\mathbf{s}, p_z) \triangleq \frac{\partial \mathbf{f}}{\partial \mathbf{p}} [\mathbf{I}_3 \ \mathbf{S}(\mathbf{p})] = \begin{bmatrix} -\frac{f_x}{p_z} & 0 & \frac{s_x - c_x}{p_z} & \frac{(s_x - c_x)(s_y - c_y)}{f_y} & -f_x - \frac{(s_x - c_x)^2}{f_x} & \frac{f_x}{f_y}(s_y - c_y) \\ 0 & -\frac{f_y}{p_z} & \frac{s_y - c_y}{p_z} & f_y + \frac{(s_y - c_y)^2}{f_y} & -\frac{(s_x - c_x)(s_y - c_y)}{f_x} & -\frac{f_y}{f_x}(s_x - c_x) \end{bmatrix}, \quad (6)$$

given that $p_x = \frac{p_z(s_x - c_x)}{f_x}$ and $p_y = \frac{p_z(s_y - c_y)}{f_y}$, from (1). Notice that $\mathbf{L}$ is not square and therefore it is not directly invertible. However, as the columns of $\mathbf{L}$ are more than its rows, the inverse mapping of (5) is a redundant problem, i.e. there are infinitely many possible solution for executing any arbitrary (i.e. commanded control) $\dot{\mathbf{s}}_c$ by moving the camera. That means that one can exploit the null-space of $\mathbf{L}$ in order to execute possibly any additional task after projecting it to the null-space. In particular, the general solution of (5) is the following:

$$\mathbf{v}_c = \mathbf{L}^\dagger(\mathbf{s}, p_z) \dot{\mathbf{s}}_c + \mathbf{N}_L(\mathbf{s}, p_z) \mathbf{v}_n, \quad (7)$$

where

$$\mathbf{N}_L(\mathbf{s}, p_z) = \mathbf{I}_6 - \mathbf{L}^\dagger(\mathbf{s}, p_z) \mathbf{L}(\mathbf{s}, p_z) \in \mathbb{R}^{6 \times 6} \quad (8)$$

is the null-space projection matrix, $\mathbf{L}^\dagger \in \mathbb{R}^{6 \times 2}$ is a pseudo-inverse of $\mathbf{L}$, for which it holds $\mathbf{L} \mathbf{L}^\dagger = \mathbf{I}_2$ and $\mathbf{v}_n \in \mathbb{R}^6$ any secondary velocity for the camera.

Remark 2 It is easy to prove, after multiplying each side of (7) with $\mathbf{L}(\mathbf{s}, p_z)$ from the left, that the velocity of the point in the pixel space, i.e. $\dot{\mathbf{s}}_c$, will remain unaffected by the second term of (7), in light of (8); as this calculation gives $\dot{\mathbf{s}} = \dot{\mathbf{s}}_c$. Therefore, this second term essentially constitutes a secondary task that one can exploit in order to optimize the observation. This is the case considered in this work.

Now, let us consider the visual servoing problem, in which the camera is actuated in order for the projected point, i.e. $\mathbf{s}$, to reach a desired pre-defined position on the pixel space. Let $\mathbf{s}^* \in \mathbb{R}^2$ be this desired target position in the pixel space. Based on the inverse mapping provided in (7), a simple proportional signal for achieving the aforementioned objective is the following:

$$\dot{\mathbf{s}}_c \triangleq -k_p(\mathbf{s} - \mathbf{s}^*), \quad (9)$$

where $k_p \in \mathbb{R}_{>0}$ is a positive tunable gain. One can easily prove the exponential convergence of $\mathbf{s}$ to $\mathbf{s}^*$ by multiplying each side of (7) with $\mathbf{L}$ from the left, after substituting $\dot{\mathbf{s}}_c$ from (9), which gives $\dot{\mathbf{s}} = -k_p(\mathbf{s} - \mathbf{s}^*)$.

Remark 3 The linear proportional control signal provided in (9) constitutes just a simple example. However, more sophisticated and probably non-linear control signals can be potentially utilized. For instance, for ensuring that the point will remain within the FOV or any prespecified region, one could enforce hard constraint by utilizing notions such as the Prescribed Performance Control (PPC) (Bechlioulis et al. 2019), or in general Barrier Artificial Potential (BAP) fields, such as the ones provided in (Kastritsi et al. 2018). Employing such methods can guarantee, in a continuous-time consideration, that the POI will in any case remain within the FOV, for bounded velocities of the target.

3.2 Preliminaries of Van der Merwe’s Scaled Unscented Transform (SUT) algorithm

In the proposed method, the region of the observed behavior, i.e. the ROM, is estimated/approximated, in space, through an ellipsoid constituting the level-set of a Gaussian distribution in the 3-dimensional space; more details are provided in the next section. For tackling the problem of projecting this ellipsoid to the visual plane, based on the non-linear projection dictated by the pin-hole principle (see (1)), we employ the Van der Merwe’s Scaled Unscented Transform (SUT) algorithm (Wan and Van Der Merwe 2000). Based on SUT, the ellipsoid is characterized using a deterministic set of sigma points, as well as an uncertainty regarding that, making the behavior variance able to be projected to the visual plane.

Let us study the case which is applied to this specific work, i.e. considering a motion of $\mathbf{p}(t)$ which is executed in $N = 3$ dimensions. According to the SUT, by definition, $M = 7$ (i.e., $2N + 1$) sigma points are generated, $\boldsymbol{\chi}^{(i)}(t) \in \mathbb{R}^3$, where $i \in \{0, \dots, 6\}$, that are computed at each time instance t as follows:

$$\chi^{(i)}(t) \triangleq \begin{cases} \mathbf{p}(t), & \text{for } i = 0, \\ \mathbf{p}(t) + \left(\sqrt{(N + \lambda)\mathbf{\Sigma}(t)} \right)_{|i}, & \text{for } i = 1, \dots, N, \\ \mathbf{p}(t) - \left(\sqrt{(N + \lambda)\mathbf{\Sigma}(t)} \right)_{|i-N}, & \text{for } i = N + 1, \dots, 2N, \end{cases} \quad (10)$$

where $(\mathbf{A})_{|i} : \mathbb{R}^{3 \times 3} \rightarrow \mathbb{R}^3$ is the i^{th} column of any matrix $\mathbf{A} \in \mathbb{R}^{3 \times 3}$, $\mathbf{\Sigma} \in S_{++}^3$ the positive definite variance-covariance matrix of the Gaussian distribution (in our work, of the one that characterizes the perceived behavior) and $\lambda \in \mathbb{R}_{>0}$ is the scaling parameter, which is calculated as follows:

$$\lambda = \alpha^2(N + \kappa) - N,$$

with $\alpha, \kappa \in \mathbb{R}_{>0}$ constant parameters that are used to control the projected sigma point spread and scaling, respectively. Notice that the square root of the matrix included in (10) can be computed through the Cholesky decomposition.

Given the sigma points calculated through (10), the projected (on the visual plane) sigma points, denoted with $\psi^{(i)} \in \mathbb{R}^2$, are calculated utilizing the pin-hole projection mapping, given in (1); namely $\psi^{(i)} = \mathbf{f}(\chi^{(i)})$, $\forall i = 0, \dots, 6$. These points are used to reconstruct the image Gaussian distribution using the specific weights of the SUT algorithm. In particular, the mean, $\mu_p(t) \in \mathbb{R}^2$, and the variance-covariance matrix, $\Sigma_p(t) \in S_{++}^2$, of the projected normal distribution are then calculated respectively by:

$$\begin{aligned} \mu_p(t) &= w_m \psi^{(0)} + \sum_{i=1}^{2N} w_i \psi^{(i)}, \\ \Sigma_p(t) &= w_c (\psi^{(0)} - \mu_p)(\psi^{(0)} - \mu_p)^\top + \sum_{i=1}^{2N} w_i (\psi^{(i)} - \mu_p)(\psi^{(i)} - \mu_p)^\top, \end{aligned} \quad (11)$$

where the sigma point weights are calculated as follows:

$$\begin{aligned} w_m &= \frac{\lambda}{N + \lambda}, \\ w_c &= \frac{\lambda}{N + \lambda} + (1 - \alpha^2) + \beta, \\ w_i &= \frac{1}{2(N + \lambda)}, \quad \text{for } i > 0. \end{aligned} \quad (12)$$

Definition 1 Let the SUT mapping of normal distributions, described by (10)-(12), be symbolized, for compactness, as follows:

$$\mathcal{N}(\mu_p, \Sigma_p) = \mathbf{g}(\mathcal{N}(\mu_S, \Sigma_S)), \quad (13)$$

where $\mathbf{g}(\cdot) : \mathbb{R}^3 \times S_{++}^3 \rightarrow \mathbb{R}^2 \times S_{++}^2$ the above SUT-based projection of the 3D distribution to the 2D pixel space of the camera.

4 Proposed method

The proposed method is composed of two control parts, the one that involves the primary control signal, i.e. $\dot{\mathbf{s}}_c$, and the one that involves the secondary (enhancement) signal, i.e. $\mathbf{v}_n$, which are applied to the two terms of the right-hand side of (7) respectively. A block diagram of the proposed method is shown in Fig. 2. For the primary control signal we consider the gradient of a Barrier Artificial Potential (BAP), in order to ensure that the moving point, i.e. $\mathbf{p}(t)$, will remain instantaneously within the field of view and therefore instantaneously detectable. For the secondary signal, we propose the use of an estimator for the Region of Motion (ROM) of the POI, i.e. $\mathcal{S}$, that is modeled through an ellipsoid in 3D space, namely a level-set of a Gaussian distribution. All these components of the proposed method are presented in the next subsections.

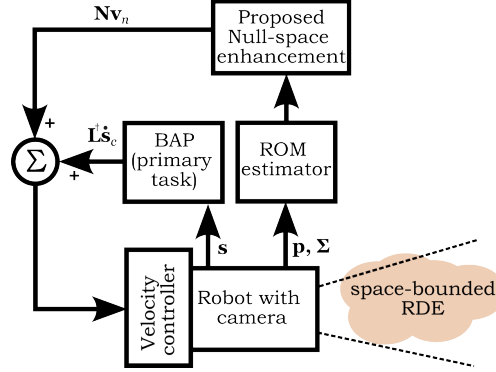


Fig. 2: Block diagram of the proposed null-space enhancement

4.1 Instantaneous FOV-ensuring control (primary task)

For ensuring that the POI will instantaneously remain within the FOV, we propose a control signal that is based on a Barrier Artificial Potential (BAP) field, in the pixel space. The employed BAP is designed such that no motion is generated for the camera when the POI lies within a prespecified safe region, while, on the other hand, it tends to infinity as the POI approaches the boundary of the image imposed by the FOV constraints, as depicted in Fig. 3a. More specifically, the primary control signal is selected to be the following:

$$\dot{\mathbf{s}}_c \triangleq - \left(\frac{\partial V(\mathbf{s})}{\partial \mathbf{s}} \right)^\top, \quad (14)$$

where $V(\mathbf{s}) : \mathbb{R}^2 \rightarrow \mathbb{R}_{\geq 0}$ the BAP function. In this work we employ the Coulomb-forces inspired BAP (Kastritsi et al. 2018), which is applied as follows:

$$V(\mathbf{s}) \triangleq U(s_x; \underline{s}_x) + U(s_x; \bar{s}_x) + U(s_y; \underline{s}_y) + U(s_y; \bar{s}_y), \quad (15)$$

where $\underline{s}_x, \underline{s}_y, \bar{s}_x, \bar{s}_y \in \mathbb{R}_{\geq 0}$ the lower and higher bounds of the image in the pixel space along the x and y directions respectively, $d_0 \in \mathbb{R}_{>0}$ is a tunable parameter defining the safe region within the boundary of the image (see Fig. 3a) and

$$U(x; x_{lim}) \triangleq \begin{cases} 0, & \text{if } |x - x_{lim}| \geq d_0 \\ \frac{k_1}{2} \left(\frac{1}{|x - x_{lim}|} - \frac{1}{d_0} \right)^2, & \text{otherwise} \end{cases} \quad (16)$$

is the utilized artificial potential function for any scalar $x \in \mathbb{R}$ for a given scalar constraint $x_{lim} \in \mathbb{R}$, depicted in Fig. 3b, with $k_1 \in \mathbb{R}_{>0}$ being the tunable gain of the control signal.

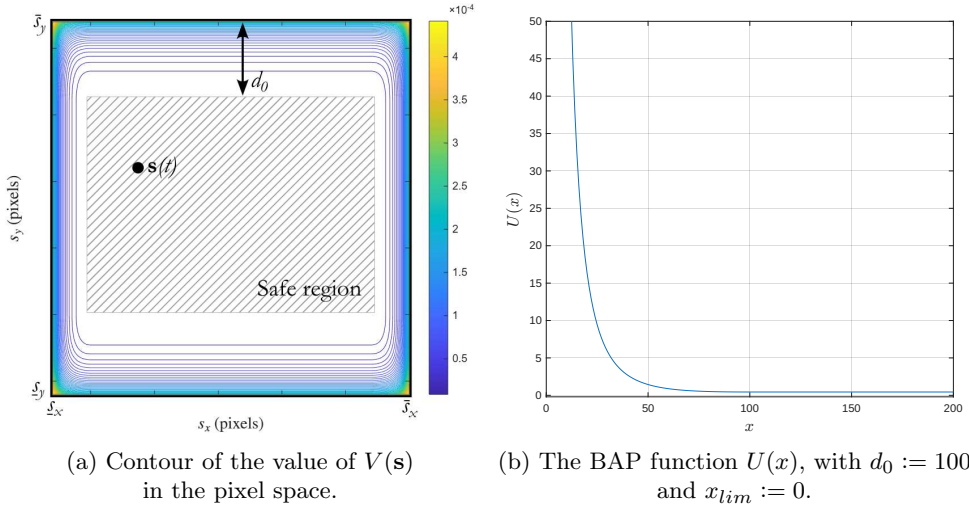


Fig. 3: Design of the Barrier Artificial Potential function

Remark 4 Notice that the selected artificial potential function has the following properties:

- $V = 0$ and $\left(\frac{\partial V}{\partial \mathbf{s}}\right)^\top = \mathbf{0}_2$ for all $\mathbf{s}$ that are within the safe region, i.e. for all $\mathbf{s}$ such that $\underline{s}_x + d_0 < s_x < \bar{s}_x - d_0$ and $\underline{s}_y + d_0 < s_y < \bar{s}_y - d_0$. That means that no motion will be induced for the camera, if the POI remains away from the boundary of the image (defined by the FOV).
- U and its derivative will tend to ∞ (or $-\infty$, depending on the direction of the constraint) as x approaches the boundary. This can ensure that the POI will remain within the FOV, for motions with bounded velocities (see (Kastritsi et al. 2018) or similarly (Heshmati-Alamdari et al. 2014), which employs the notion of PPC).

4.2 Estimating the Region of Motion of the POI

The estimation of the region of the observed behavior, i.e. the ROM, which plays an essential role on the design of the secondary control signal (described in the next subsection), is performed in real time given every new measurement of the sensor. As we consider a camera, we assume a fixed frame rate of the incoming images, and therefore the analysis in this subsection will consider discrete time, indexed through $k \in \mathbb{N}$. Let $\mathbf{p}_k \in \mathbb{R}^3$ be the position of the POI, with respect to the inertial world frame $\{0\}$, at the k -th step. For estimating the ROM of the observed behavior, one could store the whole history of motion from its beginning and utilize classical statistical methods. However, our aim is to avoid such a procedure, as it suffers from high memory and computational requirements; in fact it tends to infinity, as the time passes. Therefore, in the proposed method, we recursively update an evolving Gaussian representation of the POI's trajectory. To this aim, let $\boldsymbol{\mu}_{\mathcal{S},k} \in \mathbb{R}^3$ and $\boldsymbol{\Sigma}_{\mathcal{S},k} \in S_{++}^3$ be the mean and covariance matrix of the evolving Gaussian distribution, reflecting the statistics of the kinematic behavior of the body. The recursive update laws that approximates the features of $\mathcal{S}$ are the following:

$$\begin{aligned}\boldsymbol{\mu}_{\mathcal{S},k} &= (1-b)\boldsymbol{\mu}_{\mathcal{S},k-1} + b\mathbf{p}_k, \\ \boldsymbol{\Sigma}_{\mathcal{S},k} &= (1-b)\boldsymbol{\Sigma}_{\mathcal{S},k-1} + b\mathbf{d}_k\mathbf{d}_k^\top\end{aligned}\tag{17}$$

where $b \in \mathbb{R}_{>0}$ is a scalar forgetting factor and $\mathbf{d}_k \triangleq \mathbf{p}_k - \boldsymbol{\mu}_{\mathcal{S},k-1}$ is the displacement between the current measurement and the previous Gaussian center.

As a result, the evolving covariance, i.e., $\boldsymbol{\Sigma}_{\mathcal{S},k}$, captures the motion tendency of the target over time at a computational cost of $O(1)$. The evolving ellipsoid, which constitutes an estimate of the ROM, i.e. of $\mathcal{S}$, is therefore represented by the level-set of the Gaussian distribution as follows:

$$\hat{\mathcal{S}} = \{\mathbf{y} \in \mathbb{R}^3 : \mathbf{y}^\top \boldsymbol{\Sigma}_{\mathcal{S},k}^{-1} \mathbf{y} \leq 1\}.\tag{18}$$

The evolving distribution is projected into the image plane using the Van der Merwe's Scaled Unscented Transform (SUT) algorithm, presented in Section 3, as follows:

$$\mathcal{N}(\boldsymbol{\mu}_{p,k}, \boldsymbol{\Sigma}_{p,k}) = \mathfrak{g}(\mathcal{N}(\boldsymbol{\mu}_{\mathcal{S},k}, \boldsymbol{\Sigma}_{\mathcal{S},k})),\tag{19}$$

resulting in the 2D distribution that approximates the ROM in the pixel space, characterized by the center $\boldsymbol{\mu}_{p,k} \in \mathbb{R}^2$ and variance $\boldsymbol{\Sigma}_{p,k} \in S_{++}^2$. The level-set of this distribution, considered in this work to approximate the boundary of the projected ROM, is the following:

$$\mathcal{H}(t) \triangleq \{\mathbf{s} \in \mathbb{R}^2 : \mathbf{s}^\top \boldsymbol{\Sigma}_p^{-1}(t) \mathbf{s}_p = 1\}.\tag{20}$$

4.3 Null-space enhancement (secondary task)

The secondary task involves the motion of the camera in the null-space of the primary task, for optimally perceiving the whole kinematic behavior, based on the estimate

described in the previous subsection. In other words, our aim is to move the camera such that the ellipsoid approximating the ROM, i.e. $\mathcal{H}(t)$ which is characterized by $\boldsymbol{\mu}_p$ and $\boldsymbol{\Sigma}_p$, approaches a specific (substantial) predefined size and a specific pre-defined angle, based on the geometry of the FOV of the sensor, as show in Fig. 4. To achieve this, we aim at designing a control signal that moves the camera such that the major axis of the projected ROM ellipsoid, i.e., $\mathcal{H}(t)$, reaches a predefined desired value. Let $\theta_M^* \in \mathbb{R}$ and $a_M^* \in \mathbb{R}_{>0}$ be the angle and magnitude (polar coordinates) of this predefined desired major axis respectively, denoted in blue in Fig. 4. Furthermore, let $\theta_M(t) \in \mathbb{R}$ and $a_M(t) \in \mathbb{R}_{>0}$ be the actual current angle and magnitude of the major axis of $\mathcal{H}(t)$ respectively. Notice that the polar coordinates of the major axis of $\mathcal{H}(t)$ can be found through the eigen-decomposition of $\boldsymbol{\Sigma}_p$, i.e. by finding the major eigen vector and its corresponding eigenvalue.

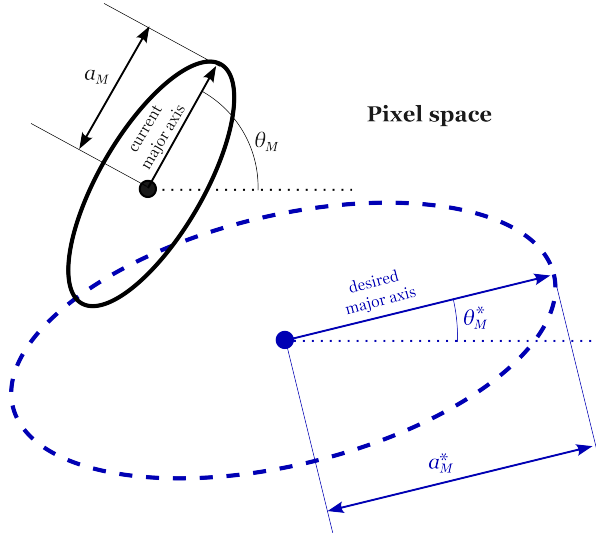


Fig. 4: Objective of the secondary control signal

Remark 5 It is worth noting that, since an ellipse (in general) is invariant under rotations of π , the angle, i.e. θ_M , is wrapped into the interval $[-\pi/2, \pi/2]$, yielding the minimum angular deviation from the desired (horizontal) orientation.

Given the above specific control objective of the secondary task, we propose the following control signal, that acts on the second term of (7):

$$\mathbf{v}_n \triangleq -\mathbf{J}_s \mathbf{K} \mathbf{e}_p, \quad (21)$$

where $\mathbf{K} \in \mathbb{R}^2$ is a diagonal matrix of tunable gains, $\mathbf{e}_p \in \mathbb{R}^2$ is the error between the current and desired major axis in polar coordinates, defined as:

$$\mathbf{e}_p \triangleq \begin{bmatrix} a_M - a_M^* \\ \theta_M - \theta_M^* \end{bmatrix}, \quad (22)$$

and $\mathbf{J}_s \in \mathbb{R}^{6 \times 2}$ is the “screw-motion” Jacobian, defined as follows:

$$\mathbf{J}_s \triangleq \begin{bmatrix} \mathbf{z} & \mathbf{0}_{3 \times 1} \\ \mathbf{0}_{3 \times 1} & \mathbf{z} \end{bmatrix}, \quad (23)$$

where $\mathbf{z} \in \mathbb{R}^3$ the depth direction of the camera, which, in most of the cases, is $\bar{\mathbf{z}} = [0 \ 0 \ 1]^\top$.

Remark 6 Notice that the secondary control signal, provided in (21), is filtered through the “screw-motion” Jacobian, which means that the dictated motion will always perform a translation and a rotation along the axis of depth of the camera. This is derived from the principle of “perspective”, according to which the image of any object is enlarged when the object is approached by the camera, while a rotation around the same axis will also mean a rotation of the image around the axis orthogonal to the image plane.

The overall proposed solution, involving both the primary and the secondary task, is summarized in Algorithm 1 for technical clarity.

Algorithm 1 Proposed solution

- 1: **Inputs:** $\boldsymbol{\mu}_{S,0}, \boldsymbol{\Sigma}_{S,0}$ ▷ Initialization
 - 2: Select values for: $\alpha, \beta, \kappa, d_0, b, a_M^*, \theta_M^*, k_1, \mathbf{K}$ ▷ Parameter selection
 - 3: **for** each k **do**
 - 4: Get the new frame from the camera
 - 5: Detect POI and get $\mathbf{p}_k$ and $\mathbf{s}_k$
 - 6: Compute $\dot{\mathbf{s}}_c$ from (14), based on $\mathbf{s}_k$ ▷ Primary task
 - 7: Update $\mathcal{N}(\boldsymbol{\mu}_{S,k}, \boldsymbol{\Sigma}_{S,k})$ using (17), based on $\mathbf{p}_k$
 - 8: Utilize the projection of (19) to get $\mathcal{N}(\boldsymbol{\mu}_{p,k}, \boldsymbol{\Sigma}_{p,k})$
 - 9: Calculate a_M, θ_M through the eigen-decomposition of $\boldsymbol{\Sigma}_{p,k}$
 - 10: Calculate $\mathbf{v}_n$ from (21) ▷ Secondary task
 - 11: Apply $\dot{\mathbf{s}}_c$ and $\mathbf{v}_n$ to (7) to get $\mathbf{v}_c$ ▷ Merge tasks
 - 12: Send $\mathbf{v}_c$ as commanded velocity for the camera attached to the robot
 - 13: **end for**
-

5 Experimental Evaluation

The proposed solution has been evaluated in two different perception scenarios, a task in which the actor is a robot whose motion is generated on-line through a linear under-damped dynamical system (coined from now on as: the "robot-actor scenario") and a task in which the actor is a human performing a wood sanding task (coined from now on as: the "human-actor scenario"). As mentioned above, the core idea is the on-line update of the view pose of the camera, while observing a randomly evolving event with unmodeled dynamics, aiming at perceiving the space-bounded RDE with minimum uncertainty. Let us highlight that the study starts from the least (robot-actor scenario) to the most unpredictable (human-actor) dynamic event. Linking them with real-life use case scenarios, such applications could be monitoring robot activity (e.g. drones) for surveillance applications, a re-occurring mechanical experiment, monitoring tasks performed by humans in order to parse knowledge from experienced workers to robots or to evaluate the performance or workload of the worker, or even monitoring live-stock. The proposed method is compared to the case in which only the primary task is utilized without the application of the secondary enhancement, i.e. the utilization of a Barrier Artificial Potential for maintaining the POI instantaneously within the FOV of the camera.

For the experimental evaluation, the parameters selected are the following: $d_0 = 300$ pixels, $k_1 = 10^6$, $b = 0.005$, $\theta_M^* = 0$ rad, $a_M^* = 200$ pixels, $\alpha = 1$, $\beta = 2$, $\kappa = 0$ and $\mathbf{K} = 2\mathbf{I}_2$, while the resolution of the image was selected to be 640x480, with a frame rate of 30 fps. Given the above, Fig. 5 depicts the meaning of the selected values for d_0 , θ_M^* and a_M^* .

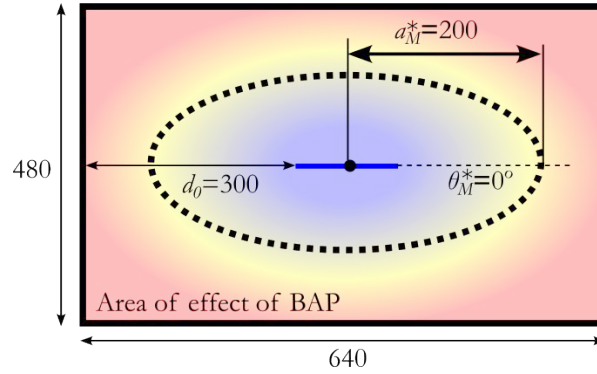


Fig. 5: The meaning of the values selected for parameters d_0 , θ_M^* and a_M^* . The dashed line provides an image of the desired ROM ellipsoid in the pixel space of the camera.

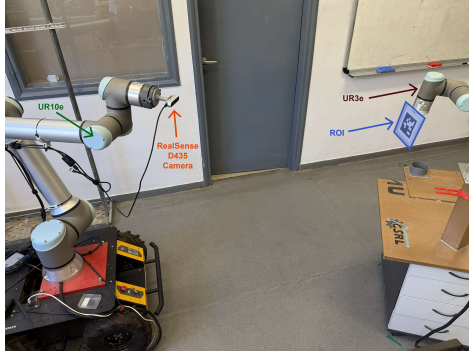
5.1 Robot-actor scenario

The goal of using a robotic actor to emulate the repetitive dynamic event is to fairly compare the proposed control scheme with other approaches, as it enables the reproduction of the experiment under the exact same conditions in each iteration. For the needs of this experiment, a UR3e robot is utilized with an Apriltag attached to its end-effector, while the control scheme is implemented using a UR10e robotic manipulator with an eye-in-hand Realsense D435 RGB-D camera, as the active observer, as depicted in Fig. 6a. For detecting the Apriltag, the dedicated algorithm from pupil Apriltag library is used, with its center being considered as the POI. Let us highlight that no information about the proprioceptively calculated position of the end-effector of the robot-actor is passed from the one robot to the other and the exact dynamics of the event is considered to be completely unknown to the observer.

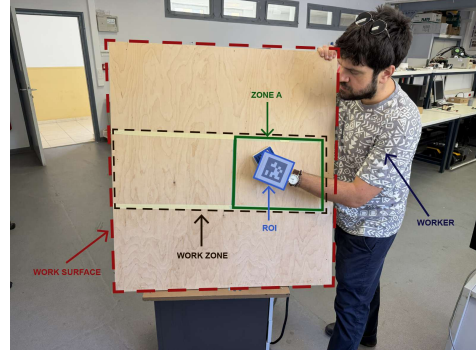
The motion trajectory for the end-effector of the UR3e robot (i.e the emulated RDE), denoted as $\mathbf{x}_1(t) \in \mathbb{R}^3$, is on-line generated through the following second order dynamical system:

$$\begin{bmatrix} \dot{\mathbf{x}}_1 \\ \dot{\mathbf{x}}_2 \end{bmatrix} = \sigma \begin{bmatrix} \mathbf{x}_2 \\ -\omega_n^2(\mathbf{x}_1 - \mathbf{p}_e) - 2\zeta\omega_n\mathbf{x}_2 \end{bmatrix}, \quad (24)$$

where $\zeta = 0.008$, $\omega_n = 0.31 \text{ rad/s}$ the damping ratio and the natural frequency of the emulated system respectively, $\sigma = 6$ the time-scaling parameter, and $\mathbf{p}_e \in \mathbb{R}^3$ the attractor of the dynamical system. Notice that the above specific selection yields an under-damped system, which asymptotically converges to $[\mathbf{p}_e^\top \mathbf{0}_3^\top]^\top$, having an oscillatory transient, i.e. emulating a mass-spring-damper system. Further, notice that the initial conditions of the system, i.e. $\mathbf{x}_1(0), \mathbf{x}_2(0)$ plays a significant role on the exact envelope within which the total motion is confined, i.e. $\mathcal{S}$. In this experiment, two different initial conditions were selected, namely the one with $\mathbf{x}_1(0) \triangleq \mathbf{p}_e - [0.3 \ 0 \ 0]^\top m$ and the other with $\mathbf{x}_1(0) \triangleq \mathbf{p}_e - [0.17 \ 0 \ -0.17]^\top m$, both with $\mathbf{x}_2(0) \triangleq [0 \ 0 \ 0.08]^\top m/s$, which results on a horizontal and a tilted by 45° ellipsoid for $\mathcal{S}$ respectively, in order to emulate and test two different cases.



(a) Robot-actor scenario.



(b) Human-actor scenario.

Fig. 6: Experimental scenarios.

5.2 Human-actor Scenario

As mentioned before, regarding the evaluation part, the intention was to gradually move from the relatively more predictable observation scenario to the completely unpredictable one. To this end, a human demonstration scenario for surface processing was also considered. The human acts as an experienced worker who performs a sanding task on a surface. For detecting the human hand, an Apriltag is attached to the worker’s hand, which is considered to be the POI (see Fig. 6b). The total work area (denoted as “Work Zone” in Fig. 6b) includes a smaller region (denoted as “Zone A” in Fig. 6b) in which the human focuses, as considered for the sake of the experiment, at the beginning and at the ending of the experiment, in order to perform detailed sanding. Notice that the areas considered are of two different sizes, in order to investigate the scenario of transitioning from a smaller $\mathcal{S}$ to a larger one and vice versa. In particular, the total duration of the experiment is selected to be one minute. The human starts by performing the sanding task, i.e. a repetitive space-bounded oscillating motion, within zone A for the first 15s, then the human expands the area of sanding by performing the task within the whole work zone, resulting in a larger $\mathcal{S}$, for another 15s, and finally he returns to zone A for the remaining experiment time (last 30s).

6 Results

Let us first investigate the behavior of the system under the proposed control scheme. Figure 7 depicts, in each sub-figure: a) the path followed by the POI (end-effector of the robot or human actor) on the left and b) the motion trajectory of the camera attached to the end-effector of UR10e on the right, with the proposed enhancement enabled (right column of Fig. 7) and without considering the proposed enhancement (left column of Fig. 7). Furthermore, snapshots of the experiments are provided in Fig. 8-10, along-side with the exact image from the camera, at some instances, without the proposed enhancement enabled (denoted in red) and with the proposed method (denoted in green). From the first and second row of Fig. 7 (robotic-actor cases) and Fig. 8-9, it is clearly visible that the observer with the proposed enhancement enabled, is initially driven backwards in order to ensure a clear view of the whole dynamic event (light grey in Fig. 7), but after this initial transient it is gradually driven towards a closer look of the event (dark grey and black in Fig. 7), as RDE’s oscillations are gradually diminished, due to the considered dynamics of the RDE (dictated by (24)). On the contrary, when the proposed enhancement is not enabled, the motion of the observer has a constant tendency to move away (backwards), due to the fact that any desired motion towards the center of the image, on the pixel plane, will result generally in a backwards motion of the camera, due to nature of the inverse of $\mathbf{L}$. Regarding the human-actor scenario, it is clearly visible from the last row of Fig. 7 and Fig. 10, that the active observer under the proposed control scheme, initially (for the first 15s) maintains a close view of the event, as the human is only moving within zone A. Afterwards, when the human expands his motion to the whole work zone, the system automatically detects that the motion envelope is enlarged, which consequently yields an abrupt backward motion in order to have a clearer and more broad view of the

event. Lastly, when the human returns working into zone A (final 30s), the observer returns to a more close view, by approaching again towards the location of the RDE. The latter observations, clearly shows how the proposed method adjusts the view-point of the observation, without however affecting the main visual servoing task, but also enhancing it, as aslo shown below.

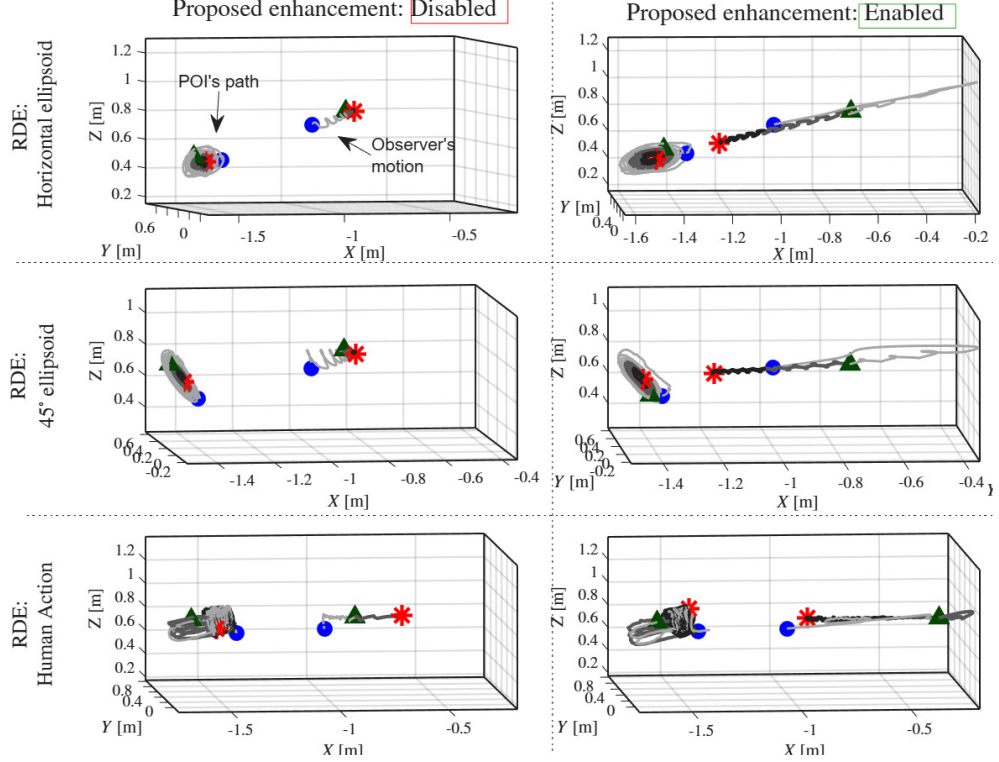


Fig. 7: The motion of the POI and the active observer. The color gradient of the path symbolizes the passing of time, starting from grey and reaching black. Blue circle: progress=0%, Green triangle: progress= 40%, Red star: progress=75%.

The estimate of the region of motion in the pixel space, namely $\mathcal{H}(t)$, is also depicted, for each one of the cases, in Fig. 8-10 respectively, highlighted as a green ellipse upon the actual image taken from the camera. Notice how, in each case, the proposed method (green configuration and image) is able to quickly adjust the distance of the sensor, i.e. even before reaching 8% of the progress, in order to ensure a clear view of the event, and subsequently ensuring that the whole $\hat{\mathcal{S}}$ lies within the field of view. Furthermore, notice, in Fig. 9, how the active observer, with the proposed enhancement enabled, aligns the major axis of the tilted by 45° ellipsoid with the x -axis of the camera, therefore fulfilling the defined control objectives, as compared to



Fig. 8: RDE: Robotic actor with horizontal ellipsoid. Red: without the proposed enhancement, Green: with the proposed enhancement enabled.

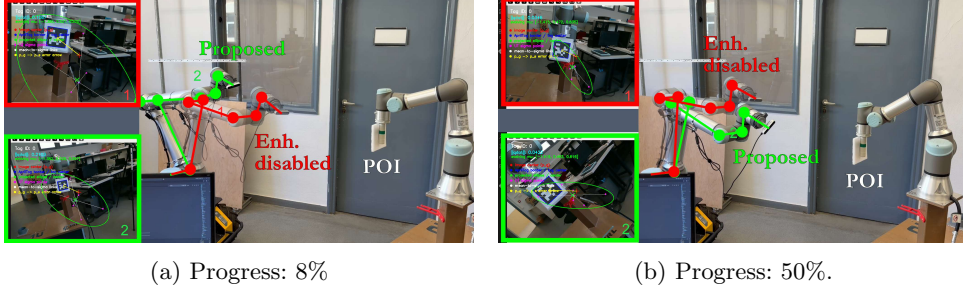


Fig. 9: RDE: Robotic actor with 45° ellipsoid. The same color code with Fig. 8 is followed.

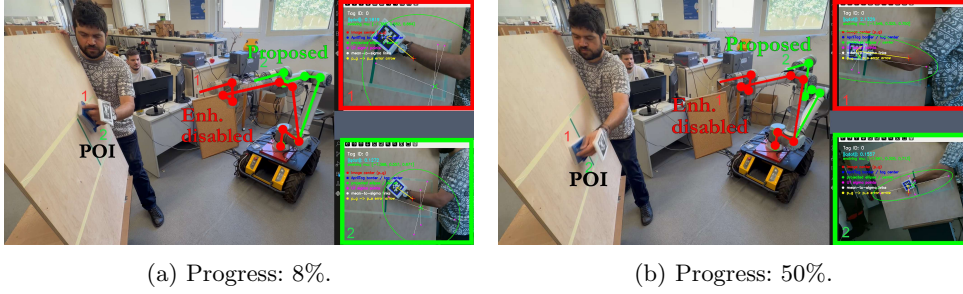


Fig. 10: RDE: Human actor. The same color code with Fig. 8 is followed.

the case in which the proposed enhancement is not enabled, which does not consider this aspect.

In Fig. 11, the trace of the time evolution of the estimated ellipsoidal RDE envelope, i.e. $\mathcal{H}(t)$, is depicted in purple color, as well as its center, i.e. $\mu_p(t)$, in the pixel space, as perceived by the in-hand camera, during its motion. The color gradient of

Table 1: Percentage of visibility of the ROM ellipse in each experimental trial.

Metrics	Proposed enhancement: Disabled			Proposed enhancement: Enabled		
	0° RDE	45° RDE	Human	0° RDE	45° RDE	Human
Center inside FOV	97.5%	98%	98.2%	100%	100%	100%
Full ellipse inside FOV	75.2%	66.6%	28.2%	96.5%	94.5%	85.8%

these ellipses symbolizes the passing of time, starting from light purple and reaching dark purple. Similarly to Fig. 7, the left column of Fig. 11 provides the results without the proposed enhancement, while its right column provides the results with the proposed enhancement enabled, in each one of the aforementioned experimental scenarios. Notice, in Fig. 11, that the initial RDE envelope in the pixel space, i.e. $\mathcal{H}(0)$, is relatively large, expanding beyond the pixel space, due to the arbitrary initialization of $\mu_{\mathcal{S},k}$, $\Sigma_{\mathcal{S},k}$ (see the first line of Algorithm 1), which however is adapted abruptly based on the next few frames. Analyzing the results at the sensor’s image plane, as depicted in Fig. 11, it is clear that when using the proposed solution, in all the experimental scenarios, the projected ellipse, i.e. $\mathcal{H}(t)$, and consequently its center, fall within sensor’s FOV, in contrast to case of not employing the proposed method (left column of Fig. 7), where there are many cases in which the system is not even able to keep the center of the ellipse inside the field of view, as it only relies on the instantaneous avoidance of the boundaries of the image. This is a significant achievement of the proposed approach, as without loss of generality, if the RDE is assumed to evolve primarily around the center of the corresponding ellipsoid, it is essential that this center remains within sensor’s FOV.

To quantify the above findings, we introduce two metrics in Table 1, provided in each one of the considered scenarios and for both control schemes. In the first row (“Center inside FOV”), we provide the percentage of time duration, during which the center of the estimated ROM, i.e. of $\mathcal{H}(t)$ was within the FOV, with respect to the total duration of the experiment, while in the second row (“Full ellipse inside FOV”) we provide the percentage of time duration in which the whole ellipse (i.e. $\mathcal{H}(t)$) was within the FOV, with respect to the total duration of the experiment. As shown in Table 1, it is clear that the proposed approach successfully (100%) manages to keep the projected center of the estimated ROM within camera’s FOV, even in the most unpredictable evaluation scenario, namely the “human-actor” one. Regarding the strict requirement presented in the second row of Table 1, one can observe the improvement achieved when the proposed enhancement is enabled, as compared to the case in which the proposed null-space signal is absent. In particular, in each one of the tested cases, the proposed method was able of significantly increasing the time in which the whole envelop of the kinematic behavior of the target lies within the FOV, with the most significant improvement being in the Human-actor scenario, in which this percentage was raised from 28% to 85%. The latter means that without the proposed enhancement one cannot reliably ensure that the kinematic behavior can be visible from the current view pose of the active observer, while when employing the proposed enhancement, this problem is significantly cured.

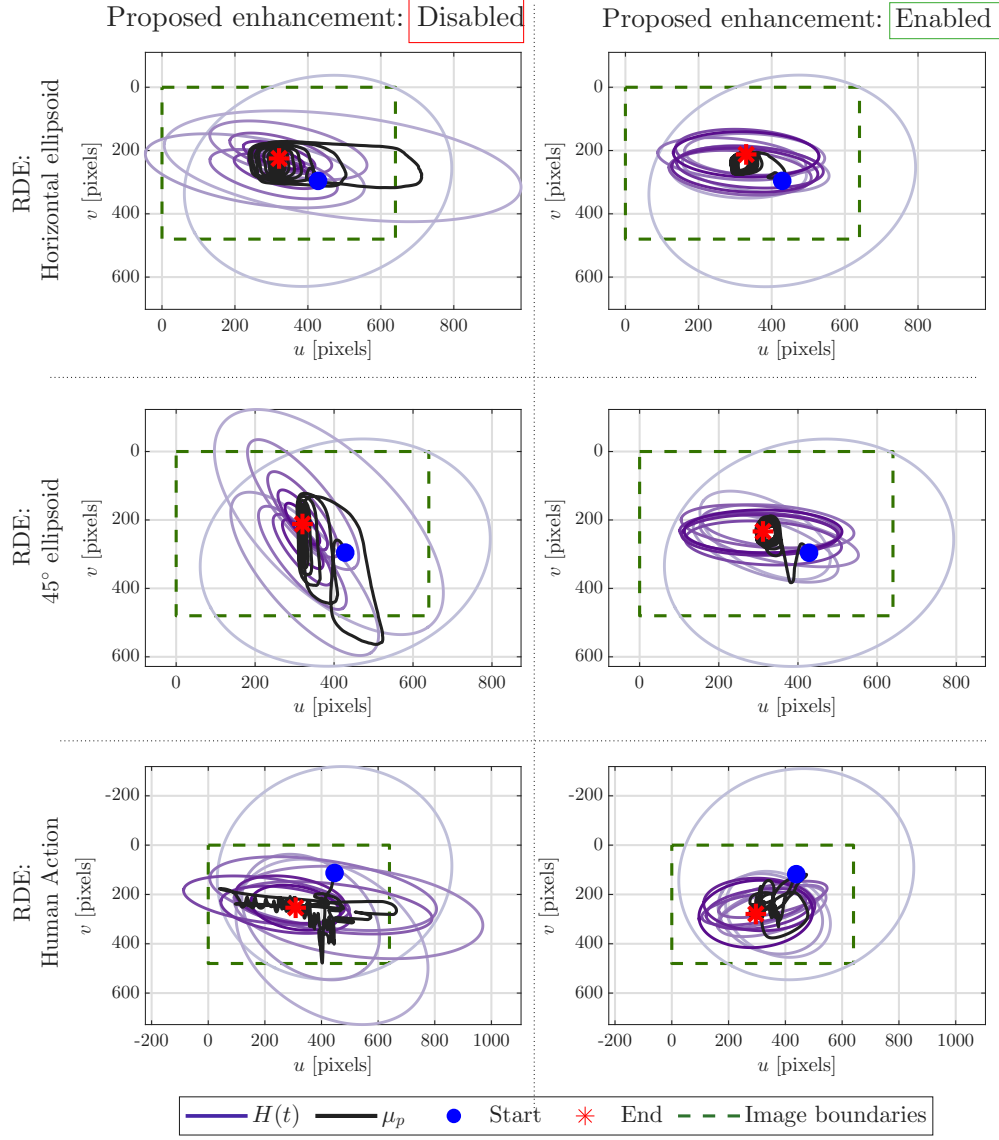


Fig. 11: The observed RDE in the pixel space. The purple ellipses symbolizes $\mathcal{H}(t)$ with its gradient denoting the passage of time, while the black line denotes the trajectory of its center.

7 Conclusion

This work introduces a null-space enhancement that can be applied in any active observation task, when tackling the problem of active perception of space-bounded

Repetitive Dynamic Events. Specifically, given that the primary objective of the observation task is to instantaneously keep the desired target’s motion within the camera’s Field of View, we incorporate additional and secondary perceptual objectives that do not require access to the full state of the observed system or state estimation. The additional control signal is able to maintain the whole kinematic behavior of the target within the Field of View of the camera, by estimating the region of motion of the target and exploiting the null-space of the main observation task. The experimental evaluation involves two scenarios, namely one scenario with a robotic actor emulating the RDE (involving two sub-cases) and one scenario with a human actor demonstrating a wood-sanding task in front of the robot observer. The experiments show that the proposed solution extends the observability of space-bounded RDEs, by driving the active observer towards view-poses that yields reliable and robust perception, even in abrupt changes of the target. The improvement achieved when the proposed method is utilized is shown to be significant in terms of maintaining the envelope of the whole kinematic behavior within the field of view of the camera.

Declarations

Competing interests. The authors have no relevant financial or non-financial interests to disclose.

Author Contributions. N. Efstathopoulos and D. Papageorgiou had an equal contribution in the theoretical development of the proposed method. Furthermore, N. Efstathopoulos is the main author and was the main editor of the text, working under the supervision of D. Papageorgiou. The implementation of the method was solely made by Nikolaos Efstathopoulos, while the experimental evaluation was handled equally by N. Efstathopoulos and D. Papageorgiou. Lastly, Vasileios Tsimpos contributed by providing the presentation of the pin-hole projection principles and the exact derivation of the pixel-space Jacobian.

Code availability. The code implemented for the experimental evaluation is openly available and can be found in the following GitHub link: <https://github.com/CSRL-HMU/Null-space-Active-Perception-Enhancement-for-observing-Space-bounded-RDE>

References

- Alsadik B, Mahdi HA, Abdulateef NA (2025) A structured review and taxonomy of next-best-view strategies for 3d reconstruction. *ISPRS Open Journal of Photogrammetry and Remote Sensing* 17:100098. <https://doi.org/https://doi.org/10.1016/j.ophoto.2025.100098>, URL <https://www.sciencedirect.com/science/article/pii/S2667393225000171>
- Bajcsy R, Aloimonos Y, Tsotsos JK (2016) Revisiting active perception. URL <https://arxiv.org/abs/1603.02729>, arXiv:1603.02729

- Bechlioulis CP, Heshmati-Alamdari S, Karras GC, et al (2019) Robust image-based visual servoing with prescribed performance under field of view constraints. *IEEE Transactions on Robotics* 35(4):1063–1070
- Chaumette F, Hutchinson S (2006) Visual servo control. i. basic approaches. *IEEE Robotics & Automation Magazine* 13(4):82–90. <https://doi.org/10.1109/MRA.2006.250573>
- Chaumette F, Hutchinson S (2007) Visual servo control. ii. advanced approaches [tutorial]. *IEEE Robotics & Automation Magazine* 14(1):109–118. <https://doi.org/10.1109/MRA.2007.339609>
- Chen CW, Hung HA, Yang PH, et al (2021) Visual servoing of a moving target by an unmanned aerial vehicle. *Sensors* 21(17). <https://doi.org/10.3390/s21175708>, URL <https://www.mdpi.com/1424-8220/21/17/5708>
- Chen Y, Wu Y, Zhang Z, et al (2023) Image-based visual servoing of unmanned aerial manipulators for tracking and grasping a moving target. *IEEE Transactions on Industrial Informatics* 19(8):8889–8899. <https://doi.org/10.1109/TII.2022.3222688>
- Connolly C (1985) The determination of next best views. In: *Proceedings. 1985 IEEE International Conference on Robotics and Automation*, pp 432–435, <https://doi.org/10.1109/ROBOT.1985.1087372>
- Corke P, Hutchinson S (2001) A new partitioned approach to image-based visual servo control. *IEEE Transactions on Robotics and Automation* 17(4):507–515. <https://doi.org/10.1109/70.954764>
- Dunn E, Van Den Berg J, Frahm JM (2009) Developing visual sensing strategies through next best view planning. In: *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, IEEE, pp 4001–4008
- Efstathopoulos N, Papageorgiou D (2025) Optimized ukf-based perception of a repetitive dynamic event. In: *2025 European Control Conference (ECC)*, pp 3089–3094, <https://doi.org/10.23919/ECC65951.2025.11186969>
- Efstathopoulos N, Kounalakis N, Balaska V, et al (2025) Graspapad: Generalized framework for optimal grasp key-points active detection. In: *2025 34th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pp 629–635, <https://doi.org/10.1109/RO-MAN63969.2025.11217682>
- Eidenberger R, Scharinger J (2010) Active perception and scene modeling by planning with probabilistic 6d object poses. In: *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp 1036–1043, <https://doi.org/10.1109/IROS.2010.5651927>

- Gaudillière V, Simon G, marie odile B (2023) Perspective-1-ellipsoid: Formulation, analysis and solutions of the camera pose estimation problem from one ellipse-ellipsoid correspondence. *International Journal of Computer Vision* 131:1–25. <https://doi.org/10.1007/s11263-023-01794-x>
- He S, Tao Y, Spasojevic I, et al (2025) An active perception game for robust information gathering. In: 2025 IEEE International Conference on Robotics and Automation (ICRA), pp 14168–14174, <https://doi.org/10.1109/ICRA55743.2025.11128798>
- Hernandez-Barragan J, Villaseñor C, Lopez-Franco C, et al (2024) Image based visual servoing with kinematic singularity avoidance for mobile manipulator. *PeerJ Computer Science* 10:e2559
- Heshmati-Alamdari S, Bechlioulis CP, Liarokapis MV, et al (2014) Prescribed performance image based visual servoing under field of view constraints. In: 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems, IEEE, pp 2721–2726
- Hill J, Park W (1979) Real time control of a robot with a mobile camera. In: Proc. 9th Int. Symp. on Industrial Robots, pp 233–245
- Jiang J, Wang Y, Jiang Y, et al (2023) Robust image-based visual servoing constraints controller with guaranteed field of view. In: 2023 China Automation Congress (CAC), pp 6120–6125, <https://doi.org/10.1109/CAC59555.2023.10451163>
- Kastritsi T, Papageorgiou D, Doulgeri Z (2018) On the stability of robot kinesthetic guidance in the presence of active constraints. In: 2018 European Control Conference (ECC), IEEE, pp 622–627
- Li Z, Li W, Pan Y (2024) Composite learning image-based visual servoing of redundant robots with nullspace compliance. *IEEE Control Systems Letters* 8:315–320. <https://doi.org/10.1109/LCSYS.2024.3351559>
- Loschky L, Smith M, Chandran P, et al (2026) The role of event understanding in guiding attentional selection in real-world scenes: The scene perception & event comprehension theory (spect). *Attention, perception & psychophysics* 88. <https://doi.org/10.3758/s13414-026-03234-7>
- Ma W, Hutchinson S (2010) Gradient Projection Methods for Constrained Image-based Visual Servo, Springer London, London, pp 253–274. https://doi.org/10.1007/978-1-84996-089-2_14
- Papageorgiou D, Kounalakis N, Efstathopoulos N, et al (2024) Activo: An active perception framework for skill transfer through iterative visual observations. In: 2024 32nd Mediterranean Conference on Control and Automation (MED), pp 94–100, <https://doi.org/10.1109/MED61351.2024.10566166>

- Papageorgiou D, Efstathopoulos N, Angus Y, et al (2025) Laif: Learning by demonstration through active incremental data fusion of task observations. In: 2025 IEEE International Conference on Imaging Systems and Techniques (IST), pp 1–6, <https://doi.org/10.1109/IST66504.2025.11268388>
- Rasouli A, Lanillos P, Cheng G, et al (2020) Attention-based active visual search for mobile robots. *Autonomous Robots* 44(2):131–146
- Renz H, Krämer M, Hoffmann F, et al (2025) Next-best-trajectory planning of robot manipulators for effective observation and exploration. In: 2025 IEEE International Conference on Robotics and Automation (ICRA), IEEE, pp 6696–6702
- Rudnev V, Fox G, Elgharib M, et al (2025) Dynamic eventnerf: Reconstructing general dynamic scenes from multi-view rgb and event streams. In: Proceedings of the Computer Vision and Pattern Recognition Conference, pp 4866–4876
- Safronov E, Piga N, Colledanchise M, et al (2021) Active perception for ambiguous objects classification. In: 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp 4437–4444, <https://doi.org/10.1109/IROS51168.2021.9636414>
- Spica R, Giordano PR, Chaumette F (2014) Coupling visual servoing with active structure from motion. In: 2014 IEEE International Conference on Robotics and Automation (ICRA), pp 3090–3095, <https://doi.org/10.1109/ICRA.2014.6907303>
- Triantafyllou PD, Rovithakis GA, Doulgeri Z (2017) Guaranteeing field of view constraints in visual servoing tasks under uncertain dynamics. *IFAC-PapersOnLine* 50(1):2229–2234
- Wan E, Van Der Merwe R (2000) The unscented kalman filter for nonlinear estimation. In: Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium (Cat. No.00EX373), pp 153–158, <https://doi.org/10.1109/ASSPCC.2000.882463>
- Xungao Z, Min X, Jiansheng G, et al (2018) Robot manipulation using image-based visual servoing control with robust state estimation. In: 2018 Chinese Control And Decision Conference (CCDC), IEEE, pp 445–449
- Zhong X, Zhou Q, Sun Y, et al (2025) Deep reinforcement learning-based uncalibrated visual servoing control of manipulators with fov constraints. *Applied Sciences* 15(8). <https://doi.org/10.3390/app15084447>, URL <https://www.mdpi.com/2076-3417/15/8/4447>
- Zhou Y, Zhou Y, Wang H (2025) Handling constraints in image-based visual servoing: a planning method with artificial potential fields. *Robotic Intelligence and Automation* 45(3):326–334. <https://doi.org/10.1108/RIA-10-2024-0214>, URL

<https://doi.org/10.1108/RIA-10-2024-0214>, <https://www.emerald.com/ria/article-pdf/45/3/326/10190945/ria-10-2024-0214en.pdf>

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [FinalVideo.mp4](#)