Active View Planning for Visual SLAM in Outdoor Environments Based on Continuous Information Modeling

Zhihao Wang, Haoyao Chen, Member, IEEE, Shiwu Zhang, Member, IEEE, and Yunjiang Lou, Senior Member, IEEE

Abstract—Visual simultaneous localization and mapping (vSLAM) is widely used in satellite-denied and open-field environments for ground and surface robots. However, due to frequent perception failures derived from featureless areas or the swing of robot view direction on rough terrains, the accuracy and robustness of vSLAM are still to be enhanced. This article develops a novel view planning approach of actively perceiving areas with maximal information to address the mentioned problem; a gimbal camera is used as the main sensor. First, a map representation based on feature distribution-weighted Fisher information is proposed to completely and effectively represent environmental information richness. With the map representation, a continuous environmental information model is further established to convert the discrete information space into a continuous one for numerical optimization in real time. Subsequently, receding horizon optimization is utilized to obtain the optimal informative viewpoints with simultaneously considering the robotic perception, exploration, and motion cost based on the continuous environmental model. Finally, several simulations and outdoor experiments are performed to verify the improvement of localization robustness and accuracy by the proposed approach. We release our implementation as an open-source package for the community.

Index Terms—Active view planning, continuous information modeling, localization uncertainty representation, receding horizon optimization.

I. INTRODUCTION

VISUAL simultaneous localization and mapping (vSLAM) is widely used in outdoor and field environments for environmental monitoring, resource exploration, and lakeshore inspection since the solution is of low cost and rich perceptual information [1], [2]. However, outdoor environments, such as rough terrains and featureless water surface or ground, introduce difficulties for commonly used vSLAM algorithms [3] to achieve robust performance. For example, when the surface robot is floating along the lakeshore for inspection or cleaning, the featureless water surface will occupy a large field of view (FOV); when the ground robot is moving uphill, the perception system on the robot may tilt with the body to look toward the sky. These conditions easily lead to lost features and failure of localization. The inertial measurement unit measurements and vision are fused to deal with feature loss in featureless and dynamic environments [4]. However, robot pose tracking often fails when working for a long time in the complex environments due to the loss of visual tracking [5]. Likewise, large-FOV cameras that can observe more features [6] are utilized for robust tracking in featureless scenes; however, the loss of angular resolution for higher FOVs is drastically amplified by the higher depth range in outdoor environments [7], leading to worse tracking performance than the perspective camera.

Animals with necks can turn their view flexibly to areas of interest. Similarly, actively controlling sensors like a camera to look at the places with rich features will benefit robotic navigation performance. Besides, the robotic trajectory may be defined by the operator in remote monitoring and operating applications like lakeshore inspection. Thus, the view planning problem—where to look and how to look under the predefined trajectory—is essential to improve the estimation accuracy [8]. To address the problem, this article utilizes a gimbal camera that can turn the camera toward the best view with minimal estimation uncertainty. Note that the whole-body planning of camera view and robotic trajectory together is another research topic and is out of the research scope of this article.

Generally, for active vSLAM, three challenging issues need to be considered [9]: 1) how to represent the robotic state estimation uncertainty with respect to the camera view; 2) how to evaluate the performance of the candidate views to consider the tradeoff of viewing new places (i.e., exploration) and reducing...
the estimation uncertainty by reviewing known feature points (i.e., exploitation); and 3) how to select the next best view in all candidate views in real time.

The estimation uncertainty depends on many factors such as texture, illumination [10], and dynamic or static environmental objects. Various indicators, such as feature numbers [11] and Fisher information [12], are generally used for evaluating the estimation uncertainty. The feature-number-based solution utilizes the number of feature points in the view to represent the estimation uncertainty. However, it is difficult to quantify the uncertainty of each feature point. The Fisher-information-based solution is able to quantify the uncertainty meaningfully. For information computation, the existing approaches based on feature points [12] need to calculate the Jacobian matrix of each feature, resulting in computation efficiency suffering. Localization information is summarized into voxels by using Fisher information field to accelerate the computation for online planning, but the building time of the field is still higher due to the iteration of each feature [13]. Voxelization [14] can be used to accelerate the map building process by downampling the feature points, but voxelization concentrates on the occupancy of each voxel and neglects point distribution in the voxel. This neglects results in the ambiguity of estimation uncertainty representation when calculating Fisher information with voxels. The ambiguity is described in Section III in detail. Therefore, the accurate and efficient representation of environmental estimation uncertainty for online mapping and planning remains challenging.

Besides information representation, the evaluation of candidate views for robotic exploration/exploitation is also an essential problem for active simultaneous localization and mapping (SLAM). For exploration, many researchers utilize lasers or cameras to detect geometric frontiers or calculate information gain to plan the sensor movement [15]; the purpose is to completely explore the unknown environments as soon as possible. For exploitation, some approaches plan a feature-rich trajectory to minimize the state estimation uncertainty with a camera fixed on a quadrotor [16] or a gimbal camera on a mobile robot [17]. However, due to the greedy consideration of the minimal localization uncertainty, the existing methods [16], [17] suffer from continuously revisiting known areas without exploring unknown areas. This leads to the degeneration [17] or local minimum problem [9] of planning, especially for view planning in unknown environments. Some approaches have been developed by utilizing a mode switching mechanism [8], [9], [18] to address the exploitation–exploitation dilemma. A solution presented in [8] involved switching the exploitation mode to the exploration mode when sufficient landmarks are successfully detected. Different weights were assigned to the modes in [9] and [18], and balancing the SLAM uncertainty reduction and area coverage task were performed well. By contrast, this study utilizes motion consistency [19] as an exploration indicator to deal with the degeneration problem of view planning. The motion consistency produces attractive force to keep the next best view being consistent with the motion direction. Subsequently, the evaluation of candidate views considers both the exploitation of information and the exploration indicator in an objective function; this makes it possible to solve the exploration–exploitation dilemma with continuous planning.

Based on the evaluation of candidate views, sample-based methods, e.g., RRT* [17] or dynamic movement primitive [20], were developed to select the best view with maximal utility. The methods suffer from discontinuous motion due to discrete sampling. Some researchers [21], [22] realized continuous-space planning by maximizing the utility of candidate states. However, the utility function of informative path planning is always high dimensional, nonlinear, and nonconvex, which makes solving the optimization problem of maximizing the utility difficult. The evolutionary algorithm is used to solve the complex optimization problem [21], but it is time consuming. Gradient descent is an efficient optimization technology, but it is hard to derive an analytic expression of the complex utility function’s gradient [22].

This article aims to develop a novel approach by actively and smoothly controlling a gimbal camera equipped on the robot to realize robust and accurate SLAM in unknown outdoor environments. The three challenging issues mentioned above are solved efficiently in the proposed approach. The contributions of this article are threefold.

First, a novel map representation based on feature distribution-weighted Fisher information is proposed to store the localization uncertainty of environments. This new information map overcomes the ambiguity of the localization uncertainty representation of the traditional voxelization method. Our method makes the environmental information representation more accurate and efficient for active perception and further helps realize online information mapping and motion planning.

Second, a continuous information modeling method is proposed to map environmental information, such as localization uncertainty around the robot, into multiple polynomial functions. Polynomial functions provide analytic derivatives for environmental information with respect to the action space and, therefore, solve the informative planning problem efficiently by numerical optimization with less time consumption.

Third, an information-gradient-based local view (IGLOV) planner is proposed to plan the optimal camera views in real time for obtaining maximal environmental information. The planner realizes active view planning by considering estimation uncertainty, exploration for avoiding degeneration, and motion smoothness constraints simultaneously. The experiments illustrate that our approach outperforms state of the art.

II. SYSTEM OVERVIEW

This article considers a ground or surface robot traveling along predefined trajectories in unknown outdoor environments, as illustrated in Fig. 1. A camera is equipped on the robot through a two-axis gimbal. To guarantee the robustness of trajectory tracking, robust vSLAM is required by automatically changing the camera perception direction to achieve stable state estimation. Therefore, the problem to be addressed in this article is formulated as

\[
\mathbf{u}^* := \arg \max_{\mathbf{u}} f (\mathbf{\xi}_{sv}, M, \mathbf{u})
\]

\[
\text{s.t. } h(\mathbf{\xi}_{sv}, \mathbf{u}) \leq 0
\] (1)

where \(f(\cdot)\) denotes an objective function to be designed that quantifies the estimation accuracy of the camera pose
Fig. 1. Illustration of active view planning of a gimbal camera for vSLAM in an outdoor environment.

Fig. 2. Framework of the proposed active vSLAM approach.

\( \xi_{wc} \): \( M \) denotes the estimated map; \( u \), including yaw and pitch control commands, denotes the control vector of the gimbal to be optimized; \( u^* \) is the optimal control vector of \( u \); \( h() \) represents the constraints for the gimbal camera; and \( \xi_{wc} = [\chi_{wc}, \phi_{wc}]^T \in \mathbb{R}^6 \) denotes the camera pose with respect to the world frame, where \( \chi_{wc} \in \mathbb{R}^3 \) and \( \phi_{wc} \in \mathbb{R}^3 \) denote translation and rotation, respectively.

III. ENVIRONMENTAL INFORMATION MAPPING

The information richness of the surrounding environment should be evaluated to guide camera view planning. Voxelization is used to efficiently represent the information richness of the environment by downsampling the feature point cloud. However, representation with voxelization raises a new problem in active perception. The information calculated from the voxel map represents the localization uncertainty of each voxel but neglects the effect of the feature number in the voxel and the feature distribution around the voxel on the localization uncertainty.

Especially, for feature-based vSLAM, the number of features and their distribution uniformity directly affect the tracking accuracy [4]. Specifically, the neglect of feature distribution results in feature-dense areas that have the same information as the feature-sparse areas; however, feature-dense areas actually contain more features for tracking than feature-sparse areas. This problem is called as ambiguity in the representation of estimation uncertainty brought by voxelization. To deal with the ambiguity problem, feature distribution information, including the feature number in the voxel and the feature distribution around the voxel, is integrated with the Fisher information of a voxel. The integration of the two kinds of information is realized by a new map representation method to evaluate the information richness of the environment.

A. Calculation of Fisher Information

The Fisher information matrix (FIM) indicates the Cramér–Rao lower bound, the smallest covariance of an unbiased estimator [23]. Therefore, the FIM is usually used to represent the estimation uncertainty in many robotic applications, such as feature selection [24].

As our approach focuses on the view planning based on the known feature point maps obtained from SLAM, the uncertainty of camera pose estimation is evaluated from the perspective of observation uncertainty to decide the next best view. The observation uncertainty is evaluated by the Fisher information about estimating the camera pose \( \xi_{wc} \). The observation \( z_i \) of \( p_i^w \) at the camera pose \( \xi_{wc} \) is modeled as

\[
z_i = g(\xi_{wc}, p_i^w) + \omega \tag{2}
\]

where \( g(\cdot) \) is derived from the camera’s measurement model; here, the bearing vector model is implemented as the measurement model [25] and is defined as

\[
g(\xi_{wc}, p_i^w) = p_i^f/\|p_i^f\|_2, p_i^f = (\exp(\xi_{wc}h^i))^T p_i^w \tag{3}
\]

where \( p_i^w \) and \( p_i^f \) denote the center position of the \( i \)th voxel \( V_i \) in the world coordinate frame and the camera coordinate frame, respectively, \( i = 0, 1, \ldots, M \), and \( M \) is the number of occupied voxels of the voxel map \( M \); \( \omega \) denotes white noise with covariance \( Q \). Then, the FIM \( J_i \in \mathbb{R}^{6x6} \) evaluates the estimation uncertainty derived from the observation of \( p_i^w \) at \( \xi_{wc} \) and is
defined as

$$I_i = J_i^T Q^{-1} J_i$$

(4)

where $J_i = \partial g / \partial \xi_{wc}$ denotes the Jacobian of the observation function $g(\cdot)$ with respect to $\xi_{wc}$. For detailed deduction, refer to [26]. Because $\{I_0, \ldots, I_M\}$ is a series of matrices, the memory usage increases rapidly with environmental exploring. A common way to evaluate voxels in terms of estimation accuracy is based on the theory of optimal experimental design (TOED) [27]. The TOED utilizes the T-opt optimality criterion, i.e., the trace of the FIM, to convert the matrix to a scalar metric for reducing memory usage. Moreover, it has been proved that the FIM without the visibility constraint is rotation invariant [25], which means the FIM only relates to the camera position $\chi_{wc}$ and is not concerned with the camera rotation $\phi_{wc}$. Therefore, we finally define the Fisher information metric $I_i^F$ as

$$I_i^F(\chi_{wc}, p_i^w) = \text{trace}(I_i)$$

(5)

where $I_i^F(\chi_{wc}, p_i^w)$ represents the estimation uncertainty of $\chi_{wc}$ when observing $V_i$ at $\chi_{wc}$.

### B. Calculation of Distribution-Weighted Fisher Information

The uniformity of feature distribution affects the tracking accuracy in the feature-based SLAMs. The statistics of features in the neighbor voxel set $S_{ne}$, which consisted of 27 neighbor voxels around $V_i$, are used to formulate the uniformity of feature distribution. Specifically, the mean and the standard deviation of the feature number in $S_{ne}$ are defined as

$$\mu_i = \frac{1}{N_{ne}} \sum_{k=1}^{N_{ne}} N_k$$

(6)

$$\sigma_i = \sqrt{\frac{\sum_{k=1}^{N_{ne}} (N_k - \mu_i)^2}{N_{ne}}}$$

(7)

where $N_{ne}$ is the element number in $S_{ne}$ and $N_k$ denotes the feature number within the $k$th neighbor voxel $V_k$ in $S_{ne}$. The uniformity $I_i^D$ of feature distribution around $V_i$ is defined as

$$I_i^D = \mu_i \cdot (1 + e^{-\sigma_i}).$$

(8)

Because the calculation of Fisher information is based on discrete voxels, which neglects the feature number and distribution around the voxel, we complement the Fisher information with the feature distribution of each voxel for accurate representation. $I_i^F$ concentrates on the estimation uncertainty of the camera pose in a voxel, while $I_i^D$ focuses on the local feature distribution around the voxel, which explicitly represents the localization uncertainty from the perspective of feature matching and tracking. Then, information in (5) and (8) is fused as

$$I(\chi_{wc}, p_i^w) = I_i^D \cdot I_i^F(\chi_{wc}, p_i^w)$$

(9)

where $I(\chi_{wc}, p_i^w)$ denotes the distribution-weighted Fisher information of $\chi_{wc}$ when observing $V_i$ at $\chi_{wc}$.

Based on (9), the distribution-weighted Fisher information by considering both the Fisher information and the feature distribution is able to correctly quantify the estimation uncertainty of camera poses. It provides the essential information metric for the camera view planning module in the following section.

### IV. Camera View Planning

Literature [18] has shown that the known areas of the map have more known features for tracking and contribute to low uncertainty, whereas unknown areas generally have fewer features and lead to high uncertainty. Therefore, the maximum-information-based solution allows the robot to only revisit the known areas for robust localization. This condition may lead to the degeneration of navigation, especially for active camera view planning in unknown environments. To address the problem, we develop a novel IGLOV planner to actively minimize localization uncertainty while considering the degeneration simultaneously. The IGLOV planner contains four main parts, i.e., generating sample points, evaluating information gain, conducting polynomial regression, and receding horizon optimization.

As the gimbal camera has only two degrees of freedom, solving the inverse kinematic is convenient. Furthermore, its view planning in task space can benefit the motion prediction and handling of environmental perception. Therefore, the planner optimizes the view landing points in task space; the view landing point means the intersection of the terrain surface and the camera optical axis. Followed by the gimbal’s inverse kinematics, the view landing point will be transformed into desired gimbal rotation angles. Besides, “point” is used to denote “view landing point” for simplification in the rest of this article.

#### A. Generation of Sample Points

As shown in Fig. 3, the red star denotes the current pose of the robot, and frames $W$ and $B$ represent the world and the robot base coordinate frames, respectively. The $y$-axis of $X_B - O_B - Y_B$ is along the robot’s moving direction. The anchor points are first generated along the $y$-axis. Moreover, each anchor point $a_p_i$ corresponds to a sample circle $C_i$ for generating the sample points. Sample points are uniformly generated from $a_p_i$ to both sides along the sample circle $C_i$ with radius $r_i$ and angle interval $\Delta_\theta$. The $j$th sample point on the $i$th sampling circle $C_i$ is denoted

![Fig. 3. View-landing-point sampling process in the IGLOV planner.](image-url)
as \( p_{i,j} \), and its coordinates w.r.t. the world frame is given as

\[
p_{i,j} = \mathbf{x}_{\text{wc}} + [r_i \sin(\theta_{i,j} + \theta_b), r_i \cos(\theta_{i,j} + \theta_b), 0]^T
\]  

(10)

where \( r_i = P_{\text{min}} + i \cdot \Delta d, i = 0, 1, \ldots, N_{\text{AP}} \). \( P_{\text{min}} \) is the minimal range for the anchor points, \( \Delta d \) is the distance interval between two neighboring anchor points along the radial direction, and \( N_{\text{AP}} \) is the number of anchor points. \( \theta_{i,j} = k \cdot \Delta \theta, k = -N_{\text{SP}}/2, \ldots, -1, 0, 1, \ldots, N_{\text{SP}}/2, j = 0, 1, \ldots, N_{\text{SP}}, N_{\text{SP}} \) is the number of sample points along \( C_i \). \( \theta_{i,j} \) is the angle between \( p_{i,j} \) and \( \theta \). \( \Delta \theta \) is the angle interval between two neighboring sample points (e.g., \( p_{i,j-1} \) and \( p_{i,j} \)). \( \theta_b \) is the yaw angle of the robot base calculated from \( \phi_{wb} \) and \( \phi_{wb} \) is the orientation of the robot base.

### B. Evaluation of Information Gain

To evaluate the information gain of the sample points, a function is constructed with considering both the localization uncertainty and degeneration. The localization uncertainty is quantified by the information defined in (9). Furthermore, the consistency between the view direction of the gimbal camera and the motion direction of the robot base is considered to deal with degeneration. Thus, the information gain function of the sample point \( p_{i,j} \) is defined as

\[
g_{i,j} = I(\mathbf{x}_{\text{wc}}, p_{i,j}) - \lambda_d \cdot |\theta_{i,j}|
\]  

(11)

where \( I(\mathbf{x}_{\text{wc}}, p_{i,j}) \) denotes the information of the voxel \( p_{i,j} \) calculated by (9), \( |\theta_{i,j}| \) denotes the absolute value of \( \theta_{i,j} \), and \( \lambda_d := I(\mathbf{x}_{\text{wc}}, p_{i,j})/\pi \) is a dynamic weight coefficient for balancing the two terms into the same magnitude. The term \( |\theta_{i,j}| \) represents the consistency between the view direction corresponding to the sample point \( p_{i,j} \) and the robotic motion direction, as demonstrated in Fig. 3. Thus, it penalizes the sample points deviating from the motion direction.

Equation (11) calculates the information gain of the single view landing point \( p_{i,j} \); however, it raises the problem that the information of a single point cannot wholly represent the information within the camera’s FOV. To this end, we consider the information around the sample point \( p_{i,j} \) by involving the neighboring voxel set \( S_{np} \) within a predefined distance threshold \( d_n \). Through the addition of \( S_{np} \) into the information gain function, the planner takes the 3-D environmental information into consideration even though with a 2-D sampling method. Finally, the information gain function (11) is rewritten as

\[
g_{i,j} = \sum_{p_{np} \in S_{np}} I(\mathbf{x}_{\text{wc}}, p_{np}) - \lambda_d \cdot |\theta_{i,j}|
\]  

(12)

where \( p_{np} \) denotes an element in the set \( S_{np} \).

### C. Polynomial Regression for Continuous Information Gain

Camera view planning is modeled as an optimization problem by evaluating the points in the task space. However, the information gain function defined in (12) is discrete. For the convenience of numerical optimization, the function to be optimized must be continuous and differentiable. To address the problem, we apply the methodology of polynomial regression to approximate a continuous and differentiable function about the environmental information. Compared with Gaussian process regression and other surface fitting methods, the polynomial regression for each curve is efficient, with the mathematics expression being differentiable. After the polynomial regression, the best point of the local environment can be obtained by searching the point with the maximal information gain in multiple polynomial curves. Besides, the optimal result based on multiple curves is approximately equivalent to the result on a surface when the distance \( \Delta d \) between two adjacent curves is small enough.

For each sample circle \( C_i \), a polynomial function \( f^b_i \) is obtained by fitting the information gain of all the discrete sample points \( p_{i,j} \) along \( C_i \). The position of \( p_{i,j} \) is determined by \( \theta_{i,j} \) from (10), and thus, the gain \( g_{i,j} \) relates the angle \( \theta_{i,j} \) according to (12). Therefore, we define the continuous information gain function \( f^c_i(\vartheta) \) as

\[
f^c_i(\vartheta) = a_{i,0} + a_{i,1} \cdot \vartheta + a_{i,2} \cdot \vartheta^2 + \cdots + a_{i,n} \cdot \vartheta^n
\]  

(13)

where \( \vartheta \) denotes the domain of function \( f^b_i \), denotes the continuous scanning angle variable along \( C_i \), as shown in Fig. 3. \( A_i := (a_{i,0}, a_{i,1}, \ldots, a_{i,n})^T \in \mathbb{R}^{n+1} \) denotes the stacked weight parameters to be calculated, and \( n \) denotes the degree of the polynomial function.

The information gain values of all the sample points along \( C_i \) should fit the function (13). By stacking the relative sample points, we have

\[
G_i = \Theta_i A_i
\]  

(14)

where \( \Theta_i \) denotes the Vandermonde matrix of \( \theta_{i,j}, j \in (0, 1, 2, \ldots, N_{\text{SP}}) \), given as

\[
\Theta_i := \begin{bmatrix}
1 & \theta_{i,0} & \theta_{i,1,0} & \cdots & \theta_{i,n,0} \\
1 & \theta_{i,1} & \theta_{i,1,1} & \cdots & \theta_{i,n,1} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
1 & \theta_{i,N_{\text{SP}}} & \theta_{i,N_{\text{SP}}-1} & \cdots & \theta_{i,n,N_{\text{SP}}}
\end{bmatrix}
\]  

(15)

\[
G_i := (g_{i,0}, g_{i,1}, \ldots, g_{i,N_{\text{SP}}})^T
\]

The least squares method is utilized to solve \( A_i \), and we have

\[
A_i = (\Theta_i^T \Theta_i)^{-1} \Theta_i^T G_i
\]  

(16)

The solution of the weight parameters \( A_i \) is analytical, contributing to computationally efficient. By fitting curves for each sample circle \( C_i \), \( A_i \) polynomial functions are obtained for representing the information gain about the local environment. Fig. 4 illustrates the process of information curves fitting. The best point \( \vartheta^b_i \) with the maximal information gain of the single polynomial function \( f^b_i \) is obtained by

\[
\vartheta^b_i = \arg \max \ f^b_i(\vartheta).
\]  

(17)

The solution of (17) is obtained with the gradient descent method. Note that \( f^b_i(\vartheta) \) is a nonconvex function, suffering from the local minimal problem. Fortunately, according to the processes in Sections IV-A and IV-B, the best sample point \( ^*p^b_i \) is easily obtained by comparing all the sample points along \( C_i \). \( ^*p^b_i \) provides a reliable initial value for numerical optimization.
planning of the camera. Then, a method based on receding horizon optimization is developed to maximize the environmental information of future point sequence and minimize motion smoothness cost between neighbor points in the sequence.

Because the trajectory of robot base is a priori given, the future robotic positions in L steps, defined as \( \{ \chi_{k+1}, \ldots, \chi_{k+L} \} \), are available from the trajectory. Note that even if the trajectory is unknown, the robot can predict the future positions by a constant velocity motion model. We assume that the voxel map \( M \) maintains the same during the time of the horizontal sliding window. Then, the optimization problem including the information gain and motion smoothness is defined as

\[
p_{k+1:k+L}^b = \arg \min_{p_{k+1:k+L}} \lambda_{\text{info}} J_{\text{info}} + \lambda_{\text{simo}} J_{\text{simo}}
\]

where \( p_{k+1:k+L} \) denotes the view landing points to be optimized in the horizontal sliding window, \( J_{\text{info}} \) denotes the information gain, \( J_{\text{simo}} \) denotes the smoothness cost penalizing trajectory discontinuity, and \( \lambda_{\text{info}} \) and \( \lambda_{\text{simo}} \) are weight coefficients to balance the two terms \( J_{\text{info}} \) and \( J_{\text{simo}} \) respectively—the more significant value of one coefficient than the other, the more concerned about the related term. The values of \( \lambda_{\text{info}} \) and \( \lambda_{\text{simo}} \) were set empirically. From (18), the information gain term \( J_{\text{info}} \) is defined as

\[
J_{\text{info}} = \sum_{t=k+1}^{k+L} \left( \frac{1}{2} F(\chi_{t}^{\text{wc}}, M, \tilde{p}_{t})^2 \right)
\]

This term constrains the neighbor displacement vectors in both the direction and length. The cost indicates smoothness and distance distribution of \( \tilde{p}_{k+1}, \tilde{p}_{t}, \) and \( \tilde{p}_{t-1} \).

Since the complex information gain function in (12) is converted to the formulation in (18), which is differentiable, the optimization problem (20) is solved by using the gradient descent method [28]. And the single-step optimized point \( \tilde{p}_{t}^b \) by (19) is used as initial state. \( \tilde{p}_{t} \) iterates as follows:

\[
\tilde{p}_{t} := \tilde{p}_{t} - \left( \lambda_{\text{info}} \cdot \frac{\partial J_{\text{info}}}{\partial \tilde{p}_{t}} + \lambda_{\text{simo}} \cdot \frac{\partial J_{\text{simo}}}{\partial \tilde{p}_{t}} \right).
\]

The gradient of \( J_{\text{info}} \) with respect to \( \tilde{p}_{t} \) is calculated as

\[
\frac{\partial J_{\text{info}}}{\partial \tilde{p}_{t}} \iff \frac{\partial J_{\text{info}}}{\partial \theta_{t}} \iff \frac{\partial f_{t}^b}{\partial \theta_{t}} = \sum_{s=1}^{n} a_{i,s} \cdot (\theta_{t,s})^{s-1}
\]

where \( \iff \) denotes logical equivalence; the first \( \iff \) indicates that \( \tilde{p}_{t} \) and \( \theta_{t} \) represent the same point, and the second \( \iff \) indicates that only one polynomial function \( f_{t}^b \) in \( F \) relates to the variable \( \tilde{p}_{t} \) when \( \delta(\tilde{p}_{t}) = 1 \) according to (18), because \( f_{t}^b \) is fitting from the sampling circle \( C_{t} \) where \( \tilde{p}_{t} \) lies in and \( C_{t} \) is determined by

---

**Fig. 4.** Illustration of information gain curve fitting. (a) Yellow squares denote the sample points \( p_{i,j} \). The example environment includes feature-rich grass (left area) and featureless ground (right area). (b) Sample points are transferred into polar coordinates with \( \theta \) and \( r \) as horizontal and vertical axes. The units of \( r \) and \( \theta \) axes are m and rad, respectively. (c) Information gain of the sample points. (d) Each curve is fitted as a polynomial with degree \( m = 6 \) from the sample points along corresponding sampling circle. (a) Sampling in Cartesian frame, (b) Sampling in polar frame. (c) Gain of sample points. (d) Information gain curves.
the distance between $\tilde{p}_t$ and $\chi^w_k$ through (10). The gradient of $J_{\text{smo}}$ with respect to $\tilde{p}_t$ is calculated as
\[
\frac{\partial J_{\text{smo}}}{\partial \tilde{p}_t} = -4 \left( \tilde{p}_{t+1} - 2 \cdot \tilde{p}_t + \tilde{p}_{t-1} \right).
\]

For receding horizon optimization, the optimizing variable is the point sequence $\tilde{p}_{k+1:k+L}$, where each point iterates according to (23). And the iteration step index of (23) is omitted for simplicity of statement. After iterating $\tilde{p}_{k+1:k+L}$ until convergence, the optimal solution of the best point sequence is obtained and remarked as $\hat{p}_{k+1:k+L}$. The solution balanced information gain and motion smoothness in the horizontal time window. Finally, the first point $\hat{p}_{k+1}^b$ in $\hat{p}_{k+1:k+L}$ is selected as the next desired best point. The bottom tracking controller is then utilized to output control vector $u^*$ according to the inverse kinematics and drive the gimbal camera toward $\hat{p}_{k+1}^b$.

Furthermore, the view direction may swing with the robot base in practical field environments when moving on rough terrains. The gimbal’s bottom controller is used not only to track the best point but also to improve the localization failure problem caused by the view direction swing by controlling the gimbal’s pitch. Moreover, the controller outputs control commands at a high frequency with 100 Hz.

V. SIMULATIONS AND EXPERIMENTS

A. Physics-Engine Simulation and Experimental Platforms

Several simulations and experiments were performed to verify the proposed approach. We built a mobile platform in the physics-engine-based simulator Gazebo and utilized an experimental terrain vehicle, as shown in Fig. 5. An RGB-D camera on the gimbal was equipped for perception in the simulations and experiments. Another RGB-D camera was fixed on the robot to compare the passive method. The gimbal is of two axes, with the pitch and yaw angles being controllable. A high-precision global navigation satellite system (GNSS) was used to provide ground truth in the experiments. The parameters are listed in Table I. $P_{\text{min}}$ was set 2 m, which can exclude the evaluation of the features too close to the camera.

B. Evaluation of Information Mapping Method

We first performed a simulation to evaluate the proposed information mapping method. Fig. 6(a) illustrates the top view of an example environment. The robot’s left side is texture-rich grass, and the right side is textureless ground. The robot with the fixed camera traveled along the trajectory shown as the green line in Fig. 6(a). As shown in Fig. 6(b), the information maps are built by different information formulations, i.e., Fisher information in (5), feature distribution information in (8), and distribution-weighted Fisher information in (9), but without considering the exploration factor. In the Fisher information map (Layer 1), the textureless ground near the robot also incorrectly provides high information values. Because the voxel focuses on the occupancy, the textureless ground can provide sparse feature to occupy the voxel. This results in the left texture-rich grass having the same Fisher information with the right textureless ground. The feature distribution information map (Layer 2) makes the visited area at the bottom left of Layer 2 to be informative area. The distribution-weighted Fisher information map (Layer 3) complements the Fisher information with the feature distribution information. It is seen that the information representation provided in Layer 3 is more accurate compared to the other two methods, because the area takes both the Fisher information and the feature distribution information into consideration according to (9).

Since the Fisher information in the above simulations considered the uncertainty of the voxel position by the covariance $Q$, we additionally performed simulations to verify the effects of feature uncertainty on the information mapping method. The feature uncertainty is represented by $Q^* = \text{diag}(\delta^2, \delta^2, \delta^2)$, where $\delta^2 = 0.25$ denotes the feature points’ pixel error and $\delta_z = 1.425 \times 10^{-5} \text{z}^2$ mm$^2$ depends on depth.
The proposed camera view planning approach for robust vSLAM includes information mapping and camera view planner IGLOV. The information mapping module has been verified in the previous section; therefore, we further designed several simulations to evaluate the performance of IGLOV planner.

**C. Evaluation of Camera View Planning Algorithm**

The proposed camera view planning approach for robust vSLAM includes information mapping and camera view planner IGLOV. The information mapping module has been verified in the previous section; therefore, we further designed several simulations to evaluate the performance of IGLOV planner. We
compared the proposed IGLOV planner with several existing planning methods, including the passive (PAS) method, uniform sampling in view space (USV) [31], Monte Carlo sampling in task space (MST) [32], and regular sampling considering degeneration in task space (RSDT). For fair comparison, the previous weighted Fisher information mapping method was used for these camera view planning approaches. The PAS method fixed the camera on the robot base and did not change the viewing direction. The USV method sampled ten views uniformly in $[−\pi, \pi]$ of yaw at each time stamp and evaluated each view to find the maximal information view. The MST method, also called random sampling method in [32], sampled 500 points around the robot within 5 m at each time stamp and evaluated each point to find the one with maximal information. USV and MST evaluate the information according to (9). The RSDT method sampled points according to (10) and selected the sample point with maximal information gain as the best view landing point according to (12). RSDT considered the consistency between the view and the motion direction but without applying polynomial regression and receding horizon optimization.

The first scene simulated a wild environment, as shown in Fig. 9(a). The red curve denotes a preplanned baseline trajectory of the robot; the blue and yellow dot curves are obtained by transforming the baseline trajectory with a positive offset $\Delta > 0$ and negative offset $\Delta < 0$, respectively. Positive offset zooms the trajectory to the feature-rich area, while negative offset shrinks the trajectory to the featureless area. The simulated robot moved along five trajectories with $\Delta = 0.6, 0.3, 0, -0.3, -0.6$ m. The simulation also investigated how the proportion of featureless areas in the image increased when $\Delta$ decreased, and this made the shaking rotation of planned camera view influence the localization accuracy more obviously. Fig. 10 shows that the yaw angle of the RSDT method changed with high-frequency ripples and brought sudden motions of the gimbal camera, degrading the feature matching and tracking of the SLAM. Thanks to the consideration of degeneration and motion smoothness in the horizontal time window, the IGLOV planner performed much more smoothly, and the camera views were almost consistent with the motion direction because the yaw angles were smaller than 0.5 rad. Moreover, the views also turned to the feature-rich regions by the informative planning; this makes the camera by IGLOV locate well with the smallest estimation error, as shown in Fig. 10.

The computation costs of different methods are compared in Fig. 11. The time costs of the USV and MST methods positively correlate with the map size; the sudden increase of computation time occurred at the first turn of the trajectory because many new map points were added to the maps. While the RSDT and IGLOV methods cost less time because the regular sampling method brings less sampling and evaluation. Besides,
Table II

| Method      | Max (m) | Mean (m) | Min (m) | RMSE (m) |
|-------------|---------|----------|---------|----------|
| PAS         | 0.783   | 1.036    | 0.343   | 0.194    |
| USV         | 0.954   | 1.234    | 0.413   | 0.219    |
| MST         | 0.139   | 0.139    | 0.053   | 0.073    |
| RSDT        | 0.139   | 0.139    | 0.053   | 0.073    |
| IGLOV       | 0.139   | 0.139    | 0.053   | 0.073    |

The bold values represent the best localization accuracy.

D. Experiments

Further, we designed several real-world experiments to verify the proposed approach with the experimental ground vehicle shown in Fig. 5(b). Because both the USV and MST methods led to SLAM failures in our experiments, we only presented the comparison between the passive method and the proposed one.

We performed the first experiment in a typical outdoor campus environment with the robot moving along the campus road, while the second experiment was performed on a hillside with rough grass terrain.

1) Experiment 1: The first experiment (Ex1) was performed along a trajectory about 500 m in Shenzhen University Town, as shown in Fig. 14. The estimation error under the two methods is shown in Table III. The IGLOV method performed better in mean and RMSE values. Compared to the simulation results in Section V-C, the localization accuracy of the passive method in the experiment was much closer to the IGLOV method. The reason is that the operating environment is urbanized and surrounded by feature-rich parterres, trees, and buildings; thus, the passive method also tracked well with sufficient features in each keyframe. However, when traveling along the trajectory, the IGLOV planner evaluated the information of the local environments and autonomously turned the camera view toward the areas with maximum information gain. Some typical planning results are demonstrated in Fig. 14. The IGLOV method always turned the camera toward the local feature-rich areas, like the parked bicycles at position P1, the high parterre at position P2, and the parked car at position P3. Although the improvement of localization accuracy was limited
in the urbanized environments, the proposed method efficiently guided the camera view to local feature-rich areas.

2) Experiment 2: We further designed an experiment (Ex2) in a wild terrain environment to evaluate the performance of the proposed approach. As shown in Fig. 1, the experiment was performed on a hillside; the trajectory length was about 130 m. The trajectory estimation results under the passive and the IGLOV methods, together with the ground truth, are shown in Fig. 1. It is seen that the estimated trajectory under the IGLOV planner (green line) is closer to the ground truth (red line) than that under the passive method (blue line). The proposed approach is designed based on task space and feature points in the local sampling range; this makes the gimbal camera pay more attention to the feature-rich areas once it appears in the sampling range and then decreases the localization error. Table III shows the results for quantitative analysis. vSLAM was run without loop closure. The IGLOV planner performed much better than the passive method; the reason is that the camera avoided the view to the featureless areas like the sky or the pavement and focused on local feature-rich areas like the steles or the trees.

The gimbal camera actively changed the view direction to the feature-rich regions shown in Fig. 15. Our approach stabilized the view direction toward the slope to complement the view area coverage and obstacle avoidance, IEEE/ASME Trans. Mechatron., vol. 27, no. 5, pp. 3440–3450, Oct. 2022.

VI. CONCLUSION

This article developed a novel approach to realize active VSLAM for ground or surface robots in challenging outdoor environments. An information mapping algorithm was first proposed to represent the environmental information richness efficiently; the algorithm made online active view planning possible. A continuous information modeling method that combined regular sampling and polynomial regression was proposed to map the environmental information around the robot into multiple polynomial functions. Based on the multiple polynomial functions, the informative planning problem was solved efficiently by numerical optimization. A receding-horizon-optimization-based method solved the view planning problem under degeneration and motion smoothness constraints. Finally, several physics-engine simulations and experiments in outdoor environments were performed. The comparisons to the existing state-of-the-art approaches verified the effectiveness of the proposed approach. Our future work will focus on integrating robotic trajectory planning into the active SLAM for autonomous environmental exploration and informative navigation by considering map uncertainty, estimation of motion noise, infrared distribution and traversability simultaneously. We will also study the robustness of vSLAM with the severe and high-frequency oscillations due to fast movement on rough terrains.

REFERENCES

[1] R. Duan, D. P. Paudel, C. Fu, and P. Lu, “Stereo orientation prior for UAV robust and accurate visual odometry,” IEEE/ASME Trans. Mechatron., vol. 27, no. 5, pp. 3440–3450, Oct. 2022.
[2] J. Liu, X. Li, Y. Liu, and H. Chen, “RGB-D inertial odometry for a resource-restricted robot in dynamic environments,” IEEE Robot. Autom. Lett., vol. 7, no. 4, pp. 9573–9580, Oct. 2022.
[3] R. Mur-Artal and J. D. Tardós, “ORB-SLAM2: An open-source SLAM system for monocular, stereo and RGB-D cameras,” IEEE Trans. Robot., vol. 33, no. 5, pp. 1255–1262, Oct. 2017.
[4] T. Qin, P. Li, and S. Shen, “VINS-mono: A robust and versatile monocular visual-inertial state estimator,” IEEE Trans. Robot., vol. 34, no. 4, pp. 1004–1020, Aug. 2018.
[5] Z. Wang, J. Zhang, S. Chen, C. Yuan, J. Zhang, and J. Zhang, “Robust high accuracy visual-inertial-laser SLAM system,” in Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst., 2019, pp. 6636–6641.
[6] H. Matsuki, L. Von Stumberg, V. Usenko, J. Stuckler, and D. Cremers, “Omnidirectional DSO: Direct sparse odometry with fisheye cameras,” IEEE Robot. Autom. Lett., vol. 3, no. 4, pp. 3693–3700, Oct. 2018.
[7] Z. Zhang, H. Rebecq, C. Forster, and D. Scaramuzza, “Benefit of large field-of-view cameras for visual odometry,” in Proc. IEEE Int. Conf. Robot. Autom., 2016, pp. 801–808.
[8] S. Frintrop and P. Jensfelt, “Attentional landmarks and active gaze control for visual SLAM,” IEEE Trans. Robot., vol. 24, no. 5, pp. 1054–1065, Oct. 2008.
[9] Y. Chen, S. Huang, and R. Fitch, “Active SLAM for mobile robots with area coverage and obstacle avoidance,” IEEE/ASME Trans. Mechatron., vol. 25, no. 3, pp. 1182–1192, Jun. 2020.
[10] Y. Wang, H. Chen, S. Zhang, and W. Lu, “Automated camera-exposure control for robust localization in varying illumination environments,” Auton. Robots, vol. 46, no. 4, pp. 515–534, 2022.
[11] X. Deng, Z. Zhang, A. Sintov, J. Huang, and T. Breit, “Feature-constrained active visual SLAM for mobile robot navigation,” in Proc. IEEE Int. Conf. Robot. Autom., 2018, pp. 7233–7238.
[12] K. Khouss, M. Giamou, G. S. Sukhatme, S. Huang, G. Dissanyake, and J. P. How, “Reliable graphs for SLAM,” Int. J. Robot. Res., vol. 38, no. 2/3, pp. 260–298, 2019.
[13] Z. Zhang and D. Scaramuzza, “Fisher information field: An efficient and differentiable map for perception-aware planning,” Aug. 2020, arXiv:2008.03324.
[14] A. Hornung, K. M. Wurm, M. Bemnestwitz, C. Stachniss, and W. Burgard, “OctoMap: An efficient probabilistic 3D mapping framework based on octrees,” Auton. Robots, vol. 34, no. 3, pp. 189–206, 2013.
[15] M. G. Jadidi, J. V. Miro, and G. Dissanyake, “Gaussian processes autonomous mapping and exploration for range-sensing mobile robots,” Auton. Robots, vol. 42, no. 2, pp. 273–290, 2018.
[16] Z. Zhang and D. Scaramuzza, “Perception-aware receding horizon navigation for MAVs,” in Proc. IEEE Int. Conf. Robot. Autom., 2018, pp. 2534–2541.
[24] T. D. Barfoot, *State Estimation for Robotics*. Cambridge, U.K.: Cambridge Univ. Press, 2017.

[25] A. N. Bishop, B. Fidan, B. D. Anderson, K. Do˘ garage, and Y. Gao, “Probabilistic RGB-D odometry based on points, lines and planes under depth uncertainty,” in *Proc. IEEE Int. Conf. Robot. Autom.*, 2019, pp. 5986–5992.

[26] Z. Zhang and D. Scaramuzza, “Beyond point clouds: Fisher information field for active visual localization,” in *Proc. Int. Conf. Robot. Autom.*, 2019, pp. 1488–1494.

[27] F. Pukelsheim, *Optimal Design of Experiments*. Philadelphia, PA, USA: SIAM, 2006.

[28] S. Boyd, S. P. Boyd, and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[29] P. F. Proenca and Y. Gao, “Probabilistic RGB-D odometry based on points, lines and planes under depth uncertainty,” *Robot. Auton. Syst.*, vol. 104, pp. 25–39, 2018.

[30] M. Grupp, “EVO: Python package for the evaluation of odometry and SLAM,” 2017. [Online]. Available: https://github.com/MichaelGrupp/evo

[31] R. Zeng, W. Zhao, and Y.-J. Liu, “PC-NBV: A point cloud based deep network for efficient next best view planning,” in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2020, pp. 7050–7057.

[32] N. Palomeras, N. Hurtós, M. Carreras, and P. Ridao, “Autonomous mapping of underwater 3-D structures: From view planning to execution,” *IEEE Robot. Auton. Lett.*, vol. 3, no. 3, pp. 1965–1971, Jul. 2018.

Zhihao Wang received the B.Eng. degree in mechanical engineering from the Wuhan University of Technology, Wuhan, China, in 2017, and the M.Eng. degree in control engineering from the Huazhong University of Science and Technology, Wuhan, in 2019. He is currently working toward the Ph.D. degree in robotics with the School of Mechatronics Engineering and Automation, Harbin Institute of Technology, Shenzhen, China. His research interests include robotic motion planning and control, 3-D exploration, and reconstruction.

Haoyao Chen (Member, IEEE) received the B.Eng. degree in mechatronics and automation from the University of Science and Technology of China, Hefei, China, in 2004, and the Ph.D. degree in robotics and automation from the University of Science and Technology of China and the City University of Hong Kong, Hong Kong, in 2009. He is currently a Professor with the Harbin Institute of Technology, Shenzhen, China, and the State Key Laboratory of Robotics and System, Harbin, China. His research interests include visual servoing, multirobot systems, motion control, and aerial manipulation.

Shiwu Zhang (Member, IEEE) received the B.S. degree in mechanical and electrical engineering and the Ph.D. degree in precision instrumentation and precision machinery from the University of Science and Technology of China (USTC), Hefei, China, in 1997 and 2003, respectively. He is currently a Professor with the Department of Precision Machinery and Precision Instrumentation, USTC. He is the author of more than 100 papers in different journals and conferences. His research interests include smart materials and their applications in bioinspired robots, amphibious robot, soft robots, and terrady-namics.

Yunjiang Lou (Senior Member, IEEE) received the B.S. and M.E. degrees in automation from the University of Science and Technology of China, Hefei, China, in 1997 and 2000, respectively, and the Ph.D. degree in electrical and electronic engineering from the Hong Kong University of Science and Technology, Hong Kong, in 2006. He is currently with the State Key Laboratory of Robotics and Systems and the Shenzhen Key Laboratory for Advanced Motion Control and Modern Automation Equipments, School of Mechatronics Engineering and Automation, Harbin Institute of Technology, Shenzhen, China. His research interests include motion control, mechanism design, compliant actuators, and industrial robots.