A Probabilistic Approach to Multi-Modal Adaptive Virtual Fixtures

Maximilian Mühlbauer[1,2], Thomas Hulin[2], Bernhard Weber[2], Sylvain Calinon[3,4], Freek Stulp[2], Alin Albu-Schäffer[2,1], João Silvério[2]

Abstract—Virtual Fixtures (VFs) provide haptic feedback for teleoperation, typically requiring distinct input *modalities* for different phases of a task. This often results in vision- and position-based fixtures. Vision-based fixtures, particularly, require the handling of visual uncertainty, as well as target appearance/disappearance for increased flexibility. This creates the need for principled ways to add/remove fixtures, in addition to uncertainty-aware assistance regulation. Moreover, the arbitration of different modalities plays a crucial role in providing an optimal feedback to the user throughout the task.

In this paper, we propose a Mixture of Experts (MoE) model that synthesizes visual servoing fixtures, elegantly handling full pose detection uncertainties and teleoperation goals in a unified framework. An arbitration function combining multiple vision-based fixtures arises naturally from the MoE formulation, leveraging uncertainties to modulate fixture stiffness and thus the degree of assistance. The resulting visual servoing fixtures are then fused with position-based fixtures using a Product of Experts (PoE) approach, achieving guidance throughout the complete workspace. Our results indicate that this approach not only permits human operators to accurately insert printed circuit boards (PCBs) but also offers added flexibility and retains the performance level of a baseline with carefully handtuned VFs, without requiring the manual creation of VFs for individual connectors. An exemplary video showcasing our method is available at: **https://youtu.be/6BDB3g0QyFg**

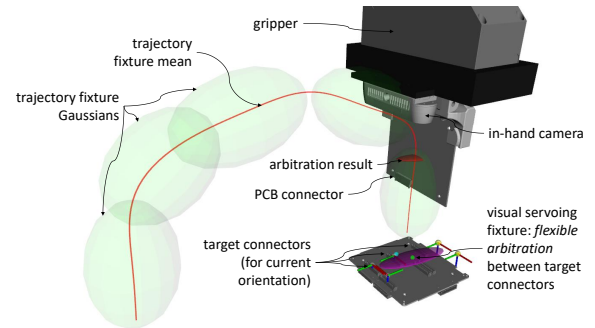*Index Terms*—Telerobotics and Teleoperation, Space Robotics and Automation, Assembly

Fig. 1: Fusion of position-based and visual servoing fixtures. Trajectory covariance (green ellipsoids) and visual servoing covariance (purple ellipsoid) are used to calculate the final wrench and uncertainty (red ellipsoid).

I. INTRODUCTION

VIRTUAL FIXTURES (VFs) play an important role in shared control as haptic aids by providing force feedback to the human operator [1, 2], assisting in the execution of tasks. The type of assistance to be provided depends on the task at hand and/or the phase of the task that the robot is currently executing. An insertion task typically requires an approach phase, where trajectory guidance is needed, followed by an insertion phase, where visual guidance takes over. Position and vision [3] or even forces [4] are examples of fixture input *modalities*. The arbitration of different modalities, i.e. when

each type of modality should be activated and by how much, is an important open problem in shared control.

Although established frameworks for position-based trajectory fixtures exist [2], visual servoing fixture formulations which are robust to object appearance/disappearance and provide assistance based on target uncertainty are lacking. Providing visual assistance in an adaptive manner is known to be challenging. On the one hand, object poses may not be known with sufficient certainty in advance, making it impossible to use constant fixtures that are designed once and rarely change. On the other hand, many external factors, such as lighting conditions, are non-trivial to model and may drastically degrade perception performance. Imperfect visual measurements are a reality in robotics – yet, we have to rely on them even though they might be uncertain. Systems using this uncertainty are thus better equipped to succeed in challenging environments, such as in-orbit scenarios.

In this work we introduce a multi-modal VFs framework that leverages probability theory to seamlessly combine vision- and position-based fixtures (Fig. 1), using probability distributions on $\mathbb{R}^3 \times \mathcal{S}^3$ (Section III) to take orientations into account. Our contribution is two-fold. First, we propose a probabilistic *Mixture of Experts* (MoE) [5] approach to automate the arbitration of uncertain visual servoing fixtures (Section IV). Under the MoE, each *expert* is a probability distribution that models the uncertainty of a detected object. State- and perception-dependent *gating functions* regulate the influence of each expert such that haptic assistance changes dynamically as the robot interacts with the environment. The formulation furthermore elegantly handles object appearance/disappearance. As a consequence, the effort to design visual servoing fixtures by hand is minimal. Second, we propose the fusion of visual servoing fixtures with position-based fixtures using a *Product of Experts* (PoE) (Section V) resulting in a principled arbitration

between multiple assistance modalities that is generic to a wide range of shared control problems. By treating position-based fixtures as mixture models [2, 6], both modalities are MoEs, allowing for their seamless fusion. Relying on Gaussian-distributed experts, this corresponds to a product of Gaussians, a well-known approach for fusing information from different sources.

We evaluate the proposed approach on a PCB connector assembly task where a position-based VF guides the operator towards the insertion region and a visual servoing VF provides fine-grained assistance guiding the insertion (Section VI). We use two torque-controlled KUKA light-weight arms as robot and haptic device (Fig. 2).

II. RELATED WORK

A. Adaptive and Probabilistic Virtual Fixtures

The first type of fixtures initialize static position-based fixtures from visual measurements. Selvaggio et al. [7] detect limit switches and plan VFs for reaching and manipulating these switches. Pruks and Ryu [8] use visual measurements to allow users to interactively define VFs based on geometric primitives, for example cylindrical fixtures for circles detected in the image. Contrarily, Hager [9] and Bettini et al. [10] calculate VF forces directly based on camera images for 2D line following thus creating dynamic visual servoing fixtures. The visual servoing fixture of Wu et al. [11] is closest to ours, however without probabilistically considering multiple fixtures. A major limitation of these works is however that fixtures can only be generated when the manipulation target is visible in the camera. Furthermore, they only consider the visual modality, unlike our approach which considers position-based fixtures as well.

Static position-based fixtures can be created using probabilistic methods. Aarno et al. [12] extract lines as trajectory fixtures from demonstrations, selecting the active VF based on the estimate of a jointly learned Hidden Markov Model. Raiola et al. [2] present a probabilistic arbitration between a library of probabilistic VFs based on Gaussian Mixture Models (GMMs). Havoutis and Calinon [6] show how Task-Parametrized Gaussian Mixture Models (TP-GMMs) [13] can be used to define fixtures which can adapt to changing start and goal points. These approaches do however not incorporate assistance based on uncertain visual measurements.

B. Virtual Fixture Arbitration

Having multiple VFs active at the same time, the need for an arbitration function that combines them arises. This extends the classical concept of *arbitration* as division of control authority between human and robot. One possibility is to have phase-dependent VFs and to activate them sequentially [14]. Other approaches allow multiple fixtures to be defined at once and use an arbitration component to switch between them. Selvaggio et al. [7] use a passivity controller to stabilize a hard assignment switching operation between fixtures. Abi-Farraj et al. [15] use fixtures guiding the operator to possible grasping poses. Manually tuned scaling factors allow them to have all fixtures active at the same time. In our previous work [3] we hand-designed an arbitration function between position-

and vision-based fixtures. A limitation of such approaches is that smooth switching between target poses requires the handcrafting of either a stabilizing controller or an arbitration function.

Also between fixtures and the operator, arbitration needs to be performed. Probabilistic formulations for arbitration have been proposed [2, 16–18]. While the implementations differ, most of these works use a scalar value for assigning weights to fixtures, ruling out degree-of-freedom-specific arbitration. In contrast, Zeestraten et al. [18] modulate human commands by a hand-designed covariance matrix allowing for a seamless arbitration with static Gaussian-based fixtures and treating each degree of freedom individually. Michel et al. [19] use a different approach by learning a full stiffness matrix, where uncertain directions generate lower VF stiffness. Our work combines the best of these approaches by using adaptively scaled stiffness matrices computed from the covariance of dynamic fixtures.

C. Machine Learning Approaches

In robotics, MoEs have been applied in locomotion learning [20], imitation learning [13, 21, 22] and shared control [2], while PoEs are a popular approach at the intersection of learning and control [23, 6, 24, 22]. The *expert* is a simple model which, combined with other experts, improves model performance over the single-expert case. Gaussian-based experts, where the expert is modelled as a Gaussian distribution, are among the most popular expert models. The MoE model corresponds to an "or" operation, performing a weighted sum of the density functions of the experts.

In contrast, a PoE model (product of Gaussians) corresponds to an "and" operation where all constraints must be approximately satisfied. Many PoE models consist of MoE-based experts. TP-GMMs [13] learn local models of skill demonstrations, encoding them as GMMs. For different input values, such as time, predictions from the local GMMs are combined using a PoE. The concept has been extended to the fusion of controllers [22] and assistive teleoperation [6, 18], where it was however mainly used to arbitrate between user and automation. We believe that the potential of MoE-based experts goes beyond modeling demonstrations and provides the flexibility to represent different fixture modalities. Although PoE approaches have been used with vision [23], to the best of our knowledge they have not been used with MoE-based vision experts nor in shared control.

III. BACKGROUND

A. Teleoperation System and Virtual Fixtures

We assume two gravity-compensated, impedance-controlled manipulators (Fig. 2) where Cartesian wrenches $\boldsymbol{w}_{\text{ee}} \in \mathbb{R}^6$ are commanded at the end effector, with joint torques computed from $\boldsymbol{\tau} = \boldsymbol{J}^\top \boldsymbol{w}_{\text{ee}}$ [25]. The Cartesian wrenches of remote and input robots are computed with

$$\boldsymbol{w}_{\text{ee,remote}} = \alpha \left(\boldsymbol{K} \Delta \boldsymbol{x} + \boldsymbol{D} \Delta \dot{\boldsymbol{x}} \right) + \boldsymbol{w}_{\text{VF}} \tag{1}$$

$$\boldsymbol{w}_{\text{ee,input}} = -\alpha \boldsymbol{Ad}_{\text{ir}} \boldsymbol{w}_{\text{ee,remote}} \tag{2}$$

where the adjoint $\boldsymbol{Ad}_{\text{ir}}$ transforms wrenches from the remote robot to the haptic input device. This position-computed force
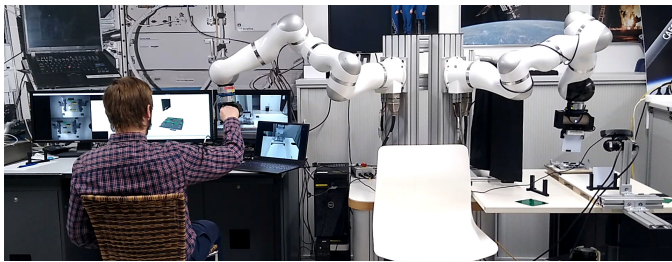
Fig. 2: The teleoperation setup with haptic input device on the left and remote device on the right side.

architecture does not require a force-torque sensor at the end effector. The factor α scales motions between both robots, $\Delta\boldsymbol{x}$ and $\Delta\dot{\boldsymbol{x}}$ corresponds to their relative displacement and $\boldsymbol{K}$, $\boldsymbol{D}$, are positive definite constant stiffness and damping gain matrices. The virtual fixture wrench $\boldsymbol{w}_{\mathrm{VF}}$ is applied to the end effector of the remote robot only, which thus achieves high accuracy. The user also receives useful feedback through the coupling introduced by α. In this work we assume that $\boldsymbol{w}_{\mathrm{VF}}$ is a combination of individual wrenches, associated with different virtual fixtures, computed as

$$\boldsymbol{w}_{\mathrm{VF},j} = \boldsymbol{K}_{\mathrm{VF},j}\mathrm{Log}_{\boldsymbol{x}_{\mathrm{ee}}}(\boldsymbol{x}_{\mathrm{VF}})\,, \qquad (3)$$

where $\boldsymbol{x}_{\mathrm{ee}}$ is the end effector pose, $\boldsymbol{K}_{\mathrm{VF},j}$ and $\boldsymbol{x}_{\mathrm{VF},j}$ are the stiffness and attractor point of the j-th fixture. $\mathrm{Log}_{\boldsymbol{x}_{\mathrm{ee}}}(\boldsymbol{x}_{\mathrm{VF}})$ denotes the $\mathbb{R}^3\times\mathcal{S}^3$ logarithmic map [21] of $\boldsymbol{x}_{\mathrm{VF}}$ at $\boldsymbol{x}_{\mathrm{ee}}$, which is the on-manifold equivalent to the Euclidean $\boldsymbol{x}_{\mathrm{VF},j} - \boldsymbol{x}_{\mathrm{ee}}$, allowing us to also take the orientation into account. We further assume that each fixture j can be based on a different input modality, e.g. vision or position and has one attractor point and stiffness matrix. We denote those attractors as $\boldsymbol{x}_{\mathrm{VS}}$ respectively $\boldsymbol{x}_{\mathrm{PB}}$ instead of $\boldsymbol{x}_{\mathrm{VF},j}$ in Sections IV and V-A. Section V introduces our proposed arbitration of different fixtures.

B. On-Manifold Probabilities

Object pose uncertainties appear at position and orientation levels. To be able to model both, we use an on-manifold approach with Gaussian distributions. We use a pose defined as the Cartesian product of the 3-dimensional Euclidean space and the unit quaternion manifold,[1] $\boldsymbol{x} \in \mathbb{R}^3 \times \mathcal{S}^3$, whose distribution is parameterized by a mean $\boldsymbol{\mu} \in \mathbb{R}^3 \times \mathcal{S}^3$ and a covariance matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{6\times6}$ in the tangent space of $\boldsymbol{\mu}$. Since $\mathcal{S}^3$ is a compact Lie group, it admits a bi-invariant metric allowing the computation of geodesics using the Lie group exponential [27, 28]. This allowed [21] to express tangent vectors and covariance matrices of $\mathcal{S}^3$ as elements of $\mathbb{R}^3$ and $\mathbb{R}^{3\times3}$ respectively, here, we follow the same approach. We employ the Gaussian distribution proposed in [21, 29] to compute the probability of $\boldsymbol{x}$:

$$\mathcal{N}\left(\boldsymbol{x}|\boldsymbol{\mu},\boldsymbol{\Sigma}\right) = \frac{1}{\sqrt{(2\pi)^d\,|\boldsymbol{\Sigma}|}}e^{-\frac{1}{2}\mathrm{Log}_{\boldsymbol{\mu}}(\boldsymbol{x})^\top\boldsymbol{\Sigma}^{-1}\mathrm{Log}_{\boldsymbol{\mu}}(\boldsymbol{x})}. \quad (4)$$

[1]To avoid issues with $\mathcal{S}^3$ double-covering $SO(3)$, we wrapped the logarithm at a full rotation, ensuring that $\mathrm{Log}_{\boldsymbol{q}}(-\boldsymbol{q}) = \boldsymbol{0}$.

From N samples, Maximum Likelihood Estimation (MLE) [30] is computed iteratively using the Fréchet mean [21]

$$\Delta = \frac{1}{N}\sum_{i=1}^{N}\mathrm{Log}_{\boldsymbol{\mu}}(\boldsymbol{x}_i)\,, \qquad \boldsymbol{\mu} \leftarrow \mathrm{Exp}_{\boldsymbol{\mu}}(\Delta)\,, \qquad (5)$$

upon convergence of (5), the covariance matrix is given as

$$\boldsymbol{\Sigma} = \frac{1}{N-1}\sum_{i=1}^{N}\mathrm{Log}_{\boldsymbol{\mu}}(\boldsymbol{x}_i)\,\mathrm{Log}_{\boldsymbol{\mu}}(\boldsymbol{x}_i)^\top\,. \qquad (6)$$

The logarithm function $\mathrm{Log}_{\boldsymbol{\mu}}(.)$ maps points from the manifold to the tangent space at $\boldsymbol{\mu}$. The exponential map $\mathrm{Exp}_{\boldsymbol{\mu}}(.)$ maps a vector from the tangent space at $\boldsymbol{\mu}$ onto the manifold. For the orientation part of the pose, we use the functions defined in [21] for unit quaternions. Vectors in tangent space can be moved from one linearization point to another using *parallel transport* compensating for different base vector orientations at different points $\boldsymbol{\mu}$. Using the parallel transport defined in [21], we transport covariance matrices between different tangent spaces. Note that other Lie-group approaches rely on expressions e.g. for the product of Gaussians [31] with very similar results on $\mathcal{S}^3$ to using the Riemannian Levi-Civita parallel transport [21] as we have experimentally verified.

IV. PROBABILISTIC VISUAL-SERVOING FIXTURES

Formally we assume that, at any moment, a number of $M_{\mathrm{VS}} \geq 0$ visual servoing fixtures may be active, each trying to bring the robot towards an object in its field of view with different $\boldsymbol{x}_{\mathrm{VS}}$. As the field of view changes with the end effector position, the number of active fixtures and their parameters depend on $\boldsymbol{x}_{\mathrm{ee}}$. Hence we treat each fixture as a conditional distribution $p_m(\boldsymbol{x}_{\mathrm{VS}}|\boldsymbol{x}_{\mathrm{ee}})$ with $m = 1,\ldots,M_{\mathrm{VS}}$ that is computed from the uncertainty of the predicted poses (Section IV-A). When $M_{\mathrm{VS}} > 1$, several fixtures pull the end effector simultaneously. In order to ensure both local assistance and the capability to switch between fixtures we propose a MoE that outputs a unimodal distribution $p(\boldsymbol{x}_{\mathrm{VS}}|\boldsymbol{x}_{\mathrm{ee}})$ from the M_{VS} candidates (Sections IV-B and IV-C). With this distribution we are able to compute not only an attractor point that drives the remote robot pose, but also stiffness gains that regulate the required precision while tracking it (Section IV-D).

Additional, desired assistive behaviors can be easily created by adding hand-parametrized Gaussian distributions, with gating functions enabling a user-defined regulation of transitions between local experts. We demonstrate these features experimentally in Section VI through the creation of *dead zones*, initialization experts and the deactivation of undesired assistance along certain axes.

A. Probabilistic Fixtures from Visual Uncertainty

In this section we propose an algorithm that outputs a probability distribution per PCB connector in the camera image, which can readily be used for the PCB connector assembly task in Section VI. Using an in-hand camera leads to an increasing accuracy when approaching the target as the connector's size in pixels increases. Previously [3], we used a fixed grayscale threshold to binarize the intensity image I
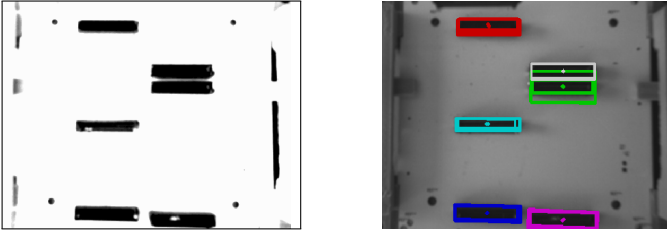
Fig. 3: Probabilistic target connector extraction using multiple grayscale threshold values. Different threshold values lead to "soft" borders (left side, intermediate gray values) while the core of the connectors and the outside, where all threshold values give the same result, are uniformly black/white. This results in different rectangles (right side). Converted to 6DoF poses, we treat the grouped detections as samples from a Gaussian distribution.

and extract targets using OpenCV [32] rectangle extraction. Depending on illumination conditions and camera settings, the optimal threshold differs. Furthermore, shadows cast by the connectors make it very difficult to find a single value for optimally extracting the target connector. Using the idea of soft grayscale thresholds [33, 34] we propose to extract potential target connectors (Fig. 3) on a symmetric range of different grayscale values $T \in T_{nom} \pm \{\Delta T_0, \Delta T_1, \dots\}$ around the nominal threshold T_{nom} with threshold increments ΔT_i. As shown in Algorithm 1, we group these extractions using their 2D coordinates (*groupByXy*), assigning exactly one matching rectangle per grayscale threshold value. Then, we convert them to 6D poses (*convertTo6D*) and, for each connector m, treat them as set of N_m individual samples drawn from a noisy measurement of the target.

Using (5), the MLE estimate of the samples is computed to approximate $p_m(\boldsymbol{x}_{\mathrm{VS}}|\boldsymbol{x}_{\mathrm{ee}}) = \mathcal{N}(\boldsymbol{x}_{\mathrm{VS}}|\boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m)$, where $\boldsymbol{\Sigma}_m$ provides a measure of the uncertainty associated with a connector. As final step of the detection, we associate new measurements with already tracked connectors based on their distance. If no existing tracked connector is found, a new tracking instance is created using the mean and covariance of the measurement as initial state. In case a tracked connector exists, we employ Kalman filtering for data fusion.

B. On-Manifold Mixture of Experts

Having represented the uncertainty of candidate VFs in the robot workspace with $p_m(\boldsymbol{x}_{\mathrm{VS}}|\boldsymbol{x}_{\mathrm{ee}})$, we express $p(\boldsymbol{x}_{\mathrm{VS}}|\boldsymbol{x}_{\mathrm{ee}})$ in a unified manner using a MoE model [5, 30]

$$p(\boldsymbol{x}_{\mathrm{VS}}|\boldsymbol{x}_{\mathrm{ee}}) = \sum_{m=1}^{M_{\mathrm{VS}}} \hat{h}_m(\boldsymbol{x}_{\mathrm{ee}}, \boldsymbol{\mu}_m) p_m(\boldsymbol{x}_{\mathrm{VS}}|\boldsymbol{x}_{\mathrm{ee}}). \quad (7)$$

Our proposed *gating function* h_m takes into account the robot end effector pose and the predicted expert locations $\boldsymbol{\mu}_m$ to compute an on-manifold, distance-based metric that determines the influence of each expert through

$$h_m(\boldsymbol{x}_{\mathrm{ee}}, \boldsymbol{\mu}_m) = \exp\left(-\frac{1}{2}\mathrm{Log}_{\boldsymbol{x}_{ee}}(\boldsymbol{\mu}_m)^\top \boldsymbol{L} \mathrm{Log}_{\boldsymbol{x}_{ee}}(\boldsymbol{\mu}_m)\right) + \gamma \quad (8)$$

where $\boldsymbol{L}$ is a hyperparameter regulating the influence of nearby points and γ is a regularization factor stabilizing (7) numerically when far from the objects, then assigning near-equal probabilities to all objects. For our experiments, we set $\boldsymbol{L} = \mathrm{diag}(l_x^2, l_y^2, l_z^2, l_{wx}^2, l_{wy}^2, l_{wz}^2)^{-1}$, enabling us to specify the

---

Algorithm 1 Probabilistic target connector detection on grayscale image $I(\boldsymbol{x}_{\mathrm{ee}})$ with threshold values T_i.

> rects $\leftarrow$ empty list
> **for** i in len(T) **do**
> $B \leftarrow I > T_i$
>   rects $\leftarrow$ rects + minAreaRects(B)   ▷ list append
> **end for**
> sorted_rects $\leftarrow$ groupByXy(rects)   ▷ one rect per T_i
> **for** m in len(*sorted_rects*) **do**
>   6d_det $\leftarrow$ $\boldsymbol{x}_{\mathrm{ee}} \cdot$ convertTo6d(sorted_rects[m])
> $\boldsymbol{\mu}_m \leftarrow$ mean(6d_det)   ▷ Eq. (5)
> $\boldsymbol{\Sigma}_m \leftarrow$ cov$_{\boldsymbol{\mu}_m}$(6d_det)   ▷ Eq. (6)
> **end for**

relevance of each direction. Our chosen gating function can be interpreted as a linear combination of a RBF kernel and a constant kernel [35]. (8) ensures a peaked assignment when close to one connector while assigning very similar weights when far from all connectors. The factors l_i and γ can be used to adjust the gating function to the scale of the problem. Smaller values l_i increase the peak while a smaller γ increases the distance required to assign similar weights to all targets. We finally normalize $\hat{h}_m(\boldsymbol{x}_{\mathrm{ee}}, \boldsymbol{\mu}_m) = h_m(\boldsymbol{x}_{\mathrm{ee}}, \boldsymbol{\mu}_m) / \sum_j^{M_{\mathrm{VS}}} h_j(\boldsymbol{x}_{\mathrm{ee}}, \boldsymbol{\mu}_m)$ to ensure that the value of all gating functions sums to 1.

C. Unimodal Approximation of the Multi-Modal MoE

Despite unifying predictions from different experts, (7) is by design multi-modal, which is not well-suited to our VF implementation requiring a single attractor point. To mitigate this issue, we rely on the expectation and covariance of $\boldsymbol{x}_{\mathrm{VS}}$ under $p(\boldsymbol{x}_{\mathrm{VS}}|\boldsymbol{x}_{\mathrm{ee}})$. Since the experts are Gaussian, the resulting distribution can be approximated as a uni-modal Gaussian. This approximation is often referred to as *moment matching*, see [13, 30] for derivations. Similarly to III-B, the mean is computed iteratively, this time using the means of each expert $\boldsymbol{\mu}_m$ and their importance $\hat{h}_m$

$$\Delta = \sum_{m=1}^{M_{\mathrm{VS}}} \hat{h}_m \mathrm{Log}_{\boldsymbol{\mu}_{\mathrm{VS}}}(\boldsymbol{\mu}_m), \quad \boldsymbol{\mu}_{\mathrm{VS}} \leftarrow \mathrm{Exp}_{\boldsymbol{\mu}_{\mathrm{VS}}}(\Delta). \quad (9)$$

The covariance computation is adapted to the manifold using

$$\boldsymbol{\Sigma}_{\mathrm{VS}} = \sum_{m=1}^{M_{\mathrm{VS}}} h_m \left(\boldsymbol{\Sigma}_{||\boldsymbol{\mu}_m}^{\boldsymbol{\mu}_{\mathrm{VS}}} + \mathrm{Log}_{\boldsymbol{\mu}_{\mathrm{VS}}}(\boldsymbol{\mu}_m) \mathrm{Log}_{\boldsymbol{\mu}_{\mathrm{VS}}}(\boldsymbol{\mu}_m)^\top\right), \quad (10)$$

where $\boldsymbol{\Sigma}_{||\boldsymbol{\mu}_m}^{\boldsymbol{\mu}_{\mathrm{VS}}}$ denotes $\boldsymbol{\Sigma}_m$ mapped from the tangent space of $\boldsymbol{\mu}_m$ to that of $\boldsymbol{\mu}_{\mathrm{VS}}$ using parallel transport. Note that this corresponds to the second moment of the multi-modal distribution [13], unlike [21] where by omitting the vector outer product $\mathrm{Log}_{\boldsymbol{\mu}_{\mathrm{VS}}}(\boldsymbol{\mu}_m) \mathrm{Log}_{\boldsymbol{\mu}_{\mathrm{VS}}}(\boldsymbol{\mu}_m)^\top$ it was the result of a linear combination of Gaussians. Under (9), we use $\boldsymbol{\mu}_{\mathrm{VS}}$ as the attractor point in (3). Due to our choice of h_m, (10) matches $\boldsymbol{\Sigma}_m$ in the vicinity of connector m, increasing as the end effector moves away. For this reason, we use $\boldsymbol{\Sigma}_{\mathrm{VS}}$ to design the stiffness $\boldsymbol{K}$ associated with the fixture.

D. Variable Stiffness Control

We use the precision matrix $\boldsymbol{P}_{\text{VS}} = \boldsymbol{\Sigma}_{\text{VS}}^{-1}$ to scale the stiffness of the resulting visual servoing fixture. With this gain design we ensure that directions that have larger variance allow for more freedom to the operator, while directions with low variance are stricter in enforcing the visual servoing fixture. For this, the elements of $\boldsymbol{K}_{\text{VS}}$ are set elementwise

$$K_{\text{VS},ij} = K_{ij}\min\left(\max\left(\eta\left(P_{ij} - \kappa\right), 0\right), 1\right) \quad (11)$$

where Pij is the entry of $\boldsymbol{P}_{\text{VS}}$ at indices i, j. Precision entries $< \kappa$ result in zero fixture stiffness, entries $> \frac{1}{\eta} + \kappa$ in full stiffness. Values in between are linearly scaled.

V. PRODUCT OF EXPERTS FOR MULTI-MODAL VIRTUAL FIXTURE ARBITRATION

The MoE-based formulation for visual servoing fixtures introduced in Section IV is well-suited to being combined with other fixtures via a *Product of Experts* (PoE)

$$p(\boldsymbol{x}) = \prod_{j=1}^{P} p_j(\boldsymbol{x}), \quad (12)$$

particularly, if other fixtures are also modeled as MoEs using Gaussian experts. Such PoE formulations are used in well-known TP-GMMs [13] and variations thereof [22], where motion demonstrations are encoded locally in GMMs and later fused via a Gaussian product to compute a global policy for the robot. We introduce a PoE formulation where experts are not necessarily learned but can also be instantiated from vision, leveraging our proposed visual servoing MoE.

A multi-modal VF formulation with position- and vision-based fixtures thus has experts responsible for vision- (Section IV) and position-based assistance. In a PoE-based formulation, the arbitration between them arises naturally from the Gaussian product. In this section we explain how we achieve multi-modal VF arbitration with position-based fixtures (section V-A) and fusion at action level (section V-B) using an on-manifold Gaussian product [21].

A. Probabilistic Position-based Fixtures

We define probabilistic position-based trajectory fixtures using GMMs [21] on the manifold $\mathbb{R}^1 \times \mathbb{R}^3 \times \mathcal{S}^3$. From a dataset of pose trajectories $\{t_i, \boldsymbol{x}_i\}_{i=1}^{N}$, where $t \in \mathbb{R}^1$ is normalized using dynamic time warping and $\boldsymbol{x} \in \mathbb{R}^3 \times \mathcal{S}^3$ represents a pose, we approximate the joint distribution between time and pose using a GMM with M_{PB} components, i.e. $\begin{bmatrix} t \\ \boldsymbol{x}^\top \end{bmatrix} \sim \sum_{m=1}^{M_{\text{PB}}} \pi_m \mathcal{N}\left(\begin{bmatrix} t \\ \boldsymbol{x}^\top \end{bmatrix} | \boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m\right)$. We subsequently treat the pose elements of the GMM as the position-based fixture attractor $\boldsymbol{x}_{\text{PB}}$ and use Gaussian mixture regression to compute the conditional distribution of $\boldsymbol{x}_{\text{PB}}$ given time,

$$p(\boldsymbol{x}_{\text{PB}}|t) = \sum_{m=1}^{M_{\text{PB}}} \pi_m(t) \mathcal{N}(\boldsymbol{x}_{\text{PB}}|\boldsymbol{\mu}_{m|t}, \boldsymbol{\Sigma}_{m|t}). \quad (13)$$

Note the similarity between (13) and (7) – both trajectory and visual servoing fixtures are MoEs. The multi-modal distribution (13) is subsequently approximated by a single Gaussian,
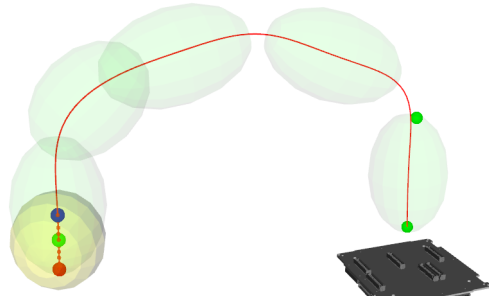

Fig. 4: Probabilistic position-based trajectory fixture based on a GMM. Individual Gaussians (light green) define a mean trajectory (red) evaluated at discrete points with corresponding covariance (yellow). A set of such points (red dots) around the projection of the current end effector pose (green dot) with closest covariance is sent to the real-time controller for interpolation.

similarly to Section IV-C, ensuring that the trajectory fixture contributes with one single expert to the PoE, i.e.

$$p(\boldsymbol{x}_{\text{PB}}|t) = \mathcal{N}(\boldsymbol{x}_{\text{PB}}|\boldsymbol{\mu}_{\text{PB}}, \boldsymbol{\Sigma}_{\text{PB}}). \quad (14)$$

For the details on the computation of (13)–(14), particularly $\boldsymbol{\mu}_{\text{PB}}, \boldsymbol{\Sigma}_{\text{PB}}$, the reader is referred to [13].

A position-based fixture provides assistance by guiding the end effector towards a trajectory. To achieve this behavior we compute D Gaussian distributions (14) given D equally spaced samples of t in the training interval, yielding $\{\boldsymbol{\mu}_d, \boldsymbol{\Sigma}_d\}_{d=1}^{D}$. We then select the two closest means to the current end effector position $\boldsymbol{x}_{ee}$ and perform on-manifold linear interpolation between them to create the expert.

B. PoE at Action Level

Inspired by [22], we perform the fusion of P different VFs on wrench level. Given the linear relationship between wrench and virtual fixtures (3), Gaussian experts result in Gaussian wrenches, i.e. $\boldsymbol{w}_{\text{VF},j} \sim \mathcal{N}(\boldsymbol{\mu}_{\text{VF},j}, \boldsymbol{\Sigma}_{\text{VF},j})$, where $\boldsymbol{\mu}_{\text{VF},j} = \boldsymbol{K}_{\text{VF},j}\text{Log}_{\boldsymbol{x}_{ee}}(\boldsymbol{\mu}_j)$ and $\boldsymbol{\Sigma}_{\text{VF},j} = \boldsymbol{K}_{\text{VF},j}\boldsymbol{\Sigma}_j\boldsymbol{K}_{\text{VF},j}^\top$. Optimal wrenches result from the optimization

$$\hat{\boldsymbol{w}}_{\text{VF}} = \arg\min_{\boldsymbol{w}_{\text{VF}}} \sum_{j=1}^{P} (\boldsymbol{w}_{\text{VF}} - \boldsymbol{\mu}_{\text{VF},j})^\top \boldsymbol{\Sigma}_{\text{VF},j}^{-1} (\boldsymbol{w}_{\text{VF}} - \boldsymbol{\mu}_{\text{VF},j}),$$

$$(15)$$

whose solution is the product of P Gaussians yielding,

$$\hat{\boldsymbol{w}}_{\text{VF}} = \hat{\boldsymbol{\Sigma}}_{\text{VF}} \sum_{j=1}^{P} \boldsymbol{\Sigma}_{\text{VF},j}^{-1} \boldsymbol{w}_{\text{VF},j}, \quad \hat{\boldsymbol{\Sigma}}_{\text{VF}} = \left(\sum_{j=1}^{P} \boldsymbol{\Sigma}_{\text{VF},j}^{-1}\right)^{-1},$$

$$(16)$$

with the Cartesian wrenches $\hat{\boldsymbol{w}}_{\text{VF}}$ being used in (1). Performing the fusion at action level has the advantage of abstracting away the local expert representations, helping to keep the overall formulation generic. For example, in [22] this was used to fuse force- and pose-based policies, which are represented in different spaces and mapped to a common space by the linear structure of the controllers. Since in this work we employ variable stiffness (Section IV-D), in order to keep the influence of the original spaces we set $\boldsymbol{\Sigma}_{\text{VF},j} = \boldsymbol{\Sigma}_j$.

VI. EVALUATION

We evaluate our method on the use case of CubeSat subsystem assembly [3]. We empirically set $l_x = l_y = l_z = 0.06$, $l_{wx} = l_{wy} = l_{wz} = 0.2$, $\gamma = 1 \times 10^{-20}$, $\kappa = 3 \times 10^3$ and $\eta = 1 \times 10^{-6}$ for stiffness scaling in the visual servoing fixture. The position-based trajectory fixture is trained on a dataset (100 trajectories) obtained during a previous user study [3] yielding a $\mathbb{R}^1 \times \mathbb{R}^3 \times \mathcal{S}^3$ GMM visualized in Fig. 1.

A. Prior-knowledge-driven Expert Customization

We customize the visual servoing MoE experts based on human prior knowledge about the task requirements.

a) Zero force along insertion axis: To allow the user to fully control the insertion, we set $K_{\text{VS},3j} = 0, \forall j$, not generating forces along the insertion axis.

b) Adding a dead-zone around connectors: To ensure strong guidance in the victinity of a connector, we add a "dead zone" with radius l_{dead} in x and y coordinates by modifying the first two entries of the vector $\text{Log}_{\boldsymbol{x}_{\text{ee}}}(\boldsymbol{\mu}_m)$

$$x' = x - l_{\text{crop}}\frac{x}{l_{\text{tot}}}, \qquad y' = y - l_{\text{crop}}\frac{y}{l_{\text{tot}}} \qquad (17)$$

with $l_{\text{tot}} = \sqrt{(x)^2 + (y)^2}$ and $l_{\text{crop}} = \min(l_{\text{tot}}, l_{\text{dead}})$. Fig. 5 shows the resulting raw weights, we used $l_{\text{dead}} = 5 \times 10^{-3}m$ in the experiments.

c) Expert initialization: As connectors appear gradually in the camera image, the visual servoing MoE (7) would yield a very certain result once the first detector is perceived only to become much less certain when a second connector is detected. To mitigate this, we initialize the MoE with an additional expert at the current end effector position with high covariance and activation (subscript *pos* for position)

$$h_{M_{\text{VS}}+1}(\boldsymbol{x}_{\text{ee}}, \boldsymbol{x}_{\text{targ}}) = 1 - \exp\left(-\frac{1}{2}\lambda||\boldsymbol{x}_{\text{ee,pos}} - \boldsymbol{x}_{\text{targ,pos}}||^2\right),$$
$$(18)$$

where $\lambda = 1 \times 10^{-2}$ in our experiments. The expected target pose $\boldsymbol{x}_{\text{targ}}$ is supplied externally based on the approximate PCB location. (18) ensures that (7) always has at least one active expert. A dead-zone with $l_{\text{dead}} = 9 \times 10^{-2}m$ is used in the experiments, this time including the z-axis, to ensure that the influence of the initialization expert dissipates in the vicinity of the connectors. Given its large covariance the initialization expert generates negligible guiding forces.

B. Qualitative Evaluation of the Visual-Servoing Fixture

For the first experiment, we only enable the visual servoing fixture which results in $P = 1$ for the PoE in (16). Fig. 6 shows different end effector poses $\boxed{\text{EE}}$ and the resulting estimated Gaussian, according to (9)–(10), given the visualized detections. The obtained results show that the proposed probabilistic fixture gives a strong positional and rotational guidance to the user when close to one target as illustrated by the barely visible purple covariance ellipsoid $\boxed{\text{A}}$ in Fig. 6a. Despite the strong guidance the user is able to 'escape' the fixture and switch to a different connector. Fig. 6b shows that the attractor point is temporarily located between connectors when switching. With our choice of gating function (8), the importance of a
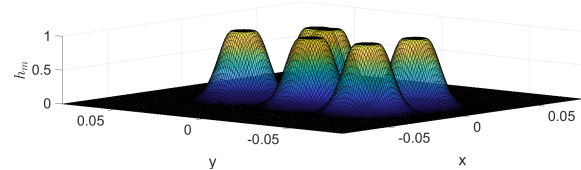

Fig. 5: Distance-based influence factors h_m of the detections in the xy-plane for distances in all other DoFs equal to 0.

connector grows exponentially with decreasing end effector distance, ensuring that the operator is, in the end, always guided towards a connector. Moment matching (10) leads to a large variance of the purple Gaussian in direction of the connectors $\boxed{\text{B}}$. Thanks to the variable stiffness (Section IV-D), this results in lower stiffness along that direction, facilitating the transition.

When far away (e.g. above the backplane PCB, Fig. 6c), the user can not only displace the end effector in the xy plane but also rotate the end effector freely around the z axis. This allows the operator to choose connectors rotated by $180°$.

Fig. 6d shows the effect of including orientation in the distance function. While the closest connector would be at $\boxed{\text{D}}$, our model knows that the most likely target is the connector $\boxed{\text{C}}$ because of a difference of $180°$ in orientation.

C. Pilot Study on CubeSat Subsystem Assembly

Fig. 7 shows the CubeSat assembly task where the subsystem connector has to be mated with the backplane connector requiring very high precision, which was not possible using our telerobotic system without VFs. A positional offset of at most $0.7\,\text{mm}$ as well as a low angular deviation ($4°$ / $2°$) must be achieved, which requires a human in the loop in addition to the fixtures. While the nominal subsystem insertion pose is assumed to be given externally, this information might be inaccurate and the user might want to choose a different connector as CubeSat production is highly individualised.

The task for the human operator is to perform this insertion using camera views and force feedback, consisting of forces from the VFs and the remote environment. Input device and remote side are based on lightweight robots (Fig. 2).

As the application is targeted for expert users, we set up a small pilot study with 15 participants who already have experience with teleoperation,[2] of which 6 already participated in a previous experiment [3]. Participants are first introduced to the system. After completing an introductory questionnaire, they perform test insertion operations with only the novel visual servoing fixture until they are confident with the teleoperation setup and the required precision.

For the actual experiments, we use three different assistive scenarios with different combinations of VFs (Table I). Users perform three trials with each method in one block, the order of the conditions being systematically varied. *Probabilistic Multi-Modal* ($P = 2$) denotes the proposed combination of fixtures (V), *Probabilistic Visual-Servoing* ($P = 1$) only the visual servoing fixture (IV) and *Multi-Modal* the multi-modal

---

[2]All participants are DLR employees with experience in haptic interac- tion with lightweight robots so no special ethic permission is required. Permission for the questionnaires was obtained from the works council and data protection of DLR, participants gave their informed consent.

(a) Close to one connector. (b) In-between two connectors. (c) Far from the board. (d) Orientation influence.
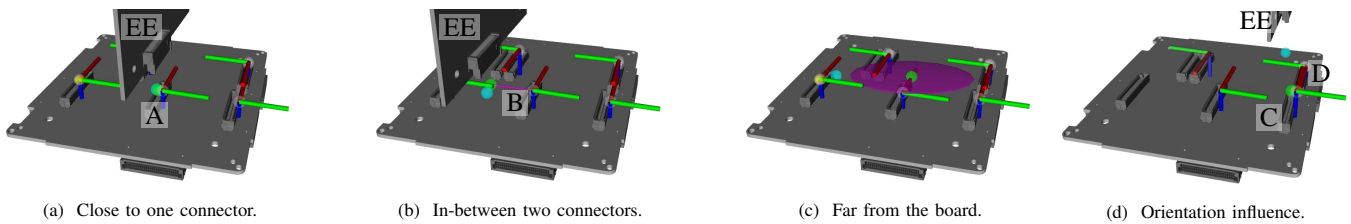
Fig. 6: Probabilistic visual servoing fixture estimation. White spheres represent detected connectors with their Gaussian covariance displayed as small yellow ellipsoid inside them and orientation shown as coordinate frames. These so-called *experts* act as individual candidate fixtures to drive the robot end effector. The purple ellipsoid depicts the 3D Gaussian distribution corresponding to the unimodal approximation of the MoE. Its mean which acts as attractor for the end effector is shown as green sphere. Finally, the blue sphere shows the 3D end effector position $\boxed{\text{EE}}$ projected on the horizontal plane.
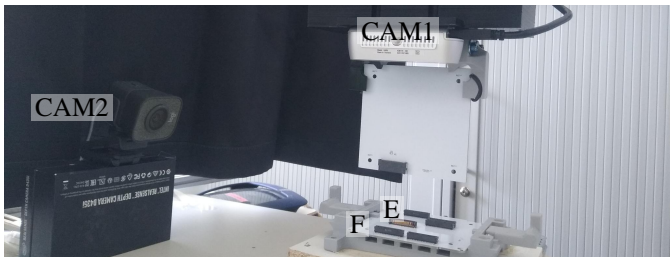

Fig. 7: CubeSat assembly scenario. The gripper holds a Subsystem to be inserted into the backplane (damaged connector at $\boxed{\text{E}}$) mounted on the table. An in-hand camera $\boxed{\text{CAM1}}$ is used for the visual servoing fixture ensuring the high precision required to successfully insert the connector. On the left side, one of the cameras $\boxed{\text{CAM2}}$ provided to the human operator is visible.

TABLE I: FLEXIBILITY, MANIP. TIME, WORKLOAD (LOWER=BETTER), SYSTEM USABILITY SCALE (SUS) SCORE (HIGHER=BETTER), IN BRACKETS: VARIANCE.

Method	Flexibility	Manip. t [s]	Workload	SUS score
Prob. Multi-Mod. (V)	**high**	27.8 (18.1)	5.9 (3.4)	70.2 (15.6)
Prob. Visual-Serv. (IV)	**high**	29.0 (18.8)	6.5 (3.6)	70.2 (15.8)
Multi-Modal [3]	low	**22.9** (15.7)	**4.7** (2.9)	**81** (12.4)

fixture of our previous approach [3]. Unlike our proposed approach, [3] does not provide the flexibility to switch between connectors automatically during task execution. Instead, it requires a manual programming of the vision when the insertion target changes. To simulate a more realistic scenario, where the flexibility of our expert-based approach is required, we ask users to insert into the front left connector $\boxed{\text{F}}$ (Fig. 7), while the position-based trajectory fixture guides them near a damaged middle connector $\boxed{\text{E}}$. This requires switching between target connectors near the PCB online, which is not possible in [3]. Our aim is to show that our proposed approach performs favorably when compared to the hand-coded approach in [3] despite the added flexibility. As such, in *Multi-Modal* the user is directly guided to the front left connector $\boxed{\text{F}}$, resulting in a very favorable baseline.

Subjects report their workload using the NASA TLX questionnaire [36] after each trial and the usability using the SUS [37] after each block. *Manipulation Time* (10 cm above the PCB until successful insertion) and subjective results are summarized in Table I. Results of a repeated-measures ANOVA on *Manip. Time* and *workload* are shown in Table II. With 15 participants and a partial $\eta^2 = 0.0714$ we achieve a sufficient statistical power of 0.91 for the *Manip. Time* analysis. For analyzing the workload, values for the within factor *Fixture* have been Greenhouse-Geissner corrected. Post-hoc comparisons with Bonferroni adjustment for the effect of the within factor *Fixture* on the average TLX score revealed a significant difference ($p < .05$) between *Prob. Visual-Serv.* and *Multi-Modal*. The SUS scores were not normally distributed

TABLE II: ANOVA results comparing the different fixtures.

Dependent Variable	Factor	df1	df2	F	p
Manipulation Time	Fixture	2	28	1.0764	n.s.
	Run	2	28	1.5299	n.s.
	Fixt. * Run	8	56	0.7131	n.s.
Average TLX	Fixture	1.39	19.46	4.1347	$< .05$
	Run	2	28	9.8455	$< .001$
	Fixt. * Run	8	56	1.1958	n.s.

and thus the Friedman test was performed ($\chi^2 = 10.29$, $p < .05$). Post-hoc comparisons with the Wilcoxon test indicated that scores were significantly higher for *Multi-Modal* compared to the other conditions.

D. Discussion of the Pilot Study Results

As expected, the lower **workload** of *Multi-Modal* when compared to *Probabilistic Visual-Servoing* reflects the difference in available guidance between both fixtures, since the former does not provide guidance towards the PCB. However, no significant difference between *Probabilistic Multi-Modal*, where a position-based expert is used, and *Multi-Modal* could be found even though users switched connectors during runtime with the former. This is contrasted by the **SUS score**, where *Multi-Modal* is significantly separated from the two other methods. All methods still achieve a mean > 68 which is generally considered to be above average.

While **manipulation times** can sometimes be vastly different between different trials because of tight tolerances, the pilot study did not show significant differences between the fixtures. This suggests that fine guidance close to the target - which usually takes most of the time - is, as expected, very similar. This can also be underlined by examining the PoE result close to the target. The used position-based trajectory fixture deviates by 3 cm from the target which is precisely detected by the visual servoing fixture (< 1 mm) and has a four orders of magnitude larger covariance. Thanks to the probabilistic weighting of both fixtures, the force applied by the incorrect trajectory fixture is only 0.03 N not hindering precise telemanipulation. We thus conclude that the added flexibility of our probabilistic approach maintains the precise guidance of [3] even under unfavorable conditions.

E. Limitations of the Approach

Selecting hyperparameters in (7)–(8) currently requires expert tuning which should be automated in future work. For a heavily inclined camera pose, the rectangle extraction in IV-A might fail which we however did not observe yet. More powerful detection methods can help to overcome this limitation and also allow to interact with more difficult to perceive objects.

VII. CONCLUSION

We proposed an approach based on a *mixture of experts* model to automatically detect and arbitrate visual servoing fixtures in shared control. Our approach allows to incorporate new or disappearing targets by dynamically creating and removing fixtures. To benefit from a multi-phase guidance throughout the robot's workspace, a position-based trajectory fixture is fused using a *product of experts* approach. Our results show that with our method we could obtain a natural arbitration of multiple fixtures comparable in performance to a hand-tuned arbitration function [3] while offering much more flexibility. This was achieved by extracting a meaningful covariance which is then used to modulate the end effector stiffness, allowing to switch targets. The position-based fixture furthermore provides guidance when far from the target. The experimental evaluation shows that the method supports the insertion of CubeSat subsystems into multiple target connectors, providing strong and useful guidance as well as giving the user the choice of different possible targets.

In future work, we plan to extend our method to other geometries and applications as well as to investigate approaches for seamless switching from teleoperation to automation.

ACKNOWLEDGMENT

REFERENCES

[1] L. Rosenberg, "Virtual fixtures: Perceptual tools for telerobotic manipulation," in *Proceedings of IEEE Virtual Reality Annual International Symposium*, 1993, pp. 76–82.

[2] G. Raiola, S. Sanchez Restrepo, P. Chevalier, P. Rodriguez-Ayerbe, X. Lamy, S. Tliba, and F. Stulp, "Co-manipulation with a Library of Virtual Guiding Fixtures," *Autonomous Robots*, pp. 1573–7527, 2018.

[3] M. Mühlbauer, F. Steinmetz, F. Stulp, T. Hulin, and A. Albu-Schäffer, "Multi-phase multi-modal haptic teleoperation," in *Proc. IEEE/RSJ Intl Conf. on Intelligent Robots and Systems (IROS)*. IEEE, 2022.

[4] C. J. Pérez-del Pulgar, J. Smisek, V. F. Muñoz, and A. Schiele, "Using learning from demonstration to generate real-time guidance for haptic shared control," in *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2016, pp. 003 205–003 210.

[5] R. Jacobs, M. Jordan, S. Nowlan, and G. Hinton, "Adaptive mixtures of local experts," *Neural Computation*, vol. 3, no. 1, pp. 79–87, 1991.

[6] I. Havoutis and S. Calinon, "Learning assistive teleoperation behaviors from demonstration," in *IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*, 2016, pp. 258–263.

[7] M. Selvaggio, G. Notomista, F. Chen, B. Gao, F. Trapani, and D. Caldwell, "Enhancing bilateral teleoperation using camera-based online virtual fixtures generation," in *Proc. IEEE/RSJ Intl Conf. on Intelligent Robots and Systems (IROS)*, 2016, pp. 1483–1488.

[8] V. Pruks and J.-H. Ryu, "Method for generating real-time interactive virtual fixture for shared teleoperation in unknown environments," *The Int. Journal of Robot. Research*, vol. 41, no. 9-10, pp. 925–951, 2022.

[9] G. D. Hager, "Human-machine cooperative manipulation with vision-based motion constraints," in *Visual Servoing via Adv. Num. Methods*, G. Chesi and K. Hashimoto, Eds. London: Springer, 2010, pp. 55–70.

[10] A. Bettini, P. Marayong, S. Lang, A. M. Okamura, and G. D. Hager, "Vision-assisted control for manipulation using virtual fixtures," *IEEE Trans. on Robotics*, vol. 20, no. 6, pp. 953–966, 2004.

[11] L. Wu, K. Wu, and H. Ren, "Towards hybrid control of a flexible curvilinear surgical robot with visual/haptic guidance," in *Proc. IEEE/RSJ Intl Conf. on Intelligent Robots and Systems (IROS)*, 2016, pp. 501–507.

[12] D. Aarno, S. Ekvall, and D. Kragic, "Adaptive virtual fixtures for machine-assisted teleoperation tasks," in *Proc. IEEE Intl Conf. on Robotics and Automation (ICRA)*, 2005, pp. 1139–1144.

[13] S. Calinon, "A tutorial on task-parameterized movement learning and retrieval," *Intelligent service robotics*, vol. 9, no. 1, pp. 1–29, 2016.

[14] K. Hagmann, A. Hellings-Kuß, J. Klodmann, R. Richter, F. Stulp, and D. Leidner, "A digital twin approach for contextual assistance for surgeons during surgical robotics training," *Frontiers in Robotics and AI*, vol. 8, 2021.

[15] F. Abi-Farraj, C. Pacchierotti, O. Arenz, G. Neumann, and P. R. Giordano, "A haptic shared-control architecture for guided multi-target robotic grasping," *IEEE Transactions on Haptics*, vol. 13, no. 2, pp. 270–285, 2020.

[16] H. Saeidi and Y. Wang, "Incorporating trust and self-confidence analysis in the guidance and control of (semi)autonomous mobile robotic systems," *IEEE Robotics and Automation Letters (RA-L)*, vol. 4, no. 2, pp. 239–246, 2019.

[17] R. Balachandran, H. Mishra, M. Cappelli, B. Weber, C. Secchi, C. Ott, and A. Albu-Schaeffer, "Adaptive authority allocation in shared control of robots using bayesian filters," in *Proc. IEEE Intl Conf. on Robotics and Automation (ICRA)*, 2020, pp. 11 298–11 304.

[18] M. J. A. Zeestraten, I. Havoutis, and S. Calinon, "Programming by demonstration for shared control with an application in teleoperation," *IEEE Robotics and Automation Letters (RA-L)*, vol. 3, no. 3, pp. 1848–1855, 2018.

[19] Y. Michel, R. Rahal, C. Pacchierotti, P. R. Giordano, and D. Lee, "Bilateral teleoperation with adaptive impedance control for contact tasks," *IEEE Robotics and Automation Letters (RA-L)*, vol. 6, no. 3, pp. 5429–5436, 2021.

[20] C. Yang, K. Yuan, Q. Zhu, W. Yu, and Z. Li, "Multi-expert learning of adaptive legged locomotion," *Sci. Robot.*, vol. 5, no. 49, 2020.

[21] M. J. A. Zeestraten, I. Havoutis, J. Silvério, S. Calinon, and D. G. Caldwell, "An approach for imitation learning on Riemannian manifolds," *IEEE Robotics and Automation Letters (RA-L)*, vol. 2, no. 3, pp. 1240–1247, 2017.

[22] J. Silvério, Y. Huang, L. Rozo, S. Calinon, and D. G. Caldwell, "Probabilistic learning of torque controllers from kinematic and force constraints," in *Proc. IEEE/RSJ Intl Conf. on Intelligent Robots and Systems (IROS)*, 2018, pp. 1–8.

[23] E. Pignat and S. Calinon, "Bayesian gaussian mixture model for robotic policy imitation," *IEEE Robotics and Automation Letters (RA-L)*, vol. 4, no. 4, pp. 4452–4458, 2019.

[24] E. Pignat, J. Silvério, and S. Calinon, "Learning from demonstration using products of experts: Applications to manipulation and task prioritization," *The Int. Journal of Robot. Research*, vol. 41, no. 2, pp. 163–188, 2022.

[25] O. Khatib, "A unified approach for motion and force control of robot manipulators: The operational space formulation," *IEEE Journal on Robotics and Automation*, vol. 3, no. 1, pp. 43–53, 1987.

[26] A. Albu-Schaffer, C. Ott, U. Frese, and G. Hirzinger, "Cartesian impedance control of redundant robots: recent results with the dlr-light-weight-arms," in *Proc. IEEE/RSJ Intl Conf. on Intelligent Robots and Systems (IROS)*, vol. 3, 2003, pp. 3704–3709 vol.3.

[27] J. Q. Gallier and J. Quaintance, *Differential geometry and lie groups*. Springer, 2020, vol. 12.

[28] X. Pennec and M. Lorenzi, "Beyond Riemannian geometry: The affine connection setting for transformation groups," in *Riemannian Geometric Statistics in Medical Image Analysis*, X. Pennec, S. Sommer, and T. Fletcher, Eds. Academic Press, 2020, pp. 169–229.

[29] S. Calinon, "Gaussians on Riemannian manifolds: Applications for robot learning and adaptive control," *IEEE Robotics & Automation Magazine*, vol. 27, no. 2, pp. 33–45, 2020.

[30] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer, 2006.

[31] T. D. Barfoot and P. T. Furgale, "Associating uncertainty with three-dimensional poses for use in estimation problems," *IEEE Trans. on Robotics*, vol. 30, no. 3, pp. 679–693, 2014.

[32] G. Bradski, "The OpenCV Library," *Dr. Dobb's Journal of Software Tools*, 2000.

[33] C. Gillmann, P. Arbelaez, J. T. Hernandez, H. Hagen, and T. Wischgoll, "An uncertainty-aware visual system for image pre-processing," *Journal of Imaging*, vol. 4, no. 9, 2018.

[34] C. Dalitz and H. Niederrhein, "Soft thresholding for visual image enhancement," Hochschule Niederrhein, Fachbereich Elektrotechnik und Informatik, Tech. Rep., 2014.

[35] C. E. Rasmussen and C. K. I. Williams, *Gaussian processes for machine learning*. MIT Press, 2006.

[36] S. G. Hart and L. E. Staveland, "Development of nasa-tlx (task load index): Results of empirical and theoretical research," in *Advances in psychology*. Elsevier, 1988, vol. 52, pp. 139–183.

[37] J. Brooke, "SUS: a "quick and dirty" usability scale," *Usability evaluation in industry*, vol. 189, no. 3, 1996.