

Interactive Distance Field Mapping and Planning to Enable Human-Robot Collaboration

Usama Ali^{1*}, Lan Wu^{2*}, Adrian Müller¹, Fouad Sukkar², Tobias Kaupp¹, Teresa Vidal-Calleja²

Abstract—Human-robot collaborative applications require scene representations that are kept up-to-date and facilitate safe motions in dynamic scenes. In this letter, we present an interactive distance field mapping and planning (IDMP) framework that handles dynamic objects and collision avoidance through an efficient representation. We define *interactive* mapping and planning as the process of creating and updating the representation of the scene online while simultaneously planning and adapting the robot’s actions based on that representation. The key aspect of this work is an efficient Gaussian Process field that performs incremental updates and handles dynamic objects reliably by identifying moving points via a simple and elegant formulation based on queries from a temporary latent model. In terms of mapping, IDMP is able to fuse point cloud data from single and multiple sensors, query the free space at any spatial resolution, and deal with moving objects without semantics. In terms of planning, IDMP allows seamless integration with gradient-based reactive planners facilitating dynamic obstacle avoidance for safe human-robot interactions. Our mapping performance is evaluated on both real and synthetic datasets. A comparison with similar state-of-the-art frameworks shows superior performance when handling dynamic objects and comparable or better performance in the accuracy of the computed distance and gradient field. Finally, we show how the framework can be used for fast motion planning in the presence of moving objects both in simulated and real-world scenes. An accompanying video, code, and datasets are made publicly available³.

Index Terms—Interactive Mapping and Planning, Euclidean Distance Fields, Gaussian Process, Mapping, Motion Planning, Human-Robot Collaboration.

I. INTRODUCTION

Human-robot collaboration (HRC) and other applications of robots in the field call for interactive representations to deal with dynamic and evolving scenes. For true collaboration in industrial settings, humans and robots physically share the same space, e.g. working jointly and simultaneously on the assembly of a product. The environment consisting of the human, robot, workspace, and objects such as tools and assembly components needs to be monitored by sensors and

This work has been submitted to the IEEE for possible publication. Copyright may be transferred without notice, after which this version may no longer be accessible.

This work was supported by the Industrial Transformation Training Centre (ITTC) for Collaborative Robotics in Advanced Manufacturing (also known as the Australian Cobotics Centre) funded by ARC (Project ID: IC200100001) and the Bavarian Research Foundation (grant AZ-1512-21).

*These authors are co-first authors and contributed equally to this work.
Corresponding author: Lan.Wu-2@uts.edu.au

¹Authors are with the Center for Robotics (CERI) at the Technical University of Applied Sciences Würzburg-Schweinfurt (THWS), Germany.

²Authors are with the Robotics Institute, Faculty of Engineering and IT, University of Technology Sydney (UTS), Australia.

³<https://uts-ri.github.io/IDMP>

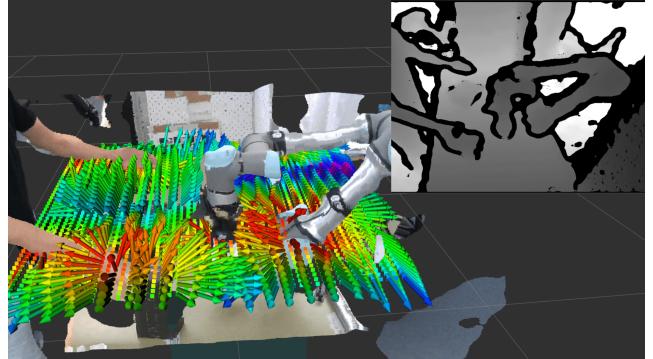


Fig. 1: Interactive generation of a distance and gradient field in an HRC setting. Top right: Depth image from an Intel Realsense camera. Bottom left: Coloured point cloud and a horizontal slice of the field (red/blue means close/far to the nearest object with arrows pointing away from it).

kept up-to-date in a timely manner. One approach is to create and continuously update an interactive representation of the changing scene, which is useful for safe and effective robot motion planning around human operators.

Recent representations that aim to fulfil some of the above-mentioned requirements have been proposed in the robotics literature [1], [2]. Euclidean Distance Field (EDF), for instance, has recently become a promising representation suitable for direct collision-checking. As an example, the methods in [3], [4], [2] compute the projective or non-projective Truncated Signed Distance Field (TSDF) first, then propagate from TSDF to the free space for updating the EDF. Han et al. [1] propose to propagate from an occupancy map instead of the TSDF to maintain efficiency while preserving accuracy for the EDF computation. EDF representations also have the ability to dynamically update the changes via the so-called free space carving method. This method performs expensive ray-casting and progressive integration of the TSDF to update the map, resulting in free space that is only gradually cleared when objects move in the scene.

This paper presents an interactive distance field mapping and planning (IDMP) framework aimed at dynamic scenes common in human-robot collaboration scenarios (see Fig. 1 as an example). By extending our previous work in [5], [6], we focus to build and maintain an efficient and up-to-date distance and gradient field using a Gaussian-Process-based method capable of integrating depth sensor measurements. Here, we present a simple and elegant method catering both dynamic updates and seamless fusion by querying distances and gradients from a temporary latent GP distance field

generated with only the points of the current frame. We name this latent representation the ‘‘Frustum Field’’. The Frustum Field is able to deal with the discrepancy in the distance between surface points that move in the current frame without the need for ray tracing. It also allows fusion with previously mapped distance and gradient fields.

To overcome the computational complexity of GP-based approaches, our framework maintains the updates and fusion in an Octree structure that runs online. Furthermore, our method handles dynamic objects reliably via the information provided by the Frustum Field. The fused representation is continuous, allowing querying of the Euclidean distance to the nearest surface and its gradient at an arbitrary spatial resolution. To demonstrate the potential of our efficient and differentiable fused distance and gradient fields for human-robot interaction tasks, we integrate IDMP with a gradient-based motion planner as well as a local reactive behaviour. We experimentally evaluate IDMP’s performance in both static and dynamic scenes and benchmark it against state-of-the-art algorithms and demonstrate its capabilities in real-world experiments for a robot arm interacting with a human.

To summarise, the key contributions of the paper are:

- A novel dynamic update and fusion method for GP-based distance and gradient fields. This method is based on a latent representation named Frustum Field that enables dealing with static, moving, and new points.
- An interactive distance field mapping and planning (IDMP) framework that generates a continuous Euclidean distance and gradient field online, and is integrated with gradient-based reactive motion and path planner for 3D collision-free interactions.
- An efficient open-sourced ROS-based implementation of the full IDMP framework.¹

To the best of our knowledge, this work is the first framework that addresses the close interactivity between perception and motion planning aimed at HRC applications.

II. RELATED WORK

EDFs are useful representations for a wide range of tasks which require distance-to-obstacle information at different points in space. For example, motion planners such as gradient-based trajectory optimisers [7], [8] utilise EDFs and their gradients to perturb trajectories smoothly in order to avoid obstacles. EDFs can be computed by solving the Eikonal equation. A major challenge in exploiting the Eikonal equation to recover the ESDF is that it is non-linear and hyperbolic. Therefore, it is difficult to solve directly, and specialised numerical approaches are often required, including discrete methods such as fast marching or label-correcting to propagate the distance field through a grid.

Voxblox [3] proposes an approach to compute the ESDF based on the projective TSDF through a wavefront propagation algorithm. Voxfield [4] and VDBblob [2] adopt the use of a non-projective TSDF to improve the accuracy. FIESTA [1] proposes to use an occupancy map instead of TSDF and a

novel data structure to maintain efficiency. VED-EDT uses a distance transform function to represent the EDF hierarchically [9]. These EDF representations both have the capability to dynamically update the changes via the so-called free space carving method, which updates every voxel along the grouped rays. This method requires performing expensive ray-casting and progressive integration and weighting of the TSDF to update the map, resulting in free space that is only gradually cleared when objects move in the scene. There are also frameworks that explicitly deal with dynamic objects [10]. However, the dynamic object handling is decoupled from mapping. Our framework deals reliably with dynamic objects and their surfaces are considered for collision avoidance, thus tightly coupled with the estimated representation.

Over the last few years, there have been numerous works exploring the potential of learning approaches such as neural networks for implicit signed distance representations. In particular, inspired by [11], the authors in [12] propose to use an implicit neural representation to constrain the learning to approximate the SDF. Recently proposed in [13], the authors learn an approximation of EDF from Neural Radiance Fields (NeRFs) with occupancy inference. Similarly, inspired from [12], iSDF [14] is proposed to use a neural signed distance field for mapping. For dynamic scenes, D-NeRF considers time as an additional input to the system and splits the learning process into two main stages: one that encodes the scene into a canonical space and another that maps this canonical representation into the deformed scene at a particular time. While these approaches have benefits such as a continuous representation and lower memory cost, they are expensive to compute and require powerful GPUs.

GP-based distance fields are appealing as they can represent complex environments with non-parametric models [15], [5]. In our previous work, we produce the continuous distance field by formulating the Gaussian Process via a latent distance function. The regularised Eikonal equation can be simply solved by applying the logarithmic transformation to a GP formulation to approximate the accurate Euclidean distance fields [5], [16]. In [16] a trajectory optimisation approach for motion planning is implemented for the 2D case. In this work, we generalise and apply this approach to 3D.

Following the theoretical formulation in [5], [16], we have introduced the so-called reverting GP distance field [6] that increased the accuracy of the distance field addressing multiple key problems in robotics with noisy measurements. A GP distance field has been applied to the active dynamic mapping framework in [17] to enable accurate distance inferences [18]. However, [17], [18] requires the inverse depth to formulate the occupancy field for dynamic updates. In contrast, we present a simple and elegant method for dynamic updates and fusion of our GP distance field by querying distance and gradients from a latent GP distance field generated with only the points of the current frame. Our proposed framework and its implementation is faster than all our previous works, making it suitable for close-proximity HRC tasks.

In the context of HRC, it is important to have an efficient representation that informs the planner of a possible collision, in particular, when objects are moving in the scene. Two

¹Code is made publicly available at <https://github.com/UTS-RI>IDMP>.

recent review papers discuss the state of the art in safe HRC [19], [20]. Both distinguish between methods that are aimed at contact detection/mitigation, and collision avoidance. Our approach follows the latter which is enabled by the interactivity between mapping and motion planning.

Many obstacle avoidance methods for HRC explicitly track the human operator using RGB-D cameras [21], [22], [23]. Human limbs are represented using skeletons [22], [23], bounding spheres [23], cylinders [24] or an occupancy box [25]. Based on the model of the human and the known state of the robot, the control strategy is often based on discrete modes, e.g. "autonomous behaviour" [21]. In some HRC control frameworks, the human position is assumed to be known, e.g. [26]. In contrast to those approaches, our framework does not rely on semantics enabling workspace monitoring of all types of objects (static and dynamic) using a continuous field representation. Our efficient map update allows an online use of standard trajectory optimisation algorithms rather than discrete control modes. Another non-semantic approach is given in [27], [28]. However, they only compute distances to discrete control points while our approach is able to recover a continuous field including non-observed areas.

III. BACKGROUND

A. Gaussian Process Distance Field

Gaussian Process (GP) [29] is a flexible and probabilistic approach to modelling distributions over functions, particularly suitable in scenarios with noisy and sparse measurements. Let us consider a set of noisy input observations (training points) $\mathbf{y} = \{y_j = f(\mathbf{x}_j) + \epsilon_j\}_{j=1}^J \in \mathbb{R}$ taken at locations $\mathbf{X} = \{\mathbf{x}_j\}_{j=1}^J \in \mathbb{R}^D$ and corrupted by additive Gaussian noise $\epsilon_j \sim \mathcal{N}(0, \sigma^2)$. We model the unknown distribution f as a GP:

$$f \sim \mathcal{GP}(0, k_{\mathbf{xx}}), \quad (1)$$

where $k_{\mathbf{xx}}$ is the covariance function that controls the properties of f . Suppose we have a set of testing points $\mathbf{X}_* = \{\mathbf{x}_{*q}\}_{q=1}^Q \in \mathbb{R}^D$. At the testing locations, we can express the joint distribution of the function values and the observed target values using the kernel function,

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{f}_* \end{bmatrix} = \mathcal{N} \left(\mathbf{0}, \begin{bmatrix} \mathbf{K}_{\mathbf{XX}} + \sigma_f^2 \mathbf{I} & \mathbf{K}_{\mathbf{XX}*} \\ \mathbf{K}_{\mathbf{X}*\mathbf{X}} & \mathbf{K}_{\mathbf{X}*\mathbf{X}*} \end{bmatrix} \right). \quad (2)$$

\mathbf{f}_* is the vector of predicted function values at \mathbf{X}_* , \mathbf{I} is the identity matrix, σ_f^2 is the variance of observation noise and $\mathbf{K}_{\mathbf{XX}}$ is the $J \times J$ covariance matrix between input points. By conditioning Eq. (2), the posterior distribution of f at an arbitrary testing point \mathbf{x}_* is given by $f(\mathbf{x}_*) \sim \mathcal{N}(\hat{f}(\mathbf{x}_*), \nabla \hat{f}(\mathbf{x}_*))$, where predictive mean and variance are:

$$\hat{f}(\mathbf{x}_*) = \mathbf{k}_{\mathbf{x}_*\mathbf{x}} (\mathbf{K}_{\mathbf{XX}} + \sigma_f^2 \mathbf{I})^{-1} \mathbf{y}, \quad (3)$$

$$\nabla \hat{f}(\mathbf{x}_*) = \mathbf{k}_{\mathbf{x}_*\mathbf{x}_*} - \mathbf{k}_{\mathbf{x}_*\mathbf{x}} (\mathbf{K}_{\mathbf{XX}} + \sigma_f^2 \mathbf{I})^{-1} \mathbf{k}_{\mathbf{x}_*\mathbf{x}}^\top. \quad (4)$$

$\mathbf{k}_{\mathbf{x}_*\mathbf{x}}$ is the vector of covariances between the J input points and the testing point. $\mathbf{k}_{\mathbf{x}_*\mathbf{x}_*}$ is the covariance function value of the testing point.

Let us now briefly introduce the so-called reverting GP distance field originally presented in [6]. Consider a surface

\mathcal{S} in a Euclidean space \mathbb{R}^D , and $\mathbf{X} = \{\mathbf{x}_j\}_{j=1}^J \in \mathbb{R}^D$ a set of discrete observations of \mathcal{S} as training points. By modelling the occupancy $o(\mathbf{x}) : \mathbb{R}^D \mapsto \mathbb{R}$ of the space with a GP as $o \sim \mathcal{GP}(0, k_{\mathbf{xx}'})$, it is possible to infer the surface occupancy $\hat{o}(\mathbf{x}_*)$ at any location in the space using Eq. (3). In other words, the covariance kernel $k_{\mathbf{xx}'}$ is a monotonic function of the distance $\|\mathbf{x} - \mathbf{x}'\|$ between \mathbf{x} and \mathbf{x}' .

We arbitrarily define the occupied area to be equal to 1. Therefore, y is equal to 1 in Eq. (3) and

$$\hat{o}(\mathbf{x}_*) = \mathbf{k}_{\mathbf{x}_*\mathbf{x}} (\mathbf{K}_{\mathbf{XX}} + \sigma_o^2 \mathbf{I})^{-1} \mathbf{1}. \quad (5)$$

The distance field $\hat{d}(\mathbf{x}_*)$ given any location \mathbf{x}_* is obtained by applying a *reverting* function r to the occupancy field as

$$\hat{d}(\mathbf{x}_*) = r(\hat{o}(\mathbf{x}_*)). \quad (6)$$

The reverting function depends on the chosen kernel and corresponds to the inverse operation, computing the distance as a function of the covariance between \mathbf{x} and \mathbf{x}' :

$$\begin{aligned} r\left(\tilde{k}(\|\mathbf{x} - \mathbf{x}'\|)\right) &= \|\mathbf{x} - \mathbf{x}'\| \\ \text{with } \tilde{k}(\|\mathbf{x} - \mathbf{x}'\|) &= k_{\mathbf{xx}'}. \end{aligned} \quad (7)$$

Considering the square exponential kernel $k_{\mathbf{xx}'} = \sigma^2 \exp\left(-\frac{\|\mathbf{x}-\mathbf{x}'\|^2}{2l^2}\right)$, substituting the reverting function of it into Eq. 6 gives us:

$$\hat{d}(\mathbf{x}_*) = \sqrt{-2l^2 \log\left(\frac{\hat{o}(\mathbf{x}_*)}{\sigma^2}\right)}. \quad (8)$$

This method assumes that the GP-based occupancy field in \mathbb{R}^D behaves similarly to the one-noiseless-point scenario where the reverting function yields the exact distance to the data point.

B. Nabla Operator for Gradient Field

The derivative of a GP is a linear operation that produces another GP. One common solution to infer the gradient field is to have surface normals as input for GP modelling. However, this requires inverting a computational-heavy joint covariance function of $k_{\mathbf{xx}'}$ with the partial derivatives of $k_{\mathbf{xx}'}$ at \mathbf{x} and \mathbf{x}' [30]. Instead, we propose to use the Nabla operator to infer the gradient along with the accurate distance field without the normal as input [31]. Applying this linear operator to Eq. 5, we get

$$\nabla \hat{o}(\mathbf{x}_*) = \nabla \mathbf{k}_{\mathbf{x}_*\mathbf{x}} (\mathbf{K}_{\mathbf{XX}} + \sigma_o^2 \mathbf{I})^{-1} \mathbf{1}, \quad (9)$$

where $\nabla \mathbf{k}_{\mathbf{x}_*\mathbf{x}}$ is the partial derivative with respect to \mathbf{x} [32]. Taking the gradient of both sides of Eq. 8 with respect to the distance shows that the gradient of $\hat{d}(\mathbf{x}_*)$ points in the same direction as the gradient of $\hat{o}(\mathbf{x}_*)$, subject to a scaling factor, so that $\nabla \hat{d}(\mathbf{x}_*) \approx \nabla \hat{o}(\mathbf{x}_*)$.

IV. PROPOSED FRAMEWORK

An overview of the proposed IDMP framework is shown in Fig. 2. Given data from one or multiple depth sensors as input, IDMP incrementally builds the *Fused Field*, a continuous GP distance and gradient field for the motion planner to query and adapt its paths in response to a changing map.

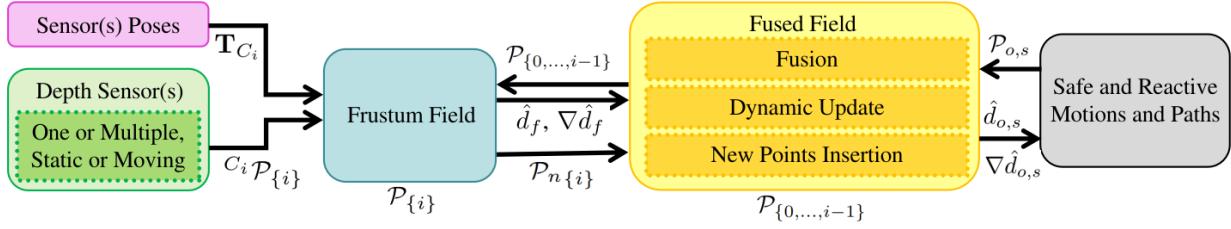


Fig. 2: System diagram of IDMP. We first model the temporary latent Frustum Field (blue) using only $\mathcal{P}_{\{i\}}$ as training points. All prior training points $\mathcal{P}_{\{0,\dots,i-1\}}$ in the Fused Field (yellow) are then passed to the Frustum Field. Given the sensor pose, the Frustum Field selects from $\mathcal{P}_{\{0,\dots,i-1\}}$ the points that are within the frustum area $\mathcal{P}_{f\{0,\dots,i-1\}}$ and returns the inferred values of $\hat{d}_f, \nabla \hat{d}_f$ to the Fused Field. These distances and gradients are used to perform fusion and dynamic updates by updating the training points that model the Fused Field. The path planner then queries for $\hat{d}_{o,s}, \nabla \hat{d}_{o,s}$ in order to compute and adapt its motion plans in response to a changing map.

The measurements in the sensor frame C_i at current time i are denoted as point cloud $C_i \mathcal{P}_{\{i\}}$. This raw point cloud is projected from current sensor frame C_i to world frame W given a transformation matrix T_{C_i} yielding a point cloud $\mathcal{P}_{\{i\}}$ in the world reference frame. In the first process (blue block), we model a temporary latent GP distance and gradient field only using $\mathcal{P}_{\{i\}}$ as training points. This field only covers the region in the sensor's field of view, thus we call it *Frustum Field*. This latent model allows us to query the distance and gradients with respect to the current measurements. Note that the Frustum Field is rebuilt at every frame i .

The next process of IDMP (yellow block) builds and maintains our main representation, the Fused Field, generated with all prior training points $\mathcal{P}_{\{0,\dots,i-1\}}$. We pass $\mathcal{P}_{\{0,\dots,i-1\}}$ to the Frustum Field to select points only within the frustum area, denoted as $\mathcal{P}_{f\{0,\dots,i-1\}}$. Then we use $\mathcal{P}_{f\{0,\dots,i-1\}}$ to query the Frustum Field for the distance \hat{d}_f and gradient $\nabla \hat{d}_f$ as per Eq. (8). Based on the distance and gradient inference, the fusion method adjusts the training points' positions, and the dynamic update method eliminates past training points of the objects moving in the scene. We call the intersection between $\mathcal{P}_{\{0,\dots,i-1\}}$ and $\mathcal{P}_{\{i\}}$ the overlapping area. Note that new measurements that are out of the range of the overlapping area, denoted as $\mathcal{P}_{n\{i\}}$, are directly merged into the Fused Field. By updating the training points, which are maintained via an Octree-based data structure, we update the Fused Field.

For planning, let us define workspace points along the robot's links as $\mathcal{P}_{o,s}$. The motion planning algorithm then queries the Fused Field for $\hat{d}_{o,s}, \nabla \hat{d}_{o,s}$ at $\mathcal{P}_{o,s}$. These distances and gradients are used to compute a safe trajectory and can be repeatedly queried to adapt to a changing map.

A. Frustum Field

Every raw point transformed to the world frame $\mathcal{P}_{\{i\}}$ is used to generate the Frustum Field. Note that the Frustum Field is temporary and therefore trained and regenerated at every frame i using the current measurements. As explained in Sec. III, we use $\mathcal{P}_{\{i\}}$ as a set of discrete observations $\mathbf{X} = \{\mathbf{x}_j\}_{j=1}^J$ of surface S . We then model the occupancy via Eq. 5 and apply the reverting function to obtain the distance field. The Frustum Field is then inferred via Eq. 8. By using the linear operator for gradient inference, we employ Eq. 9 to infer the gradients along with the distance information. After the Frustum Field

is modelled, we can query the distance and gradient from the points $\mathcal{P}_{f\{0,\dots,i-1\}}$ to pass them into the next process, the Fused Field, for the interactive updates.

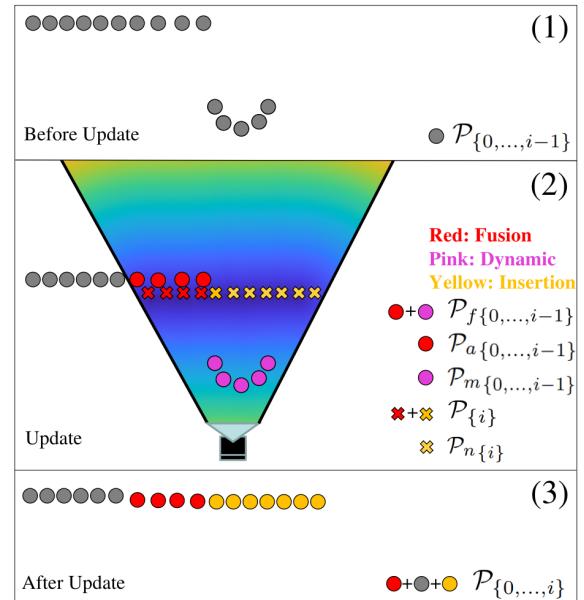


Fig. 3: A simplified illustration of the proposed method. (1) Before the update, we have the prior training points $\mathcal{P}_{\{0,\dots,i-1\}}$ in the Fused Field. (2) The coloured area is the Frustum Field (blue is close to the surface and yellow is far from it). The current points $\mathcal{P}_{\{i\}}$ are shown as red and yellow crosses. The prior training points within the frustum $\mathcal{P}_{f\{0,\dots,i-1\}}$ are marked by red and pink. Fusion, dynamic update, and insertion processes are shown in red, pink, and yellow respectively. (3) After the update, we have the updated training points $\mathcal{P}_{\{0,\dots,i\}}$ that model the Fused Field used by the planner.

B. Fused Field

The training points are used to generate the Fused Field incrementally. Before the update at frame i , the Fused Field is modelled based on the prior training points $\mathcal{P}_{\{0,\dots,i-1\}}$ from frame 0 to $i-1$. After fusion and dynamic update to allow merging $\mathcal{P}_{\{i\}}$ with $\mathcal{P}_{\{0,\dots,i-1\}}$, the Fused Field is modelled using the updated training points $\mathcal{P}_{\{0,\dots,i\}}$. There are three key processes performed at every frame: adjusting points for

fusion, removing points for dynamic update, and inserting points for measurements in new areas. Fig. 3 illustrates the way the points are selected given the Frustum Field. The top figure shows the points $\mathcal{P}_{\{0, \dots, i-1\}}$ before the update given frame i . The middle figure shows the Frustum Field and the points involved in the three processes as described in the previous section. Red and yellow crosses are the new measurements $\mathcal{P}_{\{i\}}$ captured by the sensor. The coloured background is the Frustum Field modelled using $\mathcal{P}_{\{i\}}$. The blue area means the distance is close to the surface, and the yellow area is far away from the surface. Circles coloured in red and pink are the prior training points of the Fused Field only in the frustum, denoted as $\mathcal{P}_{f\{0, \dots, i-1\}}$. Note that we use the sensor pose to select $\mathcal{P}_{f\{0, \dots, i-1\}} \subset \mathcal{P}_{\{0, \dots, i-1\}}$. We then query the Frustum Field to infer \hat{d}_f and gradient $\nabla \hat{d}_f$ to perform the fusion and dynamic update. An advantage of this approach is that we only query the Frustum Field once to perform the three further processes described as follows.

1) **Fusion:** Let us denote the prior training points within the frustum as $\mathcal{P}_{a\{0, \dots, i-1\}} \subset \mathcal{P}_{f\{0, \dots, i-1\}}$ (red circles in Fig. 3) and corresponding distances and gradients as $\hat{d}_a, \nabla \hat{d}_a$. A threshold on the distance value η is then used to indicate if $\mathcal{P}_{a\{0, \dots, i-1\}}$ are relatively close to the surface given the current Frustum Field. For each point $\mathbf{p}_{a\{0, \dots, i-1\}}$ in $\mathcal{P}_{a\{0, \dots, i-1\}}$, the proposed fusion is performed by updating its position through the distance to the surface \hat{d}_a in the direction of the gradient $\nabla \hat{d}_a$ as:

$$\hat{\mathbf{p}}_a = \mathbf{p}_a - \hat{d}_a \nabla \hat{d}_a. \quad (10)$$

Note that \hat{d}_a and $\nabla \hat{d}_a$ are the result of simply querying the Frustum Field. No data association nor iteration is required as GP distance inference produces directly the distance to the surface and it is accurate through our reverting GP model. After the fusion, $\mathcal{P}_{a\{0, \dots, i-1\}}$ is replaced by $\hat{\mathcal{P}}_{a\{0, \dots, i-1\}}$ as the training point of the Fused Field. Note that since we have $\hat{\mathcal{P}}_{a\{0, \dots, i-1\}}$, we do not need to include new points in the overlapping area (red crosses) into the Fused Field.

2) **Dynamic Update:** We show the dynamic update in Fig. 3. Dynamic objects are handled based on the information provided by the Frustum Field as well, as we only update the moving objects in the current field of view. Let us denote the prior training points to-be-removed as $\mathcal{P}_{m\{0, \dots, i-1\}} \subset \mathcal{P}_{f\{0, \dots, i-1\}}$ (pink circles). Note that in this case and without loss of generality, we can capture measurements behind $\mathcal{P}_{f\{0, \dots, i-1\}}$ represents that the points $\mathcal{P}_{f\{0, \dots, i-1\}}$ have been removed and are not in the scene anymore. The Frustum Field distance values \hat{d}_m for these points are bigger than η to indicate they are relatively far away from the current captured measurements, which means the object has moved. We then eliminate from the training set the removed points $\mathcal{P}_{m\{0, \dots, i-1\}}$ of the Fused Field. Note that in some cases, the points to adjust and remove are very close to each other and may not be separated explicitly.

3) **New Points Insertion:** Let us denote the points in $\mathcal{P}_{\{i\}}$ that are outside of the overlapping area as $\mathcal{P}_{n\{i\}}$ (yellow crosses). These points are directly added to the Fused Field training set. Points outside the overlapping area in the Fused Field remain the same (grey circles). Following our illustra-

tion, at the bottom figure in Fig. 3 after the update, we use the updated training points $\mathcal{P}_{\{0, \dots, i\}}$ to model the Fused Field and have it ready for the motion planner query.

We only use one Frustum Field for all interactive updates. The Frustum Field allows to perform the fusion in just one-step and to clear the removed object effectively.

C. Reactive Planning

Due to the differentiability and efficiency of the Fused Field, our framework naturally accommodates reactive planning. For example, a gradient-based motion planner can query \hat{d}_o and $\nabla \hat{d}_o$ and utilise these to adapt its motion plans in response to a changing map. A key advantage of our framework is that we can use the gradient inference in Eq. 9 to compute the gradients analytically, whereas conventional approaches rely on discrete grid-based approximation [7]. This results in faster and more accurate computation of gradients which is crucial for close-proximity human-robot collaborative tasks where fast replanning is required.

In addition, our framework naturally accommodates fast local reactive methods. Consider the current point \mathbf{x}_s and a goal point \mathbf{x}_g . Given the queried distance and gradient of the current point as \hat{d}_s and $\nabla \hat{d}_s$ from our IDMP Fused Field, we can formulate a repulsive vector $\mathbf{v}_{rep} = \nabla \hat{d}_s$ and an attractive vector $\mathbf{v}_{att} = \mathbf{x}_g - \mathbf{x}_s$. Note that both vectors need to be normalised. The repulsive behaviour is defined by a function $w(\hat{d}_s)$ dependent on the distance. We then can formulate the resulting vector by combining the repulsive vector and attractive vector using $w(\hat{d}_s)$ as in [33],

$$\mathbf{v}_{res} = w(\hat{d}_s) \mathbf{v}_{rep} + (1 - w(\hat{d}_s)) \mathbf{v}_{att}. \quad (11)$$

By efficiently computing Eq. 11 at each timestep, a robot has the ability to follow the resulting vector to reactively avoid moving objects present in the scene in an online manner.

V. EVALUATION

To evaluate the proposed IDMP framework, we a) quantitatively compare the mapping performance for both static and dynamic scenes with other frameworks, and b) qualitatively demonstrate online motion planning for a dynamic scene.

A. Distance Field Mapping

We compare IDMP to the two state-of-the-art algorithms: FIESTA [1] and Voxblox [3]. Both are designed to compute the distance field observed by a 3D sensor online. One key difference from both approaches to IDMP is the need to specify a fixed voxel size upfront which directly influences performance. In contrast, IDMP uses a continuous function to represent the field that can be queried at any point in space. We also demonstrate IDMP’s ability to handle dynamic scenes with unknown moving objects and compare the performance to FIESTA/Voxblox.

The evaluation is conducted on two distinct datasets: a) the widely used Cow & Lady dataset [3], and b) a custom synthetic dataset featuring a dynamically moving ball on a table. These datasets are chosen to assess IDMP’s accuracy in

static and dynamic scenarios relative to current state-of-the-art algorithms. The Cow & Lady dataset features complex surface geometries of static objects in a room recorded by a moving Kinect-1 RGB-D camera whose pose is tracked by a vicon system. The ground truth surface reconstruction is obtained by a Leica TPS-M50 laser scanner. The second dataset is used for evaluating a dynamic scene. It is generated using Gazebo and features a ball rolling on a table observed by a simulated RGB-D camera. The ground truth for distances and gradients are obtained from Gazebo.

The experiments are conducted on a laptop equipped with an AMD Ryzen 7 4800u CPU, 16GB RAM and no GPU. All three algorithms are allowed to multithread on up to 16 threads. Each algorithm is allowed to run as a ROS node with only the playback of the dataset (rosbag) running in the background. Original code implementations for Fiesta² and Voxblox³ were used for the experiments with default parameters. Root Mean Squared Error (RMSE) is used to evaluate the distance field performance. For gradients, we use

$$\text{cosine_similarity}(\mathbf{a}, \mathbf{b}) = \frac{\mathbf{a} \cdot \mathbf{b}}{\|\mathbf{a}\| \cdot \|\mathbf{b}\|} \quad (12)$$

to measure the difference between two gradient vectors \mathbf{a} and \mathbf{b} in terms of their direction.

B. Mapping Results

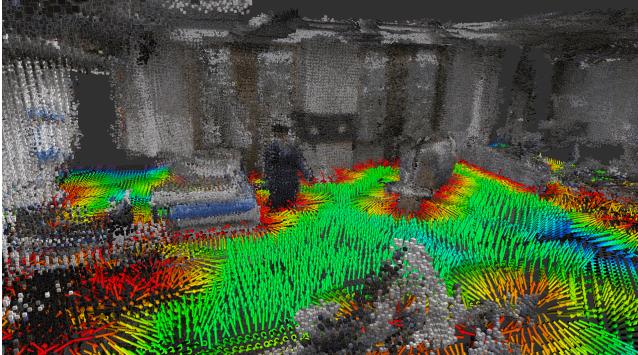


Fig. 4: Example of an IDMP query on a horizontal plane for the Cow & Lady dataset. The colour of the arrows represents the distance to the nearest object. The arrows are normalised gradients that point in the direction away from the nearest object. The updated training points are coloured using RGB data from the camera.

1) Static Scene: Fig. 1 and Fig. 4 present qualitative results where IDMP queries are visualised. For every point in space, a distance (colour) and direction (arrow) away from the closest obstacle can be computed online and on-demand. For both examples, the queries are taken using a regular grid on a plane parallel to the table and floor, respectively. Note that because of the continuous nature of IDMP, the query points can be located anywhere in space and are neither constrained to be on a plane or a regular grid nor bound to a specific spatial resolution.

²<https://github.com/HKUST-Aerial-Robotics/FIESTA>

³<https://github.com/ethz-asl/voxblox>

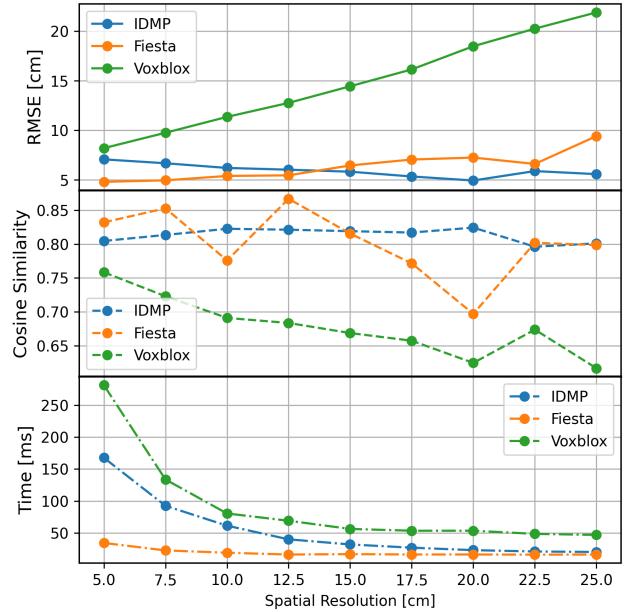


Fig. 5: Quantitative evaluation of distance RMSE (lower is better), gradients cosine similarity (higher is better) and computation time (lower is better) using the Cow & Lady dataset for IDMP, FIESTA, and Voxblox.

To compare IDMP to Voxblox and FIESTA, we compute the RMSE for each framework while varying the *spatial resolution*. For Voxblox/FIESTA, this corresponds to the voxel size. For IDMP we vary the resolution of the training points used to update the Fused Field. Figure 5 shows our experimental results. It shows the distance and gradient accuracy, and the overall time needed to process a single frame.

The accuracy for Voxblox and FIESTA decreases with increasing spatial resolution. In contrast, IDMP's accuracy remains constant, outperforming FIESTA with resolutions higher than 15cm. Note that even though we increase the resolution of our training points, IDMP generates a continuous distance field, that can be queried at any point in space independent of the spatial resolution. Voxblox and FIESTA only generate a discrete distance field in the predefined spatial resolution and where data was observed.

The evaluation of gradients shows the mean of the cosine similarities between the computed gradients and the ground truth. Both IDMP and FIESTA perform better than Voxblox for all spatial resolutions. FIESTA shows slightly better performance than IDMP in some cases; however, at the cost of higher variability.

The last part shows the average computation time for each algorithm to process a frame. For IDMP this includes generating the Frustum Field, performing the fusion, updating dynamic objects and generating the Fused Field. For Fiesta and Voxblox this includes the ESDF/TSDF generation, integration and raycasting. IDMP outperforms Voxblox for every spatial resolution while reaching similar computation times as FIESTA for resolutions larger than 15 cm.

Regarding query times, it should be noted that while Voxblox and Fiesta have constant distance and gradient query times, IDMP has a small overhead time ($2\mu\text{s}$ per point).

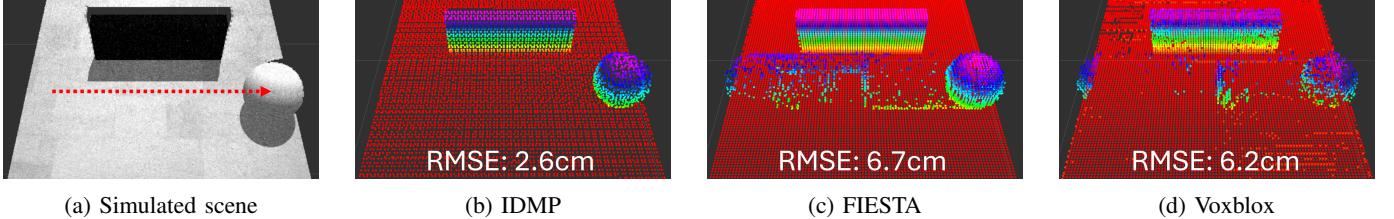


Fig. 6: Comparison of IDMP, FIESTA and Voxblox for a dynamic scene. (a) Shows the stimulation setup, and (b) to (d) show the fused surface points at the end of the sequence. Colours indicate the point's height with a spatial resolution of 1 cm. IDMP handles appropriately moving objects while FIESTA and Voxblox keep artefacts from older frames.

However, the benefit of our method is that queries can be made at any arbitrary resolution post-mapping due to our continuous representation.

2) *Dynamic Scene*: In this experiment, we evaluate how each framework handles dynamic scenes with moving objects. Fig. 6a shows the Gazebo scene with a ball rolling on the table from left to right. The last three subfigures show the surface points at the end of the run for the three frameworks. Both FIESTA and Voxblox show artefacts due to the progressive weighting and integration, while IDMP does not. IDMP results in the best RMSE of 2.6 cm which is more than a factor of 2 better than FIESTA/Voxblox. Updating the GP for a scene of this size takes on average 50 ms per frame while the query step takes 2 μ s per point. Only the time for processing the frame is dependent on the size of the scene, whereas the time needed for the query remains constant.

C. Reactive Planning in Dynamic Scenes

In this experiment, we demonstrate that our method is able to facilitate safe reactive motion and path planning of a robot arm in the presence of dynamic obstacles, which is useful for close-proximity human-robot collaboration tasks. To simulate such a scenario, we task the robot with moving to a goal pose while avoiding a moving ball object, see Fig. 7.

Trajectory optimisers are well suited to such scenarios due to their fast computation and smooth trajectory properties. We utilise CHOMP [7], a gradient-based trajectory optimisation method that iteratively perturbs a given initial trajectory away from obstacles. To do so, it utilises the gradient field to minimise its optimisation objective, which consists of trajectory smoothness and obstacle avoidance terms. Due to a) the ability of our framework to compute updated gradients analytically and b) the need to only query the points \mathcal{P}_o the motion planner requires, we can continuously replan online. The computation time for the query shown in Fig. 7 is approximately 6.4 ms (3190 points at 2 μ s point).

We additionally validate IDMP in combination with CHOMP in a real-world scenario where a human moves into the robot's workspace mid-way during its trajectory (see Fig. 8). Furthermore, we demonstrate the proposed local reactive method in a similar scenario. Fig. 9 shows the repulsive vector in red, the attractive vector in blue and the resulting vector in green. The robot arm follows the resulting vector to quickly and safely avoid the human's moving arm. Because the repulsive vector only uses the information at the robots

current position, querying the Fused Field takes 2 μ s, enabling real-time obstacle avoidance.

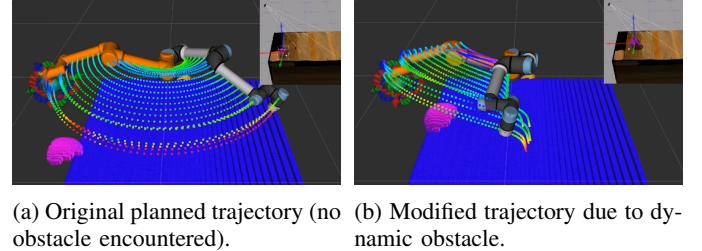


Fig. 7: Re-planning in the presence of a dynamic obstacle. CHOMP approximates the manipulator via 29 spheres. For the trajectory shown, 110 points per sphere are used, resulting in $29 \times 110 = 3190$ query points.

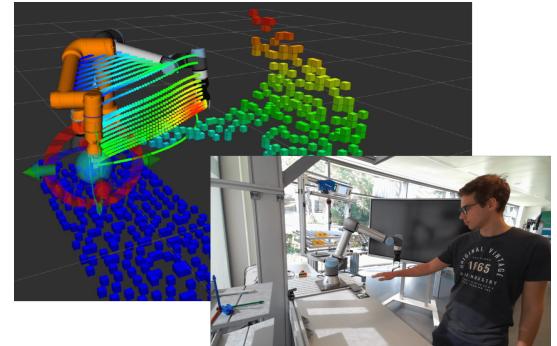


Fig. 8: Real-world experiment of IDMP with CHOMP. Re-planning in the presence of a human arm.

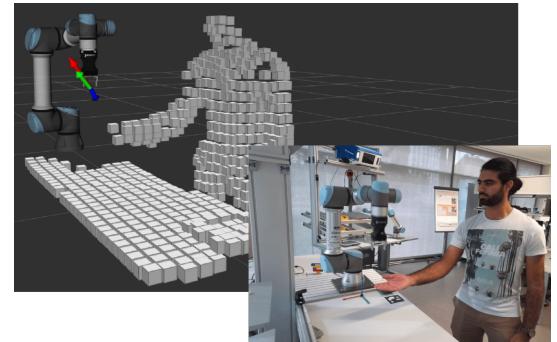


Fig. 9: IDMP and local reactive obstacle avoidance.

VI. CONCLUSION

In this letter, we presented an efficient interactive mapping and planning framework named IDMP. The simple and elegant dynamic updates and fusion are performed by querying

distances and gradients from a temporary latent GP distance field generated with only the points of the current frame. Our framework can run online on a modest PC without a GPU and handles dynamic objects effectively. The fused representation is continuous enabling queries of the Euclidean distance with gradient at an arbitrary spatial resolution. This property enables seamless integration with a gradient-based motion planning framework for HRC applications. Going forward, we aim to further explore IDMP's capabilities, such as multi-resolution sampling and mapping uncertainty, for enhancing motion planning, particularly in HRC applications.

VII. ACKNOWLEDGEMENT

The authors would like to thank Cedric Le Gentil for useful discussions and code collaboration.

REFERENCES

- [1] L. Han, F. Gao, B. Zhou, and S. Shen, "Fiesta: Fast incremental euclidean distance fields for online motion planning of aerial robots," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2019, pp. 4423–4430.
- [2] Y. Bai, Z. Miao, X. Wang, Y. Liu, H. Wang, and Y. Wang, "Vdbblox: Accurate and efficient distance fields for path planning and mesh reconstruction," in *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2023, pp. 7187–7194.
- [3] H. Oleynikova, Z. Taylor, M. Fehr, R. Siegwart, and J. Nieto, "Voxblox: Incremental 3D Euclidean Signed Distance Fields for on-board MAV planning," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. Vancouver, BC: IEEE, Sep. 2017.
- [4] Y. Pan, Y. Kompis, L. Bartolomei, R. Mascaro, C. Stachniss, and M. Chli, "Voxfield: Non-Projective Signed Distance Fields for Online Planning and 3D Reconstruction," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2022.
- [5] L. Wu, K. M. B. Lee, L. Liu, and T. Vidal-Calleja, "Faithful euclidean distance field from log-gaussian process implicit surfaces," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 2461–2468, 2021.
- [6] C. Le Gentil, O.-L. Ouabi, L. Wu, C. Pradalier, and T. Vidal-Calleja, "Accurate gaussian-process-based distance fields with applications to echolocation and mapping," *IEEE Robotics and Automation Letters*, 2023.
- [7] M. Zucker, N. Ratliff, A. D. Dragan, M. Pivtoraiko, M. Klingensmith, C. M. Dellin, J. A. Bagnell, and S. S. Srinivasa, "CHOMP: Covariant hamiltonian optimization for motion planning," *International Journal of Robotics Research*, vol. 32, no. 9-10, pp. 1164–1193, 2013.
- [8] M. Mukadam, J. Dong, X. Yan, F. Dellaert, and B. Boots, "Continuous-time gaussian process motion planning via probabilistic inference," *International Journal of Robotics Research*, vol. 37, no. 11, 2018.
- [9] D. Zhu, C. Wang, W. Wang, R. Garg, S. Scherer, and M. Q.-H. Meng, "VDB-EDT: An Efficient Euclidean Distance Transform Algorithm Based on VDB Data Structure," May 2021.
- [10] L. Schmid, O. Andersson, A. Sulser, P. Pfreundschuh, and R. Siegwart, "Dynablox: Real-time detection of diverse dynamic objects in complex environments," *IEEE Robotics and Automation Letters*, 2023.
- [11] J. J. Park, P. Florence, J. Straub, R. Newcombe, and S. Lovegrove, "Deep sdf: Learning continuous signed distance functions for shape representation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 165–174.
- [12] A. Gropp, L. Yariv, N. Haim, M. Atzmon, and Y. Lipman, "Implicit geometric regularization for learning shapes," *arXiv preprint arXiv:2002.10099*, 2020.
- [13] M. Pantic, C. Cadena, R. Siegwart, and L. Ott, "Sampling-free obstacle gradients and reactive planning in neural radiance fields (nerf)," *arXiv preprint arXiv:2205.01389*, 2022.
- [14] J. Ortiz, A. Clegg, J. Dong, E. Sucar, D. Novotny, M. Zollhoefer, and M. Mukadam, "isdf: Real-time neural signed distance fields for robot perception," *arXiv preprint arXiv:2204.02296*, 2022.
- [15] B. Lee, C. Zhang, Z. Huang, and D. D. Lee, "Online continuous mapping using gaussian process implicit surfaces," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019.
- [16] L. Wu, K. M. B. Lee, C. Le Gentil, and T. Vidal-Calleja, "Log-GPIS-MOP: A Unified Representation for Mapping, Odometry, and Planning," *IEEE Transactions on Robotics*, pp. 1–17, 2023.
- [17] L. Liu, S. Fryc, L. Wu, T. L. Vu, G. Paul, and T. Vidal-Calleja, "Active and Interactive Mapping With Dynamic Gaussian Process Implicit Surfaces for Mobile Manipulators," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 3679–3686, Apr. 2021.
- [18] L. Wu, C. Le Gentil, and T. Vidal-Calleja, "Pseudo inputs optimisation for efficient gaussian process distance fields," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2023.
- [19] S. Robla-Gomez, V. M. Becerra, J. R. Llata, E. Gonzalez-Sarabia, C. Torre-Ferrero, and J. Perez-Oria, "Working Together: A Review on Safe Human-Robot Collaboration in Industrial Environments," *IEEE Access*, vol. 5, pp. 26754–26773, 2017.
- [20] S. Patil, V. Vasu, and K. V. S. Srinadh, "Advances and perspectives in collaborative robotics: a review of key technologies and emerging trends," *Discover Mechanical Engineering*, vol. 2, no. 1, p. 13, 2023.
- [21] A. M. Zanchettin, N. M. Ceriani, P. Rocco, H. Ding, and B. Matthias, "Safety in human-robot collaborative manufacturing environments: Metrics and control," *IEEE Transactions on Automation Science and Engineering*, vol. 13, no. 2, pp. 882–893, 2016.
- [22] F. Schirmer, P. Kranz, J. Schmitt, and T. Kaupp, "Anomaly detection for dynamic human-robot assembly: Application of an lstm-based autoencoder to interpret uncertain human behavior in hrc," in *Companion of ACM/IEEE International Conf. on Human-Robot Interaction*, 2023.
- [23] C. Morato, K. N. Kaipa, B. Zhao, and S. K. Gupta, "Toward Safe Human Robot Collaboration by Using Multiple Kinects Based Real-Time Human Tracking," *Journal of Computing and Information Science in Engineering*, vol. 14, no. 1, p. 011006, Mar. 2014.
- [24] F. Schirmer, P. Kranz, B. Bhat, C. Rose, J. Schmitt, and T. Kaupp, "Towards a path planning and communication framework for seamless human-robot assembly," in *19th Annual ACM/IEEE International Conference on Human-Robot Interaction*, 2024.
- [25] C. Tonola, M. Faroni, N. Pedrocchi, and M. Beschi, "Anytime informed path re-planning and optimization for human-robot collaboration," in *2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN)*. IEEE, 2021, pp. 997–1002.
- [26] M. Faroni, M. Beschi, and N. Pedrocchi, "An MPC Framework for Online Motion Planning in Human-Robot Collaborative Tasks," in *2019 24th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*. Zaragoza, Spain: IEEE, Sep. 2019.
- [27] F. Flacco, T. Kroeger, A. De Luca, and O. Khatib, "A Depth Space Approach for Evaluating Distance to Objects," *Journal of Intelligent & Robotic Systems*, vol. 80, no. 1, pp. 7–22, Dec. 2015.
- [28] F. Flacco and A. De Luca, "Real-Time Computation of Distance to Dynamic Obstacles With Multiple Depth Sensors," *IEEE Robotics and Automation Letters*, vol. 2, no. 1, pp. 56–63, Jan. 2017.
- [29] C. E. Rasmussen and C. K. Williams, *Gaussian Processes for Machine Learning*. Cambridge, Mass.: MIT Press, 2006.
- [30] W. Martens, Y. Poffet, P. R. Soria, R. Fitch, and S. Sukkarieh, "Geometric priors for gaussian process implicit surfaces," *IEEE Robotics and Automation Letters (RA-L)*, pp. 373–380, 2017.
- [31] C. Le Gentil, M. Vayugundla, R. Giubilato, W. Stürzl, T. Vidal-Calleja, and R. Triebel, "Gaussian process gradient maps for loop-closure detection in unstructured planetary environments," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 1895–1902.
- [32] S. Särkkä, "Linear operators and stochastic partial differential equations in gaussian process regression," in *Artificial Neural Networks and Machine Learning—ICANN 2011: 21st International Conference on Artificial Neural Networks, Espoo, Finland, June 14–17, 2011, Proceedings, Part II 21*. Springer, 2011, pp. 151–158.
- [33] O. Khatib, "Real-time obstacle avoidance for manipulators and mobile robots," *The International Journal of Robotics Research*, vol. 5, no. 1, 1986.