

# GPA-Teleoperation: Gaze Enhanced Perception-aware Safe Assistive Aerial Teleoperation

Qianhao Wang <sup>†1,2</sup>, Botao He <sup>†2,3</sup>, Zhiren Xun <sup>1,2</sup>, Chao Xu <sup>1,2</sup>, and Fei Gao<sup>1,2</sup>

**Abstract**—Gaze is an intuitive and direct way to represent the intentions of an individual. However, when it comes to assistive aerial teleoperation which aims to perform operators' intention, rare attention has been paid to gaze. Existing methods obtain intention directly from the remote controller (RC) input, which is unfriendly to non-professional operators, as the experimental results show in Sec. VI-C. Further, most teleoperation works do not consider environment perception which is vital to guarantee safety. In this paper, we present GPA-Teleoperation, a gaze enhanced perception-aware assistive teleoperation framework, which addresses the above issues systematically. We capture the intention utilizing gaze information, and generate a topological path matching it. Then we refine the path into a safe and feasible trajectory which simultaneously enhances the perception awareness to the environment which the operator is interested in. Additionally, the proposed method is integrated into a customized quadrotor system. Extensive challenging indoor and outdoor real-world experiments and benchmark comparisons verify that the proposed system is reliable, robust and applicable to even unskilled users. We will release the source code<sup>1</sup> of our system to benefit related researches.

**Index Terms**—Aerial Systems; Applications; Telerobotics and Teleoperation; Motion and Path Planning

## I. INTRODUCTION

WITH the rapid development of aerial autonomy, UAV assistive teleoperation gradually shows huge potentials to reduce the difficulty of operations and raise the safety of missions. Assistive teleoperation, which aims to accomplish humans' intention by machines, puts forward twofold requirements. Firstly, it should capture the intention of operators precisely and timely. Secondly, it should be accompanied by a proper acting strategy under the promise of safety.

Gaze fixations are strongly correlated to the operator's intention, which has been observed in various fields, including car driving [1], gaming [2], and drone racing [3]. Naturally, gaze

Manuscript received: September, 9, 2021; Revised: December, 21, 2021; Accepted: January, 23, 2022. This paper was recommended for publication by Editor Pauline Pounds upon evaluation of the Associate Editor and Reviewers' comments. This work was supported by National Natural Science Foundation of China under Grant 62003299 and Grant 62088101. (Corresponding author: Fei Gao, Chao Xu.)

<sup>†</sup> Equal contribution.

<sup>1</sup> State Key Laboratory of Industrial Control Technology, Institute of Cyber-Systems and Control, Zhejiang University, Hangzhou, 310027, China.

<sup>2</sup> Huzhou Institute of Zhejiang University, Huzhou, 313000, China.

<sup>3</sup> School of Automation, Nanjing Institute of Technology, Nanjing, 211112, China.

Email: {qhwangaa, fgaoaa}@zju.edu.cn

Digital Object Identifier (DOI): see top of this page.

<sup>1</sup><https://github.com/ZJU-FAST-Lab/GPA-Teleoperation>

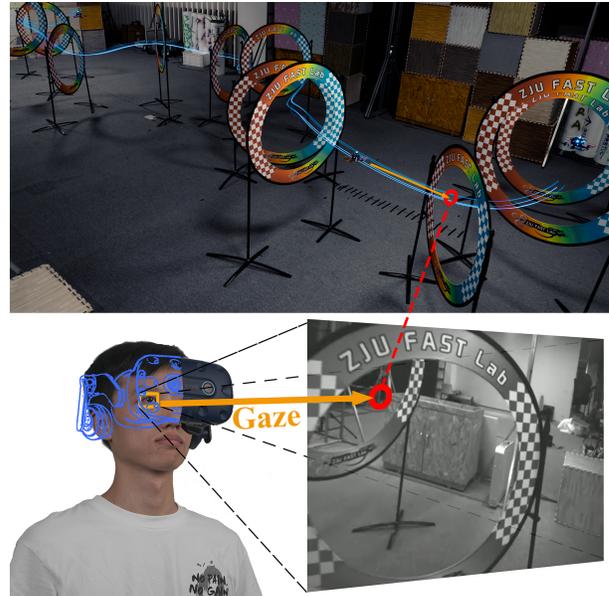


Fig. 1: Illustration of the real-world drone-racing experiment, which will be detailed in Sec. VI-C. **Top**: the composite long exposure trajectory of quadrotor flying with our assistive teleoperation system. **Bottom**: the operator with an eye tracker and the image he receives now is from the first view of the quadrotor in the yellow box. Red circles in the top and bottom figure represent gaze data in three dimensions and the first view respectively.

has a non-negligible potential for representing intentions. However, in the literature on aerial teleoperation, few works try to utilize the gaze of humans. Instead, most existing works [4]–[6] directly read the intention input from RC, which may not be able to reflect the true intention of naive operators. This is because using RC requires the operator to simultaneously operate at least four channels to control drone's direction and speed, which is hard for those who do not have long-term practices. These unstable intentions are particularly obvious when the drone is operated by an unskilled man in an unfamiliar environment and may prevent the assistive flight system from being truly accepted by consumers. In this paper, inspired by the work [3] which finds the high correlation between gaze and intention, we propose an aerial assistive teleoperation system, which stably captures the human intention by using the gaze information, and react to this intention in a user-friendly way.

In addition to intention inference, safety is another pivotal element for aerial assistive teleoperation. Since the intention may originate from an unskilled operator, increasing the safety margin in occluded and complex environments is vital.

Therefore, a robust flying assistance system should not only generate safe motions using operator's intention, but also assist operators to do timely decisions with environmental awareness. Recently, many works [7], [8], including our previous research [9] start to consider perception awareness by planning the drone to observe and avoid unknown obstacles actively. Nevertheless, these works cannot be directly applied to assistive aerial teleoperation (details in Sec. II-C). To address this issue, we propose a novel perception-aware planner by improving our previous work [9] to enhance the visibility of where the operator's intention refers.

Furthermore, keeping the speed of the drone on par with the operator's expectation is also essential in a user-friendly teleoperation system. This is because frequent unexpected speed changes will make the operator feel like they are losing effective control of the drone. To address this problem, we upgrade our trajectory planner to maintain consistent with the operator's expected speed as much as possible.

In this paper, we propose a **Gaze enhanced Perception-aware safe Assistive aerial teleoperation system called GPA-Teleoperation**. Based on eye movements and RC input, we plan a topological path that matches operator's intention. Derived from this path, we further propose the back-end optimizer which refines the path into a trajectory which simultaneously ensures safety and increases the perception awareness to environment of interest. Additionally, we integrate our proposed system into a customized quadrotor.

Additionally, we compare our system with existing consumer class assistive aerial teleoperation by several operators. Extensive real-world experiments show our method performs efficiently and safely. The contributions of this paper are:

- 1) A novel intention capture method utilizing gaze fixation, which generates a guiding topological path for trajectory optimization. Precise intention obtained from this method makes the system more friendly to operators.
- 2) A perception-aware trajectory optimization method, which simultaneously optimizes the visibility to environments where the operator focuses, and maintains the speed that the operator desires. It promotes the safety and user experience of the proposed system.
- 3) A complete gaze enhanced assistive teleoperation system with code open source. Extensive real-world tests validate that our method is practical and effective.

## II. RELATED WORKS

### A. Gaze for Intention Inference

There have been many works, mainly for car driving scenarios, studying gaze-based intention inference. With Land et al. [1] first illustrate the high relevance between gaze and future trajectory, several works repeatedly confirm this both in real-world [10], [11] and simulation [12]. For UAVs, the phenomenon has also been observed. In [13], the authors notice that eye movement patterns are specific in different flight stages like take-off and landing. In [3], Pfeiffer et al. find that gaze fixations are highly correlated with operator's control commands, which means eye movements can well reflect the operator's intention. These works provide solid evidence that

humans tend to use specific gaze patterns to indicate their intents. However, none of these works integrates human gaze into a closed control loop. While several works in visual servoing [14], [15] indicate that using gaze to control robots for specific tasks are intuitive and operators do not need any long-term professional training.

### B. Assistive Aerial Teleoperation

Several previous works [16], [17] formulate the assistive teleoperation of UAVs as a local control problem to avoid collisions. However, these works do not take operator's intention into account. In recent years, some works have been proposed to incorporate operator's intention into assistive teleoperation. Yang et al. [4] handle the intention of an operator by modeling it as an indicative direction and follow it by generating trajectories from a set of motion primitives. However, this sampling-based method is time-consuming and cannot guarantee a promising trajectory due to the inherent discretization error. Recently, Yang et al. [6] propose a hierarchical teleoperation framework considering global intentions. They capture intention of the operator to generate a global path and plan local safe trajectories while following it. The above works capture intentions directly from the inputs of a RC, which is unfriendly to operators without long-term training in challenging environments. However, by capturing intentions precisely and stably with gaze information, our system is applicable to even naive users.

### C. Perception-aware Navigation in Unknown Environments

For UAVs, generating safe flight trajectories is the key to developing a navigation system. Richter et al. [18] propose a learning-based method to plan for better visibility to unknown areas. However, the huge amount of data and training prevents it from being generalized to complex 3D environments. Zhou et al. [8] propose a perception-aware planning strategy. They use an iterative task-specific trajectory refinement method to achieve risk awareness. This method guarantees sufficient safe reaction distance for suddenly appeared obstacles. In this paper, we extend our previous work [9] to equip our assistive teleoperation system with environmental awareness. Our method successfully increases confidence and safety when operators teleoperate the drone in complex environments.

## III. SYSTEM OVERVIEW

The problem statement of our work is: capturing operator's direction intention from gaze and speed intention from RC input, our system generates a safe trajectory that conforms to the intention for the UAV to execute. The pipeline of this work is illustrated in Fig. 2. Our teleoperation system is divided into two subsystems, operator interaction subsystem and UAV subsystem.

For the operator interaction subsystem, eye tracker and RC are used respectively to obtain operator's expected direction and speed. Since RC is only used for speed control, we only need a single channel input from it, which is similar to the

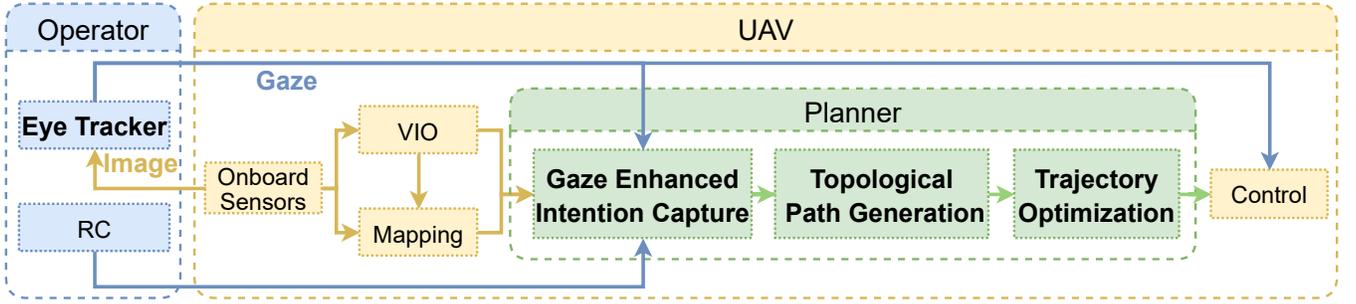


Fig. 2: A diagram of our teleoperation system architecture.

accelerator of a vehicle. Eye tracker receives the first view image from the UAV’s onboard sensor.

For the UAV subsystem, state estimation, mapping, planning and control modules are all working onboard. Based on expected direction and speed, the planner generates a local target and the operator’s desired topological path, which will be detailed in Sec. IV. Then by refining this topological path, the planner optimizes a perception-aware trajectory that makes the UAV keep at operator’s desired speed, as presented in Sec. V. Additionally, to allow operator to view the intended topology better, we control the yaw of UAV to place the topology around the center of the view.

---

**Algorithm 1: Topological Path Generation**


---

```

let  $Q$  be a queue; label  $g_D$  as explored;
 $Q.enqueue(g_D)$ ;
while  $Q$  is not empty do
   $v := Q.dequeue()$ ;
  if  $v.depth$  in perception range and  $v$  is
  unclustered then
     $c_{temp} := DBSCAN(v, D)$ ;
    label  $c_{temp}$  as clustered;
    if  $g_D \in \text{BoundingBox}(c_{temp})$  then
       $C.enqueue(c_{temp})$ ;
    else
      if  $c_m^1$  is empty then
         $c_m^1 = (c_{temp})$ ;
         $C.enqueue(c_m^1)$ ;
      else
         $M = \text{Merge}(c_m^1, c_{temp})$ ;
        if  $g_D \in \text{BoundingBox}(M)$  then
           $c_m^2 = c_{temp}$ ;
           $C.enqueue(c_m^2)$ ;
          Break;
     $Q.enqueue(\text{VisitNeighbours}(v))$ ; // BFS
return  $C$ ;

```

---

#### IV. GAZE ENHANCED INTENTION GUIDING TOPOLOGICAL PATH GENERATION

In order to optimize a perception-aware trajectory, we need to firstly generate a topological path that is consistent with the operator’s intention. In this section, we propose an intention

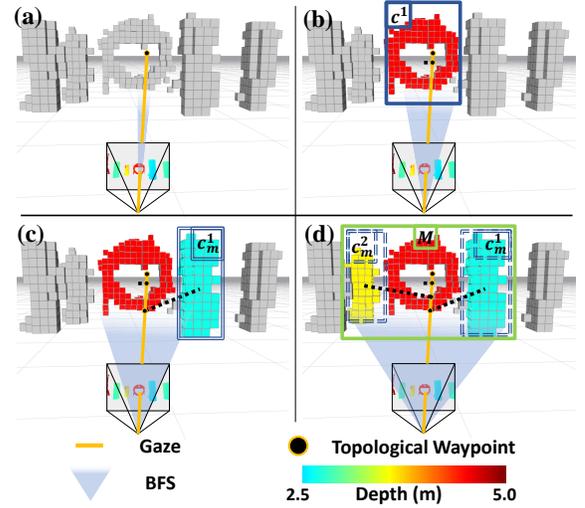


Fig. 3: Working principle of our topological path generation algorithm. (a): BFS is conducted on  $D$  with gaze as the origin. (b): Object is clustered and signed as  $c^i$  ( $i = 1, 2, \dots$ ) if the bounding-box of it contains the gaze point. (c): The first object that does not contain the gaze point is clustered and marked as  $c_m^1$ . (d): BFS is continued until the bounding box of  $M$  includes the gaze point.

guiding topological path generation module utilizing gaze fixation. An illustration of this section can be found in Fig. 3. The input of the proposed module is a control signal from RC, a gaze point  $g_I$  on the first person view (FPV) image plane  $I$  and a depth image  $D$ . For gaze, we register  $g_I$  to  $D$  according to the intrinsic and extrinsic matrix of  $D$  and  $I$ , sign as  $g_D$ . Notably, since  $g_I$  is a 2-D vector, it cannot be directly projected to 3-D space. Therefore, we use the depth of maximum reliable perception range as its depth and project it to 3-D space as a local target. After that, we apply a smoothing filter for the past ten inputs to eliminate the noise of gaze like blink and glance.

After all the pre-processes, we have captured the operator’s intention. To ensure that the planned path is consistent with intention, we need to utilize the intention to constrain the topological structure of the path. To achieve this, **Algorithm 1** is proposed. Firstly,  $g_D$  is utilized as the origin for Breadth First Search (BFS) [19] on  $D$ . The BFS would halt once it visits a pixel with its depth in the perception range and is unclustered before. Then, we take the pixel as the anchor point for DBSCAN [20] clustering, resulting in a cluster  $c_{temp}$ , as illustrated in Fig. 3(b). For each pixel in  $c_{temp}$ , we label it as clustered. If  $g_D$  is in the boundingbox of  $c_{temp}$ , meaning the gaze hit the object, we enqueue  $c_{temp}$  into the cluster-set  $C$  and label it as  $c^i$  ( $i = 1, 2, \dots$ ). Otherwise, we check

if it is the first object that is not hit by the gaze, if so, we labeled  $c_{temp}$  as the first margin cluster  $c_m^1$ , as is shown in Fig. 3(c). After assign  $c_m^1$ , for each cluster that is not hit by the gaze, we merge it with the margin object as  $\mathcal{M}$ , like Fig. 3(d) does, and check if  $g_D$  locates in the bounding-box of  $\mathcal{M}$ . If so, the latest clustered object is signed as  $c_m^2$  and the BFS is stopped. Then, we project  $g_D$  as a line to Body Frame, signed as  $g_B$ . Finally, we project the central points of clusters  $\mathcal{C} = \{c^1, c^2 \dots c^i, c_m^1, c_m^2\}$  to  $g_B$  as topological waypoints and  $g_B$  as the topological path.

## V. PERCEPTION-AWARE SPATIAL-TEMPORAL TRAJECTORY OPTIMIZATION

In this section, we refine the topological path planned in Sec. IV to generate a perception-aware trajectory, which is also expected to keep the UAV at the operator's desired speed. For perception-awareness, based on our previous work [9], we describe an analytical visibility metric (Sec. V-A) that can be easily formulated into a differentiable penalty (Sec. V-C3). For keeping desired speed, we penalize the trajectory execution time (Sec. V-C1) and limit the speed by constraining the maximum allowed speed (Sec. V-C2). To directly control both the spatial and temporal profile of trajectory, we adopt MINCO [21] representation (Sec. V-B).

### A. Analytical Visibility Metric

As the two-dimension profile Fig. 4 shows,  $\mathbf{p}_{vis}$  is a point on the quadrotor's trajectory and  $\mathbf{v}$  is the position of the target which depends on the gaze. Since the attitude of quadrotor is controlled by the direction of gaze as mentioned in Sec. III, we assume that the aircraft is always facing the target. The blue range represents the FOV of quadrotor. The blue dashed enveloped area represents the **confident FOV** which is desired to be obstacle-free. But it is hard and costly to represent this requirement analytically. Therefore, we generate a sequence of spherical areas  $\{\mathcal{B}_1, \mathcal{B}_2, \dots, \mathcal{B}_N\}$  to approximate the confident FOV. The  $\mathcal{B}_k$  is shown as the red circle, and its center  $\mathbf{v}_k$  and radius  $r_k$  are calculated by

$$\mathbf{v}_k = \mathbf{p}_{vis} + \psi_k(\mathbf{v} - \mathbf{p}_{vis}), \quad r_k = \rho \cdot \psi_k \cdot \|\mathbf{v} - \mathbf{p}_{vis}\|, \quad (1)$$

where  $\psi_k = k/N, k \in \{1, 2, \dots, N\}$ , and  $\rho$  is a constant parameter that determines size of the confident FOV.

Then we can enhance the visibility of target analytically by requiring each ball to satisfy

$$\Xi(\mathbf{v}_k) > r_k, \quad \Xi(\mathbf{c}_i) : \mathbb{R}^3 \rightarrow \mathbb{R}, \quad (2)$$

which is used to construct visibility penalty function in [9] and  $\Xi(\mathbf{c}_i)$  is the distance to the closest obstacle. However, in this paper we transform the requirement of visibility Eq.2 into

$$(\Xi(\mathbf{v}_k)/l_k) > \rho, \quad l_k = \psi_k \|\mathbf{v} - \mathbf{p}_{vis}\|, \quad (3)$$

based on which we design the visibility penalty in this work. This penalty performs better than [9] in terms of enhancing visibility, which is detailed in Sec. V-C3.

As shown in Fig. 4(b), by expelling obstacles from the spherical area, this method effectively increase the visibility of the target.

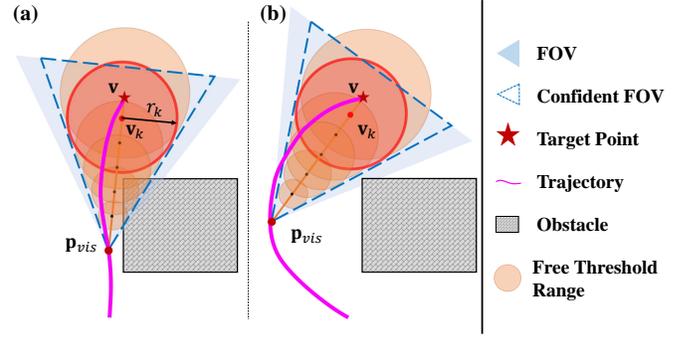


Fig. 4: Illustration of the visibility metric definition. (a): the trajectory ignoring visibility. (b): the trajectory considering visibility.

### B. Trajectory Representation

In this paper, we adopt the MINCO [21] to conduct spatial-temporal deformation of the flat-output trajectory. An  $s$ -order MINCO trajectory is indeed a  $2s$ -order polynomial spline with constant boundary conditions, which is defined as:

$$\mathfrak{T}_{\text{MINCO}} = \left\{ p(t) : [0, T] \mapsto \mathbb{R}^m \mid \mathbf{c} = \mathcal{M}(\mathbf{q}, \mathbf{T}), \right. \\ \left. \mathbf{q} \in \mathbb{R}^{m(M-1)}, \quad \mathbf{T} \in \mathbb{R}_{>0}^M \right\}, \quad (4)$$

where  $p(t)$  is an  $m$ -dimensional  $M$ -piece polynomial trajectory.  $\mathbf{c} = (\mathbf{c}_1^T, \dots, \mathbf{c}_M^T)^T \in \mathbb{R}^{2Ms \times m}$  is the polynomial coefficient and  $\mathcal{M}(\mathbf{q}, \mathbf{T})$  is a linear-complexity smooth map from intermediate points  $\mathbf{q}$  and a time allocation  $\mathbf{T}$  for all pieces to the coefficients of splines. A spline with  $\mathbf{c} = \mathcal{M}(\mathbf{q}, \mathbf{T})$  is exactly the unique control effort minimizer of an  $s$ -integrator that passes  $\mathbf{q}$ . For any penalty function  $F(\mathbf{c}, \mathbf{T})$  with available gradients, MINCO can also serve as a linear-complexity differentiable layer  $H(\mathbf{q}, \mathbf{T}) = F(\mathcal{M}(\mathbf{q}, \mathbf{T}), \mathbf{T})$ . Then  $\partial H / \partial \mathbf{q}$  and  $\partial H / \partial \mathbf{T}$  can be obtained efficiently from the corresponding  $\partial F / \partial \mathbf{c}$  and  $\partial F / \partial \mathbf{T}$ .

The  $i$ -th piece  $p_i(t)$  of trajectory is defined by

$$p(t) = p_i(t - t_{i-1}), \quad \forall t \in [t_{i-1}, t_i]. \quad (5)$$

$$p_i(t) = \mathbf{c}_i^T \beta(t), \quad \forall t \in [0, T_i], \quad (6)$$

where  $\beta(t) := [1, t, \dots, t^N]^T$  is the natural basis,  $\mathbf{c}_i$  the coefficient matrix,  $T_i = t_i - t_{i-1}$  and  $T = \sum_{i=1}^M T_i$ .

Specifically, we define **constraint points**  $\hat{\mathbf{p}}_{i,j}$  which are sampled on each piece of the trajectory by  $\hat{\mathbf{p}}_{i,j} = p_i((j/\kappa_i)T_i)$ ,  $j \in \{0, 1, 2, \dots, \kappa_i\}$ , where  $\kappa_i$  is the sample number on the  $i$ -th piece. These points are used to transform the time integral of some constraint functions, such as the Dynamical Feasibility Penalty in Sec. V-C2, into weighted sum of sampled penalty functions.

### C. Problem Formulation

Inspired by [8], first we generate an intermediate warm-up trajectory as the initial trajectory utilizing the topological path from Sec. IV. Then taking our requirements into account, we formulate the trajectory optimization problem as

$$\min_{\mathbf{q}, \mathbf{T}} [J_e, J_d, J_t, J_{vis}, J_c, J_u] \cdot \lambda, \quad (7)$$

where  $\lambda$  is the a weighting vector to trade off each penalty which will be described below. In this paper, we use the quasi-Newton method proposed in [22] which supports nonsmooth cost functions to solve this unconstrained optimization problem.

1) *Execution Time Penalty  $J_t$* : To increase the speed of UAV, we penalty the total execution time  $J_t = \sum_{i=1}^M T_i$ . Its gradient is  $\partial J_t / \partial \mathbf{c} = \mathbf{0}$ ,  $\partial J_t / \partial \mathbf{T} = \mathbf{1}$ .

2) *Dynamical Feasibility Penalty  $J_d$* : To prevent the speed from increasing excessively introduced by  $J_t$ , we use the desired speed to set the value of the maximum allowed speed. Additionally violent shifts of speed will lead to a blurred first view, which is unfriendly to operator. Thus not only velocity, we also limit the amplitude of acceleration and jerk. The constraints are written by

$$\mathcal{G}_* = p_i^{(n)}(t)^2 - \max_*^2, \quad \forall t \in [0, T_i], \quad (8)$$

where  $*$  =  $\{v, a, j\}$ ,  $n = \{1, 2, 3\}$ , and  $t = jT_i/\kappa_i$ , respectively.  $\max_v, \max_a, \max_j$  are maximum allowed velocity, acceleration and jerk. Inspired by the method in [23], the dynamical feasibility penalty is obtained by computing weighted sum of sampled constraint function:

$$(J_d)_* = \sum_{i=0}^M \frac{T_i}{\kappa_i} \sum_{j=0}^{\kappa_i} \bar{\omega}_j \max(\mathcal{G}_*(\hat{\mathbf{p}}_{i,j}), \mathbf{0})^3, \quad (9)$$

where  $(\bar{\omega}_0, \bar{\omega}_1, \dots, \bar{\omega}_{\kappa_i-1}, \bar{\omega}_{\kappa_i}) = (1/2, 1, \dots, 1, 1/2)$  are the quadrature coefficients following the trapezoidal rule [24] and  $\hat{\mathbf{p}}_{i,j}$  are constrain points. The gradient can be written by

$$\frac{\partial (J_d)_*}{\partial \mathbf{c}_i} = \frac{\partial (J_d)_*}{\partial \mathcal{G}} \frac{\partial \mathcal{G}}{\partial \mathbf{c}_i}, \quad (10)$$

$$\frac{\partial (J_d)_*}{\partial T_i} = \frac{\partial (J_d)_*}{\partial \mathcal{G}} \frac{\partial \mathcal{G}}{\partial t} \frac{\partial t}{\partial T_i} + \sum_{j=0}^{\kappa_i} \frac{\bar{\omega}_j}{\kappa_i} \kappa_i \max(\mathcal{G}_*(\hat{\mathbf{p}}_{i,j}), \mathbf{0})^3, \quad (11)$$

$$\frac{\partial \mathcal{G}_*}{\partial \mathbf{c}_i} = 2\beta^{(n)}(t)p_i^{(n)}(t)^T, \quad \frac{\partial \mathcal{G}_*}{\partial t} = 2\beta^{(n+1)}(t)^T \mathbf{c}_i p_i^{(n)}(t), \quad (12)$$

3) *Visibility Penalty  $J_{vis}$* : To achieve the visibility demand defined in Sec. V-A, we design visibility penalty as

$$J_{vis} = \sum_{i=0}^M \frac{T_i}{\kappa_i} \sum_{j=0}^{\kappa_i} \bar{\omega}_j F_{vis}(\hat{\mathbf{p}}_{i,j}). \quad (13)$$

where these constraint points  $\hat{\mathbf{p}}_{i,j}$  are selected as the visibility points  $\mathbf{p}_{vis}$ . For each visibility point  $\mathbf{p}_{vis}$ , we define

$$F_{vis}(\mathbf{p}_{vis}) = \sum_{k=1}^N \max(f_{vis}(\mathbf{v}_k), 0)^3, \quad (14)$$

$$f_{vis}(\mathbf{v}_k) = (\rho - \Xi(\mathbf{v}_k)/l_k), \quad (15)$$

where  $N, \mathbf{v}_k, l_k$  and  $\rho$  are the parameters of spherical areas mentioned in Sec. V-A and  $\Xi(\mathbf{v}_k)$  is obtained from ESDF.

Then the gradient of  $f_{vis}$  cost can be calculated by

$$\frac{\partial f_{vis}}{\partial \mathbf{p}_{vis}} = \frac{1}{l_k^2} \left[ \frac{\psi_k}{l_k} \Xi(\mathbf{v}_k)(\mathbf{p}_{vis} - \mathbf{v}_k) - (1 - \psi_k) l_k \frac{\partial \Xi(\mathbf{v}_k)}{\partial \mathbf{v}_k} \right], \quad (16)$$

where  $\partial \Xi(\mathbf{v}_k) / \partial \mathbf{v}_k$  can be efficiently acquired from ESDF.

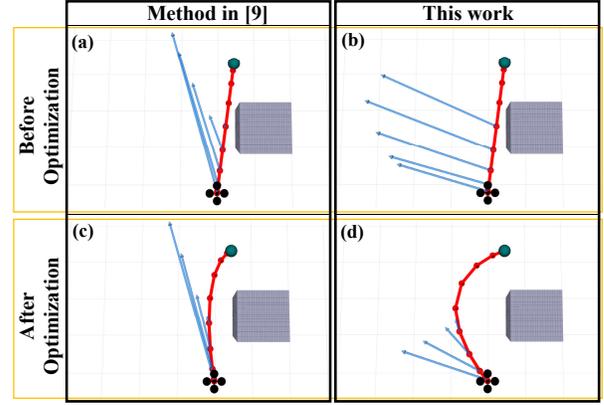


Fig. 5: Comparison of [9] and this work. The gray square is an obstacle, the black quadrotor is the initial position of the drone, the green ball indicates the target, the red curve indicates the trajectory and the points on it indicate the constraint points, and the light blue arrows indicate the negative gradients of the visibility costs.

Note that each visibility point  $\mathbf{p}_{vis}$  is a constraint point  $\hat{\mathbf{p}}_{i,j} = p_i(t)$ ,  $t = jT_i/\kappa_i$ , whose gradient is written by

$$\frac{\partial \hat{\mathbf{p}}_{i,j}}{\partial \mathbf{c}_i} = \beta(t), \quad \frac{\partial \hat{\mathbf{p}}_{i,j}}{\partial t} = \dot{\beta}(t)^T \mathbf{c}_i. \quad (17)$$

Then with the above Eq. 16, 17 the gradient of the visibility penalty  $J_{vis}$  can be calculated easily using the chain rule.

Notably, in this work, we upgrade the formulation of the visibility requirements compared to [9]. Specifically, compared to Eq. 3, the method in [9] formulates the visibility requirement into  $\Xi(\mathbf{v}_k) > r_k$ , in which the visibility cost for each visibility point is designed as

$$\tilde{F}_{vis}(\mathbf{p}_{vis}) = \sum_{k=1}^N \max(\tilde{f}_{vis}(\mathbf{v}_k), 0)^3, \quad (18)$$

$$\tilde{f}_{vis}(\mathbf{v}_k) = (r_k^2 - \Xi(\mathbf{v}_k))^2. \quad (19)$$

Comparing Eq. 15 and Eq. 19, the key improvement is that in Eq. 15 the direct effect of  $\|\mathbf{v} - \mathbf{p}_{vis}\|$  on the cost function is weakened, which makes the cost more focused on occlusion, rather than directly shortening  $\|\mathbf{v} - \mathbf{p}_{vis}\|$  to reduce the cost.

To show the significance of the improvement intuitively, in Fig. 5 we visualize these two visibility cost functions' negative gradients, which are  $-\partial \tilde{F}_{vis} / \partial \mathbf{p}_{vis}$  and  $-\partial F_{vis} / \partial \mathbf{p}_{vis}$ . We use these two costs separately for comparison in the trajectory optimization problem in Sec. V-C. Both of them are set up with the same optimization parameters, and each penalty functions of the optimization is the same except for the visibility penalty. To fairly compare these two methods, two visibility costs are adjusted to the same order of magnitude.

As shown in Fig. 5(a) and Fig. 5(c), the gradient of visibility cost in [9] has a very large component in the direction of  $(\mathbf{v} - \mathbf{p}_{vis})$ , which leads to the trajectory after optimization with less effective visibility improvement. As shown Fig. 5(b) and Fig. 5(d), compared to [9], the visibility cost in this work can enhance visibility more remarkably.

4) *Other Penaltise*: In order to guide trajectory with a collision-free path, we adopt the collision evaluation of EGO-Planner [25] to construct Collision Avoidance Penalty  $J_c$ . We also formulate Control Effort  $J_e$  and Uniform Distribution Penalty  $J_u$  to make trajectory smooth and reliable. Readers can refer to [25] and [26] for more detailed explanations.

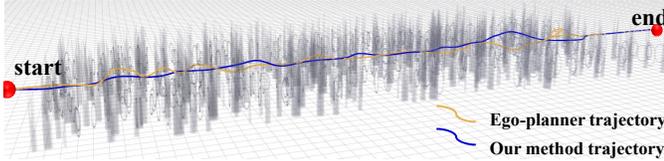


Fig. 6: Comparison of the trajectory of Ego-planner and our method in a complex environment, which has an obstacle density of  $0.625 \text{ obs}/m^2$ . The size of the environment is  $60 \times 20 \times 2m$ .

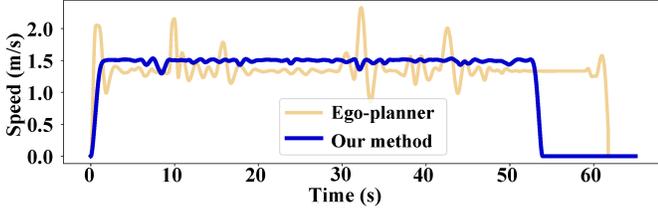


Fig. 7: Speed comparison between Ego-planner and our method.

## VI. EVALUATION AND EXPERIMENT

### A. Implementation Details

All real-world evaluations are completed by multiple volunteers and most of them have little flight experience. The experimental procedure is as follows. Upon arrival at the experiment site, participants are informed of the task. Then they are invited to practice three times before each experiment. Finally, they are asked to perform each experiment three times repeatedly and the relevant data is recorded.

Our quadrotor system's hardware setups is based on our previous work [25]. The parameters of visibility metric are  $N = 20$ ,  $\rho = 0.8$ . Our method can run in real time and cost about 5 ms to solve the optimization problem formulated in Sec. V on an onboard computer DJI Manifold 2C. Readers can get a better understanding of each experiment from the attached video<sup>2</sup>.

### B. Evaluation for Keeping Desired Speed

To show the significance of the improvement in our trajectory optimization for maintaining the operator's desired speed, as Fig 6 shows, we compare our method with Ego-planner [25], which is the basis of the trajectory planning in [9]. We set the same start and end points, and the same desired speed as 1.5 m/s. Since only the performance of keeping expected speed is compared, in this experiment, we set the Visibility Penalty (Sec. V-C3) to 0 in our method.

As Fig. 7 shows, with the aids the newly added Execution Time Penalty (Sec. V-C1) and Dynamical Feasibility Penalty (Sec. V-C2), our method performs remarkably better at keeping desired speed. However, Ego-planner often requires a change in speed for obstacle avoidance due to the limitations of the uniform B-spline curve.

### C. Evaluation for Gaze Enhanced Intention Capture

To demonstrate the significance of gaze in intention inference and topological path generation, we implement this experiment in a challenging drone racing scenario as shown in Fig. 1 and Fig. 8. We compare our gaze enhanced method with

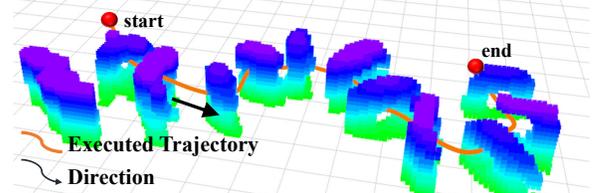


Fig. 8: Trajectories executed by the quadrotor and online generated map which is colored by height in drone racing experiment.

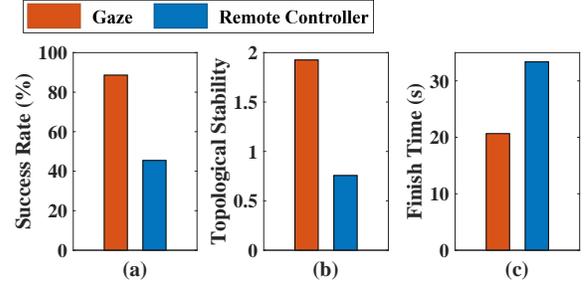


Fig. 9: Results of the comparisons in drone racing experiment.

the commonly used RC method: assistive teleoperation system which captures intention directly from RC. Note that both methods use the same perception-aware trajectory optimization method which is described in Sec. V.

To evaluate the stability and accuracy of intention inference, we proposed three independent indicators: average success rate of passing rings  $\bar{S}$ , topological stability  $\mathcal{T}_s$  and average finish time  $\bar{T}$ . We define  $\bar{S}$  and  $\mathcal{T}_s$  as

$$\bar{S} = \frac{N_{success}}{N_{sum}}, \quad \mathcal{T}_s = \frac{N_{sum}}{\sum_{i=0}^{N_{sum}} (N_{switch})_i}, \quad (20)$$

where  $(N_{switch})_i$  counts times the topology of the planned path is inconsistent with the intent as it passes through the  $i$ -th ring.  $N_{success}$  is the number of rings that quadrotor passes through successfully and  $N_{sum}$  is the total number of rings.

As the results show in Fig. 9, benefit from gaze, our method capture precise intention, which helps us have better topological stability and a higher success rate. However, for the RC method, inaccurate intention causes operators to adjust the RC input repeatedly. This leads to low topological stability, which is the major cause of the low success rate.

### D. Evaluation for Perception-aware Trajectory Optimization

To validate the superiority of the proposed perception-aware trajectory optimization method for flight safety, we compare it with the one which is not equipped with this feature. And **PA** and **NoPA** are used to refer to whether the perception-aware method is used in our system or not.

The volunteers are asked to fly quadrotor following the white route (from ① to ④) in the maze, as Fig. 11 shows. There are three corners where we place **surprise obstacles (SO)** right behind, and these surprise obstacles ( $SO_1$ ,  $SO_2$  and  $SO_3$ ) are marked by green, yellow and blue boxes respectively in Fig. 11 and Fig. 10. As Fig. 10 shows, PA always turns on a safer trajectory that enhances the visibility of the target, which allows the quadrotor to see SO earlier.

Additionally, we record the distance between SO and the quadrotor when volunteers see SO for the first time. In Fig. 12, the results presented as box charts show that our

<sup>2</sup><https://www.youtube.com/watch?v=WYujLePQwB8>

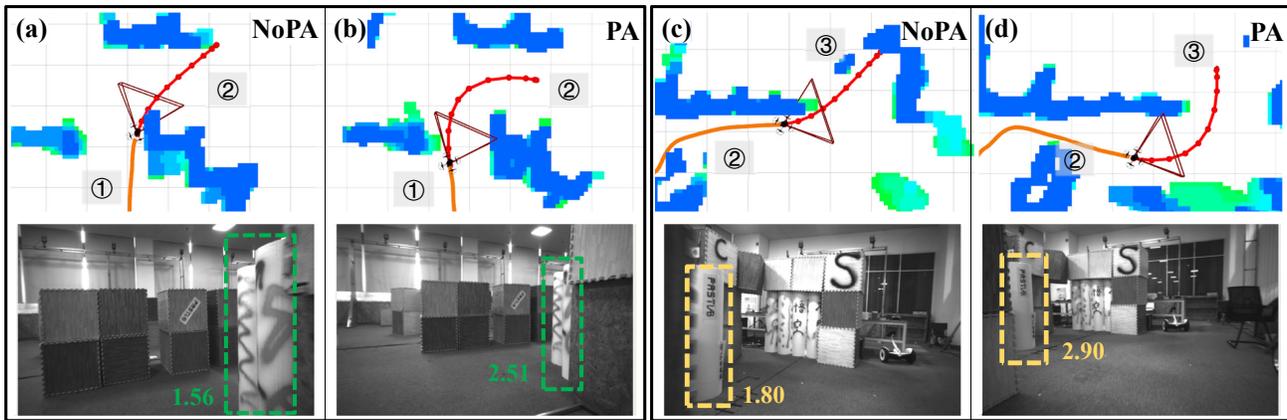


Fig. 10: Comparison of PA (perception-aware) and NoPA (without perception-aware). **Top**: the visualization of the experiment. The orange curve denotes the executed trajectory, the red curve denotes the planned trajectory, and the red pyramid stands for FOV. (a) and (b) show the quadrotor flies from ① to ②. (c) and (d) show the quadrotor flies from ② to ③. **Bottom**: the first-person perspective of the quadrotor, the obstacles marked by green and yellow are corresponding to  $SO_1$  (surprise obstacle) and  $SO_2$  in Fig. 11. The colored numbers indicate the distance between the quadrotor and the obstacles.

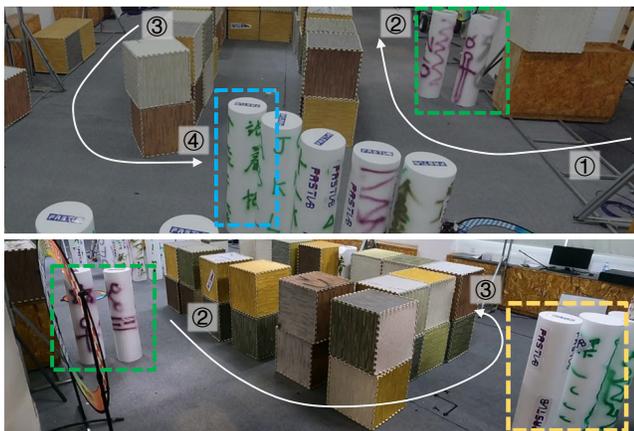


Fig. 11: Two third-person views of the experiment. Volunteers are asked to fly the quadrotor from ① to ④, and through ② and ③ in turn.  $SO_1$ ,  $SO_2$  and  $SO_3$  are marked in green, yellow and blue respectively.

PA method significantly increases the distance between SO and the quadrotor when SO is seen the first time, which demonstrates that our method assists operators to fly safer and makes more time for operators to do decisions.

### E. System-level Comparison

In this section, experimental scenarios have been divided into indoor flight and outdoor flight. Although there are some previous works on assistive teleoperation [4]–[6], none of them have runnable software available. To prove that our system is practical and robust, we compare our system with the industry-leading consumer UAV, DJI Mavic Air 2<sup>3</sup>, which is equipped with three-direction perception, and DJI’s APAS 3.0 (Advanced Pilot Assistance Systems). Note that Mavic Air 2 obtain operator’s intention from the input of RC only.

1) *Indoor Flight Test*: We conduct challenging indoor flight experiments, as shown in Fig. 14, to prove the high controllability and robustness of our system.

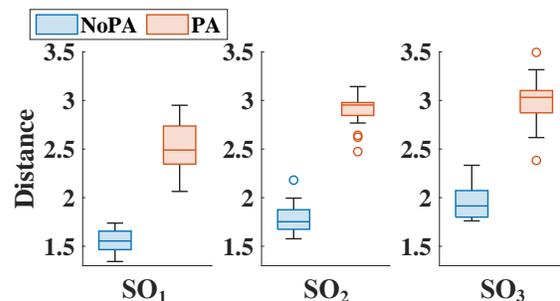


Fig. 12: Results of comparisons between NoPA and PA, the indicator refers to the distance that SO first appears fully in the field of view.

In the first half of the experiment, as Fig. 14(a) shows, the obstacle is densely placed with less than 1.5m away from each other. The task of the scene is to achieve the goal, 11m away from the start point, within the prescribed topological path as soon as possible. The results are presented in TABLE I. Both the average time consumption and maximum speed of our system outperform Mavic Air 2 significantly. For Mavic Air 2, redundant sensors and limited computing power lead to a conservative performance. Furthermore, this RC-only assistive teleoperation system cannot obtain precise intention, which is unfriendly to operator in a dense environment.

In the second half of the experiment, as Fig. 14(b) shows, we set several 2m-width gates, 1m-diameter rings and 0.8m-diameter rings. Note that our quadrotor’s diameter is 0.4m and Mavic Air 2’s diameter is about 0.5m. As Fig. 13 shows, our system can pass through various types of obstacles flexibly and fast. However, Mavic Air 2 can only pass through 2m-width gates. The result proves that our system has more applicability and robustness for different environments.

2) *Outdoor Flight Test*: Finally, to illustrate the reliability and potentiality of this work, we test in the large-scale outdoor environment, which is  $35 \times 35m$  in scale. Volunteers are asked to go through several gates and rings in order. The average time consumption and maximum speed are 49.5s and 1.96 m/s for the proposed system. For the consumer UAV, the respective data is 83s and 1.60 m/s. The attached video shows our method is more smooth and fast.

<sup>3</sup><https://www.dji.com/cn/mavic-air-2>

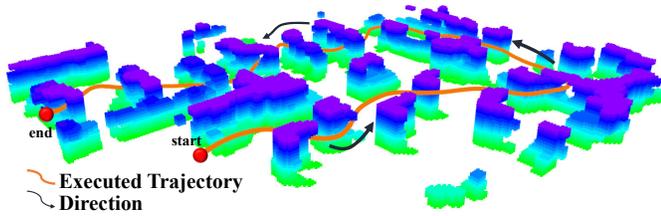


Fig. 13: Trajectories executed by the quadrotor and online generated map which is colored by height in the system-level indoor experiment.



Fig. 14: Two third-person views of the indoor experiment. (a): the first half of the experiment. The blue line is the prescribed topological path. (b): the second half of the experiment.

## VII. CONCLUSION

In this paper, we investigate the high correlation between gaze and intent and propose a gaze-enhanced perception-aware assistive aerial teleoperation. With the gaze-enhanced intent capture method, we can capture the intent more precisely. Then we generate a perception-aware trajectory that accompanies the intent timely and safe. Extensive indoor and outdoor experiments and benchmark comparisons validate that our method is robust and friendly to even unskilled users.

## REFERENCES

- [1] M. F. Land and D. N. Lee, "Where we look when we steer," *Nature*, vol. 369, no. 6483, pp. 742–744, 1994.
- [2] A. N. Belkacem, S. Saetia, K. Zintus-art, D. Shin, H. Kambara, N. Yoshimura, N. Berrached, and Y. Koike, "Real-time control of a video game using eye movements and two temporal eeg sensors," *Computational intelligence and neuroscience*, vol. 2015, 2015.
- [3] C. Pfeiffer and D. Scaramuzza, "Human-piloted drone racing: Visual processing and control," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 3467–3474, 2021.
- [4] X. Yang and N. Michael, "Assisted mobile robot teleoperation with intent-aligned trajectories via biased incremental action sampling," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 10998–11003.
- [5] J. Israelsen, M. Beall, D. Bareiss, D. Stuart, E. Keeney, and J. van den Berg, "Automatic collision avoidance for manually tele-operated unmanned aerial vehicles," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2014, pp. 6638–6643.
- [6] X. Yang, J. Cheng, and N. Michael, "An intention guided hierarchical framework for trajectory-based teleoperation of mobile robots."
- [7] E. Heiden, K. Hausman, G. S. Sukhatme, and A.-a. Agha-mohammadi, "Planning high-speed safe trajectories in confidence-rich maps," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2017, pp. 2880–2886.
- [8] B. Zhou, J. Pan, F. Gao, and S. Shen, "Raptor: Robust and perception-aware trajectory replanning for quadrotor fast flight," *IEEE Transactions on Robotics*, vol. 37, no. 6, pp. 1992–2009, 2021.
- [9] Q. Wang, Y. Gao, J. Ji, C. Xu, and F. Gao, "Visibility-aware trajectory optimization with application to aerial tracking," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2021, pp. 5249–5256.
- [10] E. R. Boer, "Tangent point oriented curve negotiation," in *Proceedings of conference on intelligent vehicles*. IEEE, 1996, pp. 7–12.



Fig. 15: Flight in forest, the blue box marks a ring and yellow box marks a gate made from two trees pasted with white paper.

TABLE I: Indoor System-level Comparison

System	Avg Time (s)	Max speed (m/s)
Ours	<b>7.5</b>	<b>1.97</b>
DJI Mavic Air 2	31.2	0.8

- [11] O. Lappi, E. Lehtonen, J. Pekkanen, and T. Itkonen, "Beyond the tangent point: Gaze targets in naturalistic driving," *Journal of vision*, vol. 13, no. 13, pp. 11–11, 2013.
- [12] N. S. Negi, P. Van Leeuwen, and R. Happee, "Differences in driver behaviour between novice and experienced drivers: A driving simulator study," in *5th International Conference on Vehicle Technology and Intelligent Transport Systems, VEHITS 2019*. SciTePress, 2019, pp. 368–379.
- [13] Z. Hubenova, K. Metodiev, S. Dimitrova, and L. Alexiev, "Usage of eye tracker technology in examining attention distribution of operators of unmanned aerial vehicles," *Aerospace Research in Bulgaria*, vol. 33, 2020.
- [14] R. Valenti, N. Sebe, and T. Gevers, "What are you looking at?" *International journal of computer vision*, vol. 98, no. 3, pp. 324–334, 2012.
- [15] L. Shi, C. Copot, and S. Vanlanduit, "Application of visual servoing and eye tracking glass in human robot interaction: A case study," in *2019 23rd International Conference on System Theory, Control and Computing (ICSTCC)*. IEEE, 2019, pp. 515–520.
- [16] M. Nieuwenhuisen, D. Droschel, J. Schneider, D. Holz, T. Läbe, and S. Behnke, "Multimodal obstacle detection and collision avoidance for micro aerial vehicles," in *2013 European Conference on Mobile Robots*. IEEE, 2013, pp. 7–12.
- [17] M. Odelga, P. Stegagno, and H. H. Bühlhoff, "Obstacle detection, tracking and avoidance for a teleoperated uav," in *2016 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2016, pp. 2984–2990.
- [18] C. Richter and N. Roy, "Learning to plan for visibility in navigation of unknown environments," in *International Symposium on Experimental Robotics*. Springer, 2016, pp. 387–398.
- [19] A. Bundy and L. Wallen, "Breadth-first search," in *Catalogue of artificial intelligence tools*. Springer, 1984, pp. 13–13.
- [20] M. Ester, H.-P. Kriegel, J. Sander, X. Xu *et al.*, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *kdd*, vol. 96, no. 34, 1996, pp. 226–231.
- [21] Z. Wang, X. Zhou, C. Xu, and F. Gao, "Geometrically constrained trajectory optimization for multicopters," *arXiv preprint arXiv:2103.00190*, 2021.
- [22] A. S. Lewis and M. L. Overton, "Nonsmooth optimization via quasi-newton methods," *Mathematical Programming*, vol. 141, no. 1, pp. 135–163, 2013.
- [23] L. S. Jennings and K. L. Teo, "A computational algorithm for functional inequality constrained optimization problems," *Automatica*, vol. 26, no. 2, pp. 371–375, 1990.
- [24] W. H. Press, H. William, S. A. Teukolsky, A. Saul, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes 3rd Edition: The Art of Scientific Computing*. Cambridge University Press, 2007.
- [25] X. Zhou, Z. Wang, H. Ye, C. Xu, and F. Gao, "Ego-planner: An esdf-free gradient-based local planner for quadrotors," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 478–485, 2021.
- [26] X. Zhou, Z. Wang, X. Wen, J. Zhu, C. Xu, and F. Gao, "Decentralized spatial-temporal trajectory planning for multicopter swarms," *arXiv preprint arXiv:2106.12481*, 2021.