2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
October 14-18, 2024. Abu Dhabi, UAE

Robot Active Vision-Based Path Planning for Localization Improvement in Indoor Environments

Sotirios Barlakas[1], Dimitrios Alexiou[1], Kosmas Tsiakas[1,2], Dimitrios Katsatos[1], Ioannis Kostavelis[1,3], Dimitrios Giakoumis[1], Antonios Gasteratos[2], Dimitrios Tzovaras[1]

***Abstract*— Reliable and robust navigation of autonomous mobile robots in indoor environments faces significant challenges due to the absence of GPS, visual degradation, repetitive structures, illumination variations, and low texture. These factors adversely affect localization systems. Current robots often use a uniform navigation approach, regardless of the varying localization uncertainties within different indoor environments. In this paper, we propose a holistic, active vision-based path planning method that produces efficient trajectories, aiming to minimize localization error and enhance navigation performance. Specifically, we utilize a 3D model of an indoor environment to derive an Artificial Potential Field (APF) with its associated localizability scores that encapsulate both visual features' richness and fiducial markers' placement. APF is employed to direct a Kinematically Constrained Bi-directional Rapidly Exploring Random Tree (KB-RRT) planner towards the calculation of optimal paths, prioritizing high localization areas. Subsequently, we use an online weight-adaptive MPC-based approach that, apart from robust path planning and obstacle avoidance, guides the robot towards areas with the most robust visual features in order to further refine the localization error. The proposed framework has been extensively tested in both simulation and real-world experiments with a mobile robot in a visually challenging indoor environment.**

I. Introduction

Autonomous mobile robots tailored for hazardous environments, such as construction sites [1] or subterranean regions [2], offer the potential to automate laborious tasks that are currently performed by humans. Accurate estimation of the robot's position is essential for ensuring navigation robustness and operational effectiveness. Visual localization methodologies have seen widespread adoption in autonomous mobile robotic systems due to the affordability of camera sensors combined with the richness of information that can be obtained from image data [3]. However, visual localization retains severe challenges, especially in indoor unstructured, featureless or repetitive environments. The utilization of fiducial markers [4] achieves the reduction of visual ambiguity in space, while more recent active localization techniques [5] aim to select the appropriate navigation actions that maximize the SLAM capabilities. Typical path planning methods [6] are still widely used in robotic applications and occupancy grid representations sufficiently cover the needs for complex navigation requirements in stationary areas with adequate geometrical and visual features. However, changing environments with dynamic actors alleviate the performance of such approaches, enabling active and perception-aware methodologies [7] [8] to have a critical role in the operation of robotic platforms.

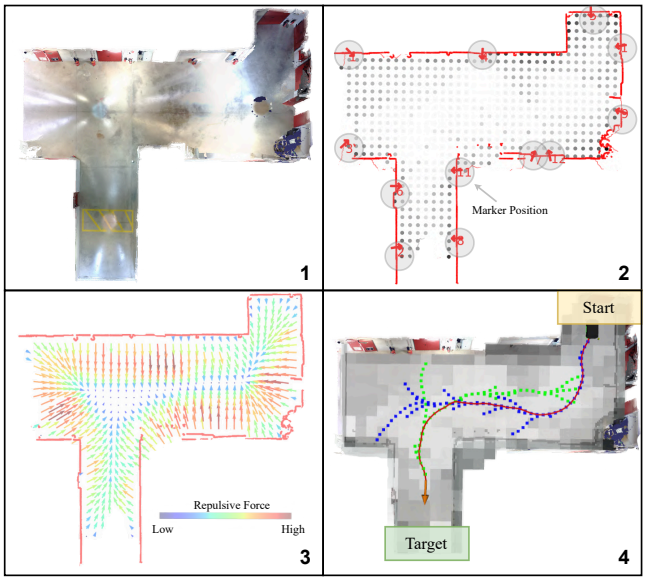


Fig. 1: Holistic view of proposed components: The input 3D model is featured in (1), while the corresponding localizability scores are shown in (2). The arrows in (3) delineate repulsive forces of the APF, whereas the RRT-based global path planning is showcased in (4).

Apart from dynamic scene elements, visual environmental features can directly affect real-time navigation efficiency, through the degradation of localization accuracy. To this end, additional contextual information should be taken into account for the optimization of path planning methods, such as the areas where the robot's field of view lacks descriptive visual features or the placement of fiducial markers. The generation of virtual replicas for physical entities, commonly mentioned as Digital Twins, has undergone great advancements [9] and offer a holistic representation of an environment. Their analysis can provide the necessary contextual information required. Recently, Huang et al. [10] presented the first framework that utilizes a 3D scene model for the extraction of optimal marker poses and the calculation of localizability scores within the operating environment.

Our method extends the outcomes of the above work

[1]Centre for Research and Technology Hellas / Information Technologies Institute (CERTH / ITI), Thessaloniki, Greece. sbarlakas@iti.gr
[2]Department of Production and Management Engineering, Democritus University of Thrace, Xanthi, Greece.
[3]Department of Supply Chain Management, International Hellenic University, Thessaloniki, Greece.

through the utilization of the localizability analysis for path planning purposes of a mobile robotic platform in a known environment, as shown in Fig. 1. Specifically, we investigate the conversion of localizability scores calculated by [10] into an APF that can facilitate a kinematically constrained, bi-directional RRT-based path planning algorithm. The localizability analysis provides insights on robot localization capabilities related to the static environment's visual features and fiducial markers' placement, however the dynamic nature of the environments must also be taken into consideration. For this reason, we apply real-time corrections to the initial plan via an adaptive MPC-based local planner. This module dynamically adjusts its weights depending on the presence and consistency of visually trackable features. The objective is to facilitate precise robot navigation within regions characterized by minimal visual localization uncertainty, while simultaneously following the initial plan and avoiding any potential obstacles.

The contributions of this paper are summarized as follows:

- A kinematically constrained, bi-directional RRT-based algorithm that utilizes the localizability scores of a known environment in an APF format.
- An MPC-based local planner, enhanced by an active robust feature tracking mechanism for localization enhancement.

The rest of the paper is organized as follows: In Section II, the current state-of-the-art on active robot path planning is presented. Section III provides a comprehensive description of the proposed method, while Section IV showcases the effectiveness of our approach with real-world and simulation experiments. Finally, Section V draws the conclusions of the present work.

II. Related work

Typical sampling-based robot path planning approaches [11] have met significant improvement in recent years, particularly regarding kinematic constraints [12], computational efficiency [13], and optimized obstacle avoidance capabilities [14]. Yet, the overall operation of autonomous mobile robots, and specifically the precise execution of computed paths requires robust localization performance. To this end, active vision-based path planning methodologies for boosting localization efficiency emerge to address this challenge [5].

Several works integrate classical path planning approaches with active vision mechanisms for localization enhancement [15]. Specifically, in [3] the authors propose an RRT*-based path planner that utilizes online-computed visual feature points and map points in order to generate optimal paths that minimize tracking failure for visual SLAM systems. The authors in [16] developed an APF-based local planner that conducts an active visual navigation scheme, based on visual feature extraction, that improves the localization accuracy. Dabin et al. [17] presented a topology-guided path planning method that enhances localization performance, thereby improving navigation precision. Each path is evaluated in respect to path length and visual information richness, requiring a pre-generated global map. A real-time perception-aware motion planning method is proposed in [18], which aims to improve robot navigation through the utilization of the localization uncertainty in the environment. In [19] an active localization approach that utilizes a topometric graph is developed. Moreover, the authors in [20] establish an MPC-based framework that simultaneously minimizes SLAM uncertainty and creates collision-free trajectories. In [21] the planning strategy comprises of a global planner, that generates target viewpoints that adequately observe visual features, alongside with a greedy local planner, which aims to maximize information acquisition from visual features within 3D deformable environments. Similarly, the authors in [22] implement an adaptive heading mechanism as part of a perception-aware local planner to improve feature tracking, thus localization accuracy. Finally, in [23], a dual-layer planning approach with a utility function to balance between exploration efficiency and localization performance is employed for trajectory generation.

Regarding deep learning-based approaches, [24] uses deep reinforcement learning to distinguish perception-informative regions within a scene, facilitating the development of a perception-aware navigation strategy aimed at improving localization accuracy for a UAV. Similarly, the authors in [25] propose a reinforcement learning-based navigation strategy to prevent localization failures. Supplementary, in [26], the authors propose an end-to-end differentiable and trainable method for selecting informative actions that maximize pose disambiguation on a reference map.

A new map representation, denoted as Fisher Information Fields (FIFs) and adapted for active vision-based path planning to enhance localization accuracy, was introduced in [27] and [28]. On that basis, in [29] an active vision-based path planning method is proposed that utilizes FIFs to plan trajectories that minimize 3D reconstruction uncertainty. Despite the efficiency of FIFs in providing a robust map representation that reflects localization capability, the influence of fiducial marker placement on localization performance is neglected. This aspect holds significant importance in addressing the indoor SLAM problem [10].

To the best of our knowledge, our work is the first one that utilizes the localizability scores [10] in the format of an APF, with the overall aim to facilitate an efficient RTT-based path planning that maximizes localization performance. Localizability scores encapsulate not only the richness of visual features within the environment, but also the influence of fiducial markers existence, and require as input solely a 3D model of the environment. In order to satisfy the requirements of constantly changing environments, we use the above extracted path as a basis, but we utilize an active vision, MPC-based, local planner that is capable of tracking areas with prominent and consistent visual features without diverging from the target goal to diminish further localization uncertainty. To this end, we propose a holistic active vision-based path planning framework consisting of both a global and a local planner, for generating efficient trajectories towards a given target and simultaneously minimizing localization uncertainty.

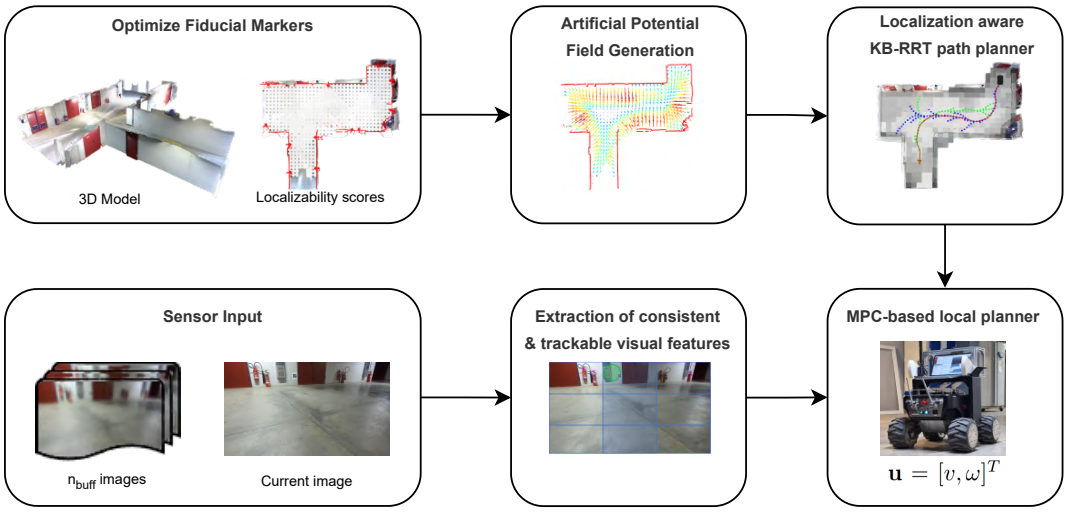


Fig. 2: Architecture of the proposed pipeline. An APF is generated based on the computed localizability scores that directs a KB-RRT path planner, while consistent and trackable visual features are extracted. Both modules serve as an input to an MPC-based local planner that generates control commands, decreasing localization uncertainty.

III. Proposed Method

The proposed method mainly consists of four separate components, with their interoperability being depicted in Fig. 2. The pipeline initiates with the creation of an APF using low localizability scores and obstacles, which acts as guidance for the KB-RRT planner, directing the expansion of the tree towards areas with high localizability and away from obstacles. Finally, the adaptive MPC-based local planner is responsible for path tracking and dynamic obstacle avoidance while taking into account the distribution of robust features during visual SLAM.

A. Artificial Potential Field Generation

Typically, the APF method consists of an attractive field that pulls robot motion towards the goal, and a repulsive field that drives the robot away from obstacles. For a given point q, the target point q_t creates an attractive field:

$$U^{att}(q) = \frac{1}{2} k_t \left\| q - q_t \right\|^2 \tag{1}$$

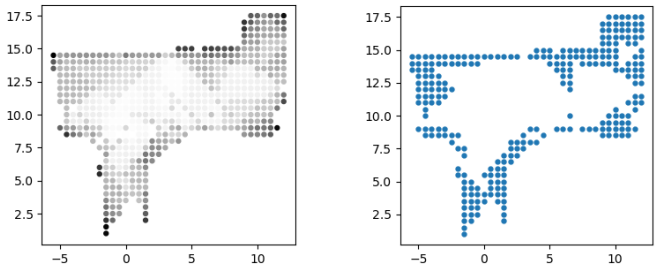
where k_t is a weight coefficient.

Each obstacle point q_i creates a repulsive field at point q:

$$U^{rep}_{o_i}(q) = \begin{cases} \frac{1}{2} k_o \left(\frac{1}{\|q-q_i\|} - \frac{1}{\rho_o} \right)^2, & \|q - q_i\| \leq \rho_o \\ 0, & \|q - q_i\| > \rho_o \end{cases} \tag{2}$$

where k_o is a weight coefficient and ρ_o is the range of action for the obstacle field.

In our approach, we introduce a third field $U^{rep}_l(q)$ that is generated by low localization regions. To compute the localizability scores of the operating environment, we rely on the work done in [10]. This step requires a precise and textured 3D model of the scene, in order to predict the optimal markers' placement positions, as well as the localizability scores for each position. This work is conducted offline, but [10] can be extended to dynamically update localizability scores, combined with online 3D mapping techniques. Hence, the pipeline can operate near real-time. These scores serve as an indicator of the localization uncertainty in a specific region, depending on the reliability of the visual features and the position of fiducial markers. Darker pixels indicate low localizability scores, while brighter ones represent high scores, as shown in Fig 3a.



(a) (b)

Fig. 3: (a) Raw (b) Filtered Localization Scores

Within the navigation context, we established a threshold to filter out cells with localization scores surpassing it, retaining only those with lower scores, as depicted in Fig. 3b. This approach is favored as straightforward and intuitive. Each score for position q_j contributes to the generation of the APF by generating a repulsive force analogous to $U^{rep}_{o_i}(q)$:

$$U^{rep}_{l_j}(q) = \begin{cases} \frac{1}{2} k_l \left(\frac{1}{\|q-q_j\|} - \frac{1}{\rho_l} \right)^2, & \|q - q_j\| \leq \rho_l \\ 0, & \|q - q_j\| > \rho_l \end{cases} \tag{3}$$

where k_l is a weight coefficient and ρ_l is the range of action for the localization field.

The total potential is the sum of the potentials generated by the target point, M obstacle points, and K low localization

areas.

$$U_{total}(q) = U^{att}(q) + \sum_{i=1}^{M} U_{o_i}^{rep}(q) + \sum_{j=1}^{K} U_{l_j}^{rep}(q) \tag{4}$$

The negative gradient of the total potential $-\nabla U_{total}$ indicates the most gainful direction of motion, which guides the expansion of the RRT.

B. Localization Aware KB-RRT Path Planner

Our path planning algorithm builds upon iKB-RRT [14], a kinematically constrained, bi-directional RRT method guided by an APF. As presented in [14], the APF guidance results in feasible path generation, reducing the number of sampled nodes and unnecessary space exploration. We extend this method by incorporating our localization-derived field. This extension serves the purpose of guiding the robot away from regions characterized by inaccurate localization estimates, thereby ensuring robust navigation while maintaining optimal path planning efficiency. The algorithm is presented below:

Algorithm 1: Localization Aware KB-RRT

```
Initialize trees T_start, T_goal with nodes X_start, X_goal
Get obstacles and localizability costmaps C_obst, C_loc
T_selected := T_start
for i ← 1 to N do
    X_rand ← Sample() the configuration space
    X_near ← FindNearest(X_rand, T_selected)
    X_guide ← APF(X_rand, X_near,
    X_new ← Steer(X_guide, X_near)
    if X_new is collision-free then
        Add X_new as a child node of X_nearest
        if T_start, T_goal distance reached threshold
          then
        end
        Swap T_start and T_goal
    end
end
```

The APF guiding step is explained in detail below. Given a node X_{near}, the negative gradient of the APF generates the magnitudes of the forces that act upon the particular node from goal node (Eq. 5), sample node (Eq. 6), obstacles (Eq. 7) and localization scores (Eq. 8).

$$|\vec{F}_{goal}^{att}(X_{near})| = k_g \left\| X_{near} - X_{goal} \right\| \tag{5}$$

$$|\vec{F}_{rand}^{att}(X_{near})| = k_s \left\| X_{near} - X_{rand} \right\| \tag{6}$$

$$|\vec{F}_{o_i}^{rep}(X_{near})| = \begin{cases} k_o \left(\frac{1}{d_i} - \frac{1}{\rho_o}\right) \frac{1}{d_i^2}, & d_i \le \rho_o \\ 0, & d_i > \rho_o \end{cases} \tag{7}$$

$$|\vec{F}_{l_j}^{rep}(X_{near})| = \begin{cases} k_l \left(\frac{1}{d_j} - \frac{1}{\rho_l}\right) \frac{1}{d_j^2}, & d_j \le \rho_l \\ 0, & d_j > \rho_l \end{cases} \tag{8}$$

where $d_i = \|X_{near} - q_i\|$, $d_j = \|X_{near} - q_j\|$ denote the euclidean distance between node X_{near} and obstacle at position q_i and low localization position q_j respectively. The total force $\vec{F}_{total}$ acting on the nearest node is computed by aggregating all force vectors:

$$\vec{F}_{total} = \vec{F}_{rand}^{att} + \vec{F}_{goal}^{att} + \sum_{i=1}^{M} \vec{F}_{o_i}^{rep} + \sum_{j=1}^{K} \vec{F}_{l_j}^{rep} \tag{9}$$

The guide node, X_{guide} is generated along the direction of the $\vec{F}_{total}$ and its coordinates are calculated as follows:

$$\begin{aligned} X_{guide_x} &= X_{near_x} + s\frac{F_x}{|\vec{F}_{total}|} \\ X_{guide_y} &= X_{near_y} + s\frac{F_y}{|\vec{F}_{total}|} \end{aligned} \tag{10}$$

where F_x and F_y are the decomposed forces of $\vec{F}_{total}$ along the X- and Y-axes of the global coordinate system, respectively, and s is the extension distance. This approach ensures that the tree expansion prioritises regions characterised by more reliable localization information, thereby enhancing overall navigation performance.

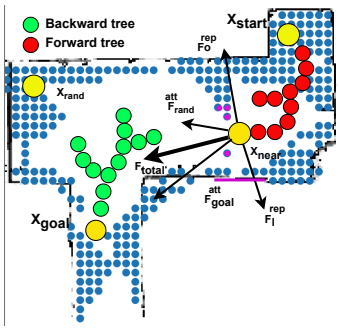


Fig. 4: Repulsive forces from nearby obstacles (magenta line) and localizability scores (magenta circles), and attractive forces towards X_{goal} and X_{rand} acting on X_{near}. The combined force $\vec{F}_{total}$ dictates the tree expansion direction.

C. Extraction of Consistent and Trackable Visual Features

Preventing failures in visual SLAM can be achieved by extracting robust visual features within the current goal's region, maximizing the probability of retaining trackable features, and enhancing SLAM performance. Incoming RGB data from an image buffer, containing n_{buff} images, is processed to extract ORB features, which are matched between the current frame and previous frames. Matches are filtered based on the Hamming distance, retaining only those that satisfy a distance criterion. Next, we compute the geometric homography transformation between images to exclude outliers from the trackable feature set that do not conform to this transformation. This filtering is applied consecutively between the current frame and each buffer image. Each matched image, of size (W, H) is split in

n_{seg} equal segments, forming a grid. For each segment, the percentage of matched features is calculated.

The spatiotemporal distribution of matched features in the previous frames and the image patches is encoded within a 2-dimensional matrix V of size $n_{buff} \times n_{seg}$. Each row of V corresponds to a buffer's frame, and each column represents a specific patch within the frame's grid layout, while V as a whole provides insights on the reliability and consistency of feature tracking. For all possible pairs of rows in V, the euclidean distance is calculated and normalized using the maximum distance within the $\mathbb{R}^{n_{seg}}$ space. Subsequently, a distance matrix D quantifies the similarity between consecutive image frames and its mean, $\bar{D}$, reflects the robustness of visual feature tracking in SLAM. Following the notion of [3] and [16], a heuristic threshold t is defined and the active localization mechanism is enabled when $\bar{D} < t$.

The above-mentioned mechanism entails mapping the current target of the global path onto image coordinates by leveraging stereo-derived depth estimation techniques. The selection of trackable visual features is performed solely on the vertical region that the current target belongs to. Within this region, the patch demonstrating the highest distribution of visual features is selected, and its corresponding average pixel values are computed. This pixel location is converted into a target position in 3D space, being the area with the dominant visual feature distribution, and provided to the MPC-based local planner. The image patches, the selected region, and the dominant patch are depicted in Fig. 5.

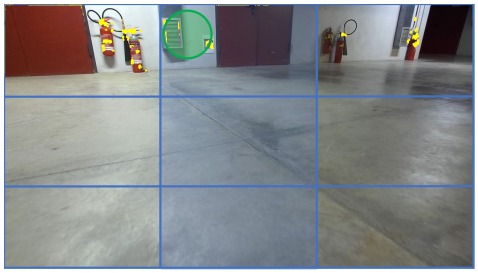


Fig. 5: ORB features are denoted in yellow, while the image's 3×3 grid layout is shown in blue. The highlighted patches represent the region where the local target belongs, and the area with the dominant visual feature distribution is annotated with green.

D. MPC-based Local Planner

The proposed module of the MPC-based local planner utilizes four components as input; the global plan, the robot's position, a depth image for obstacle detection and the 3D position of the consistent and trackable visual features.

MPC, as an optimal controller, calculates sequential control actions over a predefined finite horizon by solving an optimal control problem. At each sampling time, only the first control action is applied to the system's actuators. The process of generating sequential control actions takes place at each sampling time until the terminal condition is met. The feedback control system is then established using the concept of the receding horizon strategy. MPC is able to control nonlinear multi-input, multi-output (MIMO) systems while accounting for states and actuator limitations, so we utilize it to track the global path while maintaining the extracted features in the field of view of the robot when needed. The online optimal control problem can be formulated as follows:

$$J = \min_{\mathbf{u}(k)} \sum_{k=1}^{N} (J_{\text{track}} + J_{\text{avoid}} + J_{\text{feature}}) \tag{11}$$

Subject to:

$$\begin{aligned} &\mathbf{x}(0) = \mathbf{x}_0 \\ &\mathbf{x}(k+1) = f(\mathbf{x}(k), \mathbf{u}(k)) \\ &\mathbf{x}_{\text{min}} \leq \mathbf{x}(k) \leq \mathbf{x}_{\text{max}} \\ &\mathbf{u}_{\text{min}} \leq \mathbf{u}(k) \leq \mathbf{u}_{\text{max}} \end{aligned} \tag{12}$$

where N is the prediction horizon, equal for both prediction and control, $\mathbf{x} = [x, y, \theta]^T$ denotes the state vector, $\mathbf{u} = [v, \omega]^T$ denotes the control vector and f denotes the robot's kinematics for a non-holonomic differential system. J_{track}, J_{avoid}, J_{feature} are terms related to path tracking, obstacle avoidance and visual features tracking respectively. More specifically:

$$J_{\text{track}} = (\|\mathbf{x}(k) - \mathbf{x}^r(k)\|_Q^2) + (\|\mathbf{u}(k) - \mathbf{u}^r(k)\|_R^2) \tag{13}$$

where, for n state variables and m control variables, $Q \in \mathbb{R}^{n \times n}$ and $R \in \mathbb{R}^{m \times m}$ are positive definite symmetric weighting matrices and $\mathbf{x}^r(k)$, $\mathbf{u}^r(k)$ are the reference vectors for states and controls respectively.

$$J_{\text{avoid}} = Q_{avoid} \sum_{i=0}^{M} \frac{1}{\|\mathbf{q}(k) - q_i\|^2} \tag{14}$$

where M in the number of obstacles, $\mathbf{q}(k) = [\mathbf{x}(k), \mathbf{y}(k)]$ is the robot's position at time k, $q_i = (x_i, y_i)$ is the position of the i-th obstacle and Q_{avoid} is the weight for obstacle avoidance.

Finally, for visual features tracking, we introduce another term in the objective function:

$$J_{\text{feature}} = Q_{feature}(\theta_k - \theta_f)^2 \tag{15}$$

where θ_k is the robot's orientation at step k and θ_f is the angle at which, the robot camera view is directly aligned with the position of the most trackable visual feature distribution. $Q_{feature}$ is the weight for feature tracking, which is adapted based on the following equation:

$$Q_{feature} = \begin{cases} a(1 - \bar{D}), & \bar{D} \leq t \\ 0, & \bar{D} > t \end{cases} \tag{16}$$

where a is a weighting factor that ensures balance between path and feature tracking.

As described in Section III-C, lower values of $\bar{D}$ imply increased robustness of the visual features. The static threshold t is selected heuristically, so as to provide a valid indication regarding the possibility of features' tracking failure in localization. If $\bar{D} < t$, MPC feature tracking is

activated, with a weight proportional to the trackability of the extracted features, as described in Eq. 16. Otherwise, MPC neglects features tracking and prioritizes only path tracking and obstacle avoidance.

IV. Experiments

In order to validate the efficiency of the proposed method, we conducted experiments in simulation regarding the global planner module, and real-world experiments for the overall operation of our framework. All of our experiments refer to an indoor parking lot area within the premises of CERTH with repetitive features, variable lighting conditions and featureless areas. The operating environment, which covers an area of around $165\,\mathrm{m}^2$, was scanned by Faro Focus S-150 high resolution terrestrial laser scanner, to generate a 3D model of the scene, as illustrated in Fig. 2. Additionally, we employed the approach in [10], to place 12 fiducial markers within the scene, adjusting the height of the markers in order to be visible by the robot's camera. The placement of markers was adjusted to optimize their positioning within the real scene, without altering the integrity of localizability scores [10]. The fiducial marker placement is depicted in Fig. 9 and the corresponding localizability scores in Fig. 7. For our experiments we set the threshold for the localization scores filtering to 33, the global planner parameters to $k_l = 1.5$, $k_o = 1.5$, $k_g = 10^{-2}$, $k_s = 10^{-4}$, $\rho_o = 2m$, $\rho_l = 2m$, $s = 1m$, the min and max node number of RRT to 20 and 600 respectively. Concerning the local planner parameters, they are set to $Q = diag(200, 200, 200)$, $R = diag(10, 10, 10)$, $Q_{avoid} = 5$, $a = 500$, $t = 0.03$, $u_{max} = 0.15m/s$ and $\omega_{max} = 0.1rad/s$. Regarding the extraction of consistent and trackable visual features module, we set the maximum number of ORB features to 1000 and applied filtering with a ratio of 0.75, to retain only the good matches. The buffer stored $n_{buff} = 10$ frames, while for each matched image, a 3×3 ($n_{seg} = 9$) grid was designed, as depicted in Fig. 5.

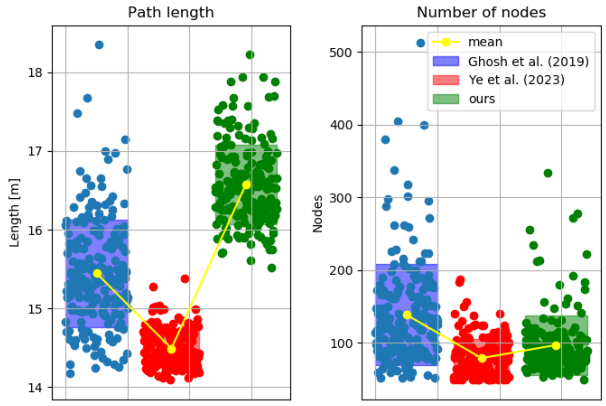


Fig. 6: Performance comparison between [12], [14] and our proposed global path planning module, in terms of path length and the number of nodes.

The localization-aware KB-RRT path planner module is directly compared with the work done in [12] and [14]. The former initially proposed the KB-RRT while the latter enhanced it utilising an APF. Both represent state-of-art optimization approaches for nonholonomic mobile robot path planning. We performed 200 trials of each method, using exactly the same start and goal positions in a common map, while the metrics calculated in each run are the path length, and the number of nodes. The results of this comparison are shown in Fig. 6. Our method appears to have a higher path length than [14] and [12], which is completely justified due to the fact that our method, in order to satisfy the localization requirements and direct the robot both away from featureless regions and towards fiducial markers, needs to generate longer paths. Regarding the number of nodes generated by RRT, our method outperforms [12], with a lower mean and deviation and achieves comparable results with [14]. This is largely due to the fact that our algorithm builds on [14], whose APF-guidance, promotes efficient path finding and minimizes unnecessary node generation. Indicative qualitative results of our method are also depicted in Fig. 7, where the generated plans of our method are compared directly with [14]. The shortest paths generated by Ye et al. present a great divergence from our paths, which tend to avoid the regions with low localizability scores.

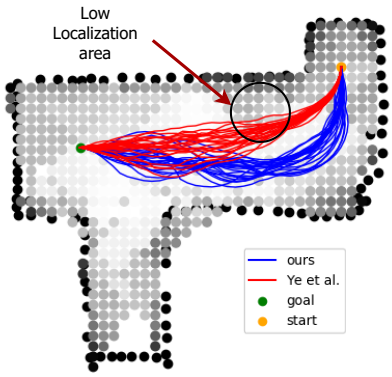
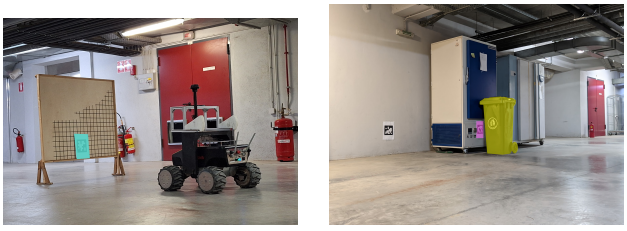


Fig. 7: Comparison of the generated global trajectories between our method and [14]. Our global paths diverge from the shortest route in order to avoid the low localizability areas.



(a) (b)

Fig. 8: (a) The robot approaching its goal. In cyan, we denote the evaluation marker (b) Special case scenario, where the marker annotated in magenta is obscured by an obstacle.

The contribution of our active vision-based path planning module to the localization accuracy of a robot was evaluated through real-world experiments. A holononomic mobile

platform was employed, equipped solely with a front-facing ZED2 stereo camera, as depicted in Fig. 8a. The localization system of the robot consists of an Extended Kalman Filter [30] integrating information from three distinct sources to impartially showcase the performance of our proposed pipeline: i) robot's wheel odometry, ii) visual SLAM position estimate using RTAB-Map [31], and iii) position estimate from fiducial markers placed in the environment, according to the analysis of [10]. In real applications, incorporating a pre-built map would enhance localization performance.

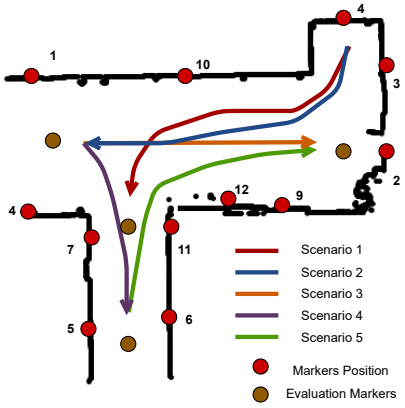


Fig. 9: The 5 different scenarios used for evaluation.

We performed 5 different scenarios ($S_1 - S_5$) within the operating environment and each one is shown in Fig. 9. Additionally, we performed a special scenario $S_5^{'}$, where one fiducial marker, namely marker 12, is obscured, as shown in Fig. 8b, and the start and the goal positions correspond to S_5. The purpose of this scenario is to evaluate our method when the localizability scores lose validity around the aforementioned fiducial marker. The performance of our method is measured in terms of localization accuracy when the robot reaches some pre-determined goal positions. Since ground truth position of the robot is not available, in order to determine this error accurately, we placed a fiducial marker near the goal position, as shown in Fig. 8a, for each scenario, so that it remains visible when the robot reaches close to the target. The evaluation markers' placement is depicted in Fig. 9. We computed the final displacement error (Eq. 17) between the ground truth position of the marker, $\bar{m}_{gt}$, and the estimated position of the marker m_{gt}. For each scenario, we calculated the average error for 10 runs of each method. It must be noted that this fiducial marker does not contribute to the localization system and is used solely for error calculation.

$$FDE = \sqrt{(m_{gt,x} - \bar{m}_{gt,x})^2 + (m_{gt,y} - \bar{m}_{gt,y})^2} \quad (17)$$

An ablation study among the proposed components was performed in order to verify the actual contribution of each module in the complete proposed pipeline. The results of this study are reported in detail in Table I. The combination of the feature tracking module with the baseline path planner of [14] results in 24.12% average improvement of the localization accuracy. Our path planning module, without the use of trackable features, presents an 54.51% average improvement compared to [14]. Finally, the combination of our path planning module with the features tracking is on average 67.74% improved compared to the baseline path planning. The integration of the trackable features extraction module within the local planner yields performance improvements for both path planners. An indicative qualitative depiction of trackable features extraction module's influence can be seen in Fig. 10. Furthermore, our global planner consistently surpasses the baseline counterpart across all scenarios. Notably, the synergetic combination of both modules yields optimal results across all evaluated scenarios. In the specific scenario $S_5^{'}$, both variants of our path planning method, with and without the incorporation of detectable features, demonstrate a decrease in performance, amounting to 42.89% and 17.07% respectively. However, both implementations demonstrate superior performance compared to the baseline path planner, both with and without feature augmentation. Scenario $S_5^{'}$ highlights the efficacy of markers, but also the robustness of our active path planner reducing reliance on them.

TABLE I: Ablation study: Module comparison with respect to average FDE (reported in meters) across different scenarios. * denotes methods using trackable features.

Method	S_1	S_2	S_3	S_4	S_5	$S_5^{'}$
Ye et al.	0.9352	1.7105	0.5846	0.3200	1.9379	1.9379
Ye et al.*	0.8802	1.4118	0.5587	0.2595	0.5063	0.5063
Ours	0.2863	0.9222	0.4456	0.1473	0.4001	0.4684
Ours*	**0.2103**	**0.5336**	**0.3570**	**0.0992**	**0.3016**	**0.4308**

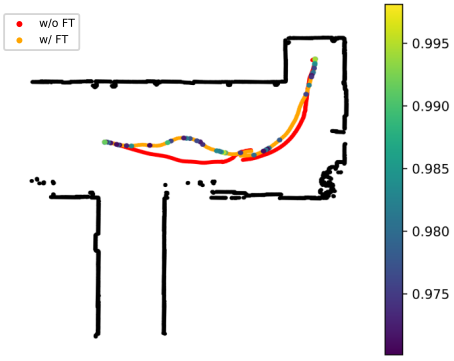


Fig. 10: Generated trajectories of our method with (orange) and without (red) trackable features consideration in scenario 2. The color variations on the orange trajectories denote the activation of trackable features consideration in the MPC-based local planner, while the corresponding colormap expresses the normalized adaptive weight, namely $\bar{D}$.

V. Conclusions

In this work, a holistic framework for active localization-aware path planning in indoor environments is presented. Localizability scores which encode information regarding

areas with rich visual features and fiducial markers' placement of a known environments, are converted into an APF and are utilized by a kinematically constrained, bi-directional RRT-based algorithm. Local planning is based on an online weight-adaptive MPC algorithm that utilizes consistent and trackable visual features to guide the robot in areas with rich information, closely to the target position. The proposed method has been evaluated both in simulation and with real-world experiments, demonstrating its ability to reduce the localization uncertainty during the operation of a robotic mobile platform in an indoor environment.

Our method is constrained by the presence of an empirically selected static threshold t, which determines the possibility of features' tracking failure in localization. Future research will focus on a data-driven optimization technique to automatically infer when localization failure occurs, replacing the above-mentioned threshold. Additionally, integration with temporal persistence modeling techniques towards the online update of the localizability scores will be explored.

ACKNOWLEDGMENT

This work has been supported by the EU Horizon Europe funded project "RobetArme" under the GA with no: 101058731.

REFERENCES

[1] I. Kostavelis, L. Nalpantidis, R. Detry, H. Bruyninckx, A. Billard, S. Christian, M. Bosch, K. Andronikidis, H. Lund-Nielsen, P. Yosefipor *et al.*, "Robétarmé project: Human-robot collaborative construction system for shotcrete digitization and automation through advanced perception, cognition, mobility and additive manufacturing skills," *Open Research Europe*, vol. 4, p. 4, 2024.

[2] M. Santiago, M. Marcos, M. Elisabeth, P. Vartholomeos, D. Giakoumis, A. Simi, and C. Balaguer, "Badger: Intelligent robotic system for underground construction," *Infrastructure Robotics: Methodologies, Robotic Systems and Applications*, pp. 205–220, 2024.

[3] X. Deng, Z. Zhang, A. Sintov, J. Huang, and T. Bretl, "Feature-constrained active visual slam for mobile robot navigation," in *2018 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2018, pp. 7233–7238.

[4] R. Munoz-Salinas and R. Medina-Carnicer, "Ucoslam: Simultaneous localization and mapping by fusion of keypoints and squared planar markers," *Pattern Recognition*, vol. 101, p. 107193, 2020.

[5] J. A. Placed, J. Strader, H. Carrillo, N. Atanasov, V. Indelman, L. Carlone, and J. A. Castellanos, "A survey on active simultaneous localization and mapping: State of the art and new frontiers," *IEEE Transactions on Robotics*, 2023.

[6] S. Macenski, F. Martín, R. White, and J. G. Clavero, "The marathon 2: A navigation system," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 2718–2725.

[7] J. Xing, G. Cioffi, J. Hidalgo-Carrió, and D. Scaramuzza, "Autonomous power line inspection with drones via perception-aware mpc," *arXiv preprint arXiv:2304.00959*, 2023.

[8] K. Tsiakas, D. Alexiou, D. Giakoumis, A. Gasteratos, and D. Tzovaras, "Leveraging multimodal sensing and topometric mapping for human-like autonomous navigation in complex environments," in *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2023, pp. 7415–7421.

[9] D. Katsatos, D. Alexiou, T. Kontodina, I. Chatzikonstantinou, I. Kostavelis, D. Giakoumis, and D. Tzovaras, "Comparative study of surface 3d reconstruction methods applied in construction sites," in *2023 IEEE International Conference on Imaging Systems and Techniques (IST)*. IEEE, 2023, pp. 1–6.

[10] Q. Huang, J. DeGol, V. Fragoso, S. N. Sinha, and J. J. Leonard, "Optimizing fiducial marker placement for improved visual localization," *IEEE Robotics and Automation Letters*, vol. 8, no. 5, pp. 2756–2763, 2023.

[11] S. M. LaValle, J. J. Kuffner, B. Donald *et al.*, "Rapidly-exploring random trees: Progress and prospects," *Algorithmic and computational robotics: new directions*, vol. 5, pp. 293–308, 2001.

[12] D. Ghosh, G. Nandakumar, K. Narayanan, V. Honkote, and S. Sharma, "Kinematic constraints based bi-directional rrt (kb-rrt) with parameterized trajectories for robot path planning in cluttered environment," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 8627–8633.

[13] J. Wang, B. Li, and M. Q.-H. Meng, "Kinematic constrained bi-directional rrt with efficient branch pruning for robot path planning," *Expert Systems with Applications*, vol. 170, p. 114541, 2021.

[14] L. Ye, F. Wu, X. Zou, and J. Li, "Path planning for mobile robots in unstructured orchard environments: An improved kinematically constrained bi-directional rrt approach," *Computers and Electronics in Agriculture*, vol. 215, p. 108453, 2023.

[15] A. Kim and R. M. Eustice, "Active visual slam for robotic area coverage: Theory and experiment," *The International Journal of Robotics Research*, vol. 34, no. 4-5, pp. 457–475, 2015.

[16] R. T. Rodrigues, M. Basiri, A. P. Aguiar, and P. Miraldo, "Low-level active visual navigation: Increasing robustness of vision-based localization using potential fields," *IEEE Robotics and Automation Letters*, vol. 3, no. 3, pp. 2079–2086, 2018.

[17] D. Kim, G. C. Kim, Y. Jang, and H. J. Kim, "Topology-guided path planning for reliable visual navigation of mavs," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2021, pp. 3117–3124.

[18] G. Sun, X. Zhang, Y. Liu, H. Wang, X. Zhang, and Y. Zhuang, "Topology-guided perception-aware receding horizon trajectory generation for uavs," in *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2023, pp. 3070–3076.

[19] W. Xue, R. Ying, F. Wen, Y. Chen, and P. Liu, "Active slam with prior topo-metric graph starting at uncertain position," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 1134–1141, 2021.

[20] Y. Chen, S. Huang, and R. Fitch, "Active slam for mobile robots with area coverage and obstacle avoidance," *IEEE/ASME Transactions on Mechatronics*, vol. 25, no. 3, pp. 1182–1192, 2020.

[21] M. Xu, L. Zhao, S. Huang, and Q. Hao, "Active slam in 3d deformable environments," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022, pp. 7952–7958.

[22] J. S. Willners, S. Katagiri, S. Xu, T. Łuczyński, J. Roe, and Y. Petillot, "Adaptive heading for perception-aware trajectory following," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2023, pp. 3161–3167.

[23] S. Zhang, R. Cui, W. Yan, and Y. Li, "Dual-layer path planning with pose slam for autonomous exploration in gps-denied environments," *IEEE Transactions on Industrial Electronics*, 2023.

[24] L. Bartolomei, L. Teixeira, and M. Chli, "Semantic-aware active perception for uavs using deep reinforcement learning," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2021, pp. 3101–3108.

[25] F. Lin, Z. Ji, C. Wei, and R. Grech, "Localisation-safe reinforcement learning for mapless navigation," in *2022 IEEE International Conference on Robotics and Biomimetics (ROBIO)*. IEEE, 2022, pp. 1327–1334.

[26] S. K. Gottipati, K. Seo, D. Bhatt, V. Mai, K. Murthy, and L. Paull, "Deep active localization," *IEEE Robotics and Automation Letters*, vol. 4, no. 4, pp. 4394–4401, 2019.

[27] Z. Zhang and D. Scaramuzza, "Fisher information field: an efficient and differentiable map for perception-aware planning," *arXiv preprint arXiv:2008.03324*, 2020.

[28] ——, "Beyond point clouds: Fisher information field for active visual localization," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 5986–5992.

[29] J. Lim, N. Lawrance, F. Achermann, T. Stastny, R. Bähnemann, and R. Siegwart, "Fisher information based active planning for aerial photogrammetry," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2023, pp. 1249–1255.

[30] T. Moore and D. Stouch, "A generalized extended kalman filter implementation for the robot operating system," in *Proceedings of the 13th International Conference on Intelligent Autonomous Systems (IAS-13)*. Springer, July 2014.

[31] M. Labbé and F. Michaud, "Rtab-map as an open-source lidar and visual simultaneous localization and mapping library for large-scale and long-term online operation," *Journal of field robotics*, vol. 36, no. 2, pp. 416–446, 2019.