2026 IEEE International Conference on Robotics and Automation (ICRA 2026)
June 1-5, 2026. Vienna, Austria

Gaze-based Teleoperation with Intent Inference Model for Robotic Manipulators*

Yanjia Yuan[1], Chong Peng[1], Dihui Chu[1,2], Qianqian Wang[1], Qiang Gao[1], Yunlong Tang[2], and Xiaoyu Wang[1*]

***Abstract*— Eye gaze-based control interfaces provide a non-invasive means of enhancing human-robot collaboration for activities of daily living and can reduce the cognitive burden on operators performing complex tasks. Eye gaze has traditionally been used for "gaze triggering," where fixating on an object activates pre-programmed robotic movements. In this work, we propose a gaze-based robotic teleoperation approach that utilizes real-time gaze data to guide the freeform movement of robotic manipulators. The proposed approach incorporates a Gaussian Mixture Regression (GMR)-based intent inference model to capture the nonlinear relationship between gaze data and the operator's intended robotic movements. For benchmarking, we further implemented a Gaussian Hidden Markov Model (G-HMM) to provide a comparable probabilistic framework for intent inference. Experimental results demonstrate that the GMR-based approach achieves a statistically significant improvement over G-HMM in terms of control efficiency, trajectory smoothness against involuntary eye fluctuations, as well as enhancing the user's sense of involvement and control.**

I. Introduction

Patients with upper limb impairments, including those affected by spinal cord injuries, strokes, multiple sclerosis, etc., often face significant challenges in independently controlling their upper limbs for daily activities. To improve the independence and quality of life for these individuals, the use of collaborative manipulator as a substitute for human hands in completing everyday tasks has emerged as an effective assistive solution [1]. However, traditional methods of controlling robotic arms require operators to frequently switch between several modes for commanding gripper position, orientation, and open/close using an unintuitive 3D Cartesian space perspective. This decreases task execution efficiency and imposes considerable cognitive burden on the operator [2].

In recent years, human-robot shared control has gradually become a prominent research focus [3]. This approach enhances task efficiency and precision by integrating the operator's intent with the robot's autonomous control [4], while maintaining the operator's sense of involvement [5].

*Research supported by the National Natural Science Foundation of China under Grant 6502008061.

This work involved human subjects in its research. Approval of all ethical and experimental procedures and protocols was granted by the Institutional Review Board of Southeast University, Zhongda Hospital, under Protocol No. 2025ZDSYLL377.

[1]Department of Intelligent Manufacturing, School of Mechanical Engineering, Southeast University, Nanjing, China.

[2]Department of Materials Science and Engineering, Department of Mechanical and Aerospace Engineering, Monash University, Melbourne, Australia.

Fig. 1. Experimental setup of the proposed gaze-based teleoperation approach. This approach utilizes gaze data to guide the freeform movement of the robotic manipulator. A Gaussian Mixture Regression (GMR) model is employed to infer the operator's intended robotic movement and mitigate motion fluctuations caused by involuntary eye movements.

A variety of human input interfaces have been leveraged for shared control such as speech [6], gesture [7], electromyography (EMG) [8], electroencephalography (EEG) [9], electrocorticography (ECoG) and gaze etc [10], [11]. Among these, gaze control offers unique advantages for being non-invasive, non-verbal, intuitive, and easy to don and doff [12].

Current gaze-based human-robot shared control systems typically employ fixation duration thresholds to initiate predefined actions [13]–[15]. As demonstrated by Wang et al., accurate 3D gaze estimation can be achieved using feature-based reconstruction methods with temporal activation thresholds [13], whereas Shafti et al. developed an alternative approach using state-machine control triggered by 1.5-second gaze fixations for context-aware grasping [14]. Building on these temporal triggering mechanisms, Zeng et al. incorporated multimodal brain-machine interfaces within conventional gaze windows [15], introducing hybrid signal fusion that modestly improves responsiveness while retaining the fundamental characteristics of gaze-based control.

While threshold-based methods simplify intent recognition, they are typically constrained by pre-programmed robotic movements and therefore lack adaptability in unstructured environments. A different line of work has sought to infer the operator's intent more directly from gaze behavior

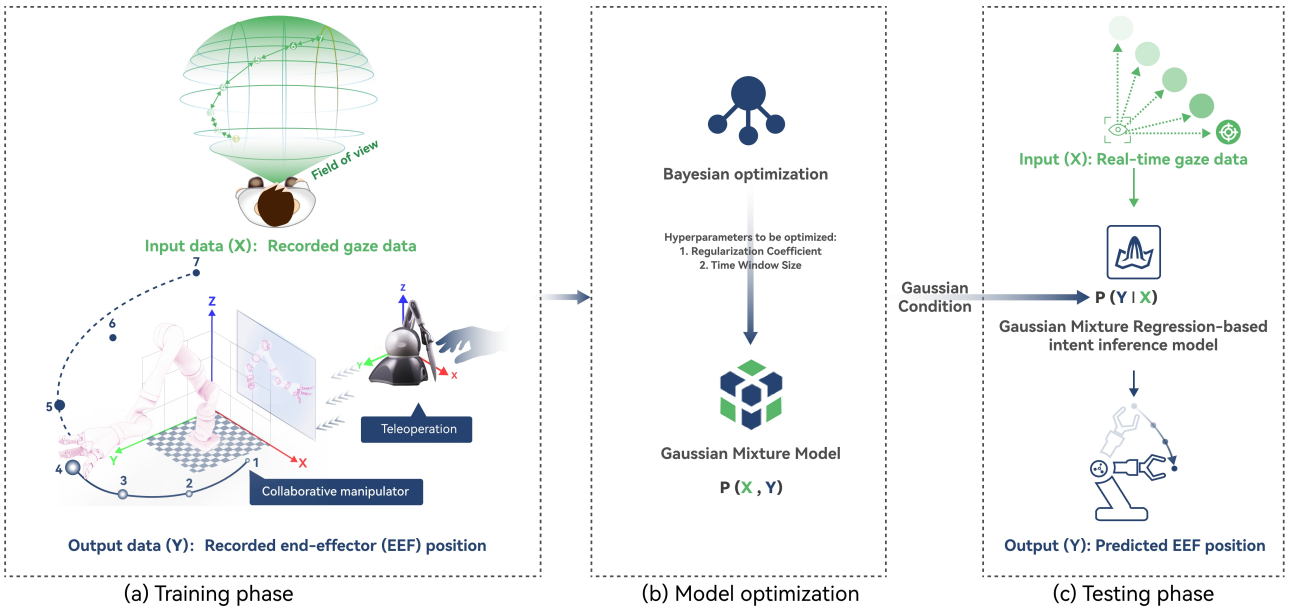

Fig. 2. Framework of the proposed gaze-based teleoperation approach. (a) Training phase: The operator teleoperates the robotic manipulator using a haptic force-feedback device, while paired gaze data and end-effector positions are recorded to train the GMR model. (b) Model optimization: The hyperparameters of the GMR model, including the regularization coefficient and time window size, are optimized to enhance training accuracy. (c) Testing phase: Real-time gaze data is input into the trained GMR model to infer the operator's intended robotic movement.

data and task context. Robot manipulators then perform tasks with autonomous path planning according to the recognized intent. Li and Zhang used Bayesian graphical models to associate gaze behavior with target objects, allowing flexible task execution in dynamic environments [16]. Subsequent methods further integrated gaze trajectories with contextual cues. Haji Fathaliyan et al. used classifiers to recognize ongoing operations and predict subsequent actions, improving performance in bimanual tasks [17]. Wang et al. proposed an RNN-based model for real-time action unit recognition using gaze and action sequences [18], enhancing adaptability in fast-changing scenarios. Despite these advances, generalization to unstructured environments remains a key challenge.

Even in approaches that infer intent for autonomous execution, robotic path planning still relies on prior environmental information, such as the identities and positions of target objects and obstacles. This information can either be preprogrammed or obtained through computer vision methods, both of which may not be readily available in unstructured environments or more generalized situations. Moreover, because these approaches depend strongly on prespecified or automatically extracted knowledge, they tend to reduce the user's sense of involvement and control, making the interaction feel less natural and less adaptive to dynamic task demands.

To address these limitations, we propose a gaze-based robotic teleoperation method with intent inference model. This model is based on Gaussian Mixture Regression (GMR) [19], [20] and learns the relationship between gaze patterns and the position of the robotic end-effector, as shown in Fig. 1. Using this method, operators can accurately guide the movement of a robotic manipulator through eye gaze alone, even in the absence of prior environmental information.

II. Method

In our preliminary implementation, a gaze-based teleoperation system was constructed in which real-time gaze coordinates were directly transmitted to the robotic end-effector, enabling a direct gaze control. However, the system exhibited significant performance limitations due to inherent physiological characteristics of human oculomotor control. Specifically, involuntary saccadic eye movements and unintended fixations caused continuous gaze point fluctuations between target objects, end-effector, and intermediate waypoints, resulting in pronounced oscillatory motion of the end-effector, substantially degrading both operational efficiency and motion smoothness.

Previous studies have explored probabilistic mappings from gaze behavior to intended robot motions, incorporating temporal filtering to improve motion smoothness [21]. The model employs temporal filtering to attenuate gaze signal noise while preserving intentional control commands, enhancing both the smoothness and efficiency of the robot's motion.

A. Gaze-based Teleoperation Framework

GMR provides an effective tool for modeling nonlinear relationships by approximating complex probability distributions from sparse datasets. Building on its demonstrated capability to learn robust mappings between gaze coordinates and manipulator target positions [19], [20], this study implements GMR as the core intent inference model for gaze-to-motion translation. The proposed system operates through sequential phases beginning with model training, where

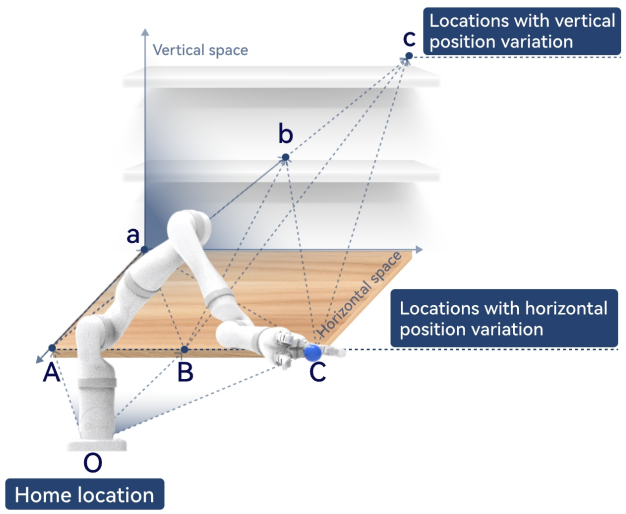

Fig. 3. Experimental procedure for training dataset collection. In Stage 1, the experimenter teleoperates the robotic manipulator using a force-feedback teleoperation device, moving it from the home position (O) to one of three locations with horizontal variation (A, B, C). In Stage 2, the manipulator is then teleoperated to one of three locations with vertical variation (a, b, c).

GMR learns the gaze-manipulator position relationship from demonstration data. This is followed by model optimization through hyperparameter tuning to minimize prediction error while preventing overfitting. The final phase involves online testing, where real-time inference converts gaze signals into smooth robotic motions. The complete pipeline, including training, optimization, and testing phases, is illustrated in Fig. 2.

The intent inference model was trained using a teleoperation platform integrating a haptic force-feedback device with a collaborative robotic manipulator. This configuration enabled natural and intuitive control of the end-effector while simultaneously recording synchronized gaze and manipulator motion data for model training.

The training process employed a probabilistic framework beginning with Gaussian Mixture Model (GMM) fitting to capture the joint distribution between gaze patterns and manipulator trajectories. Key hyperparameters, including time window size and regularization coefficients, were optimized through Bayesian Optimization to maximize model accuracy while preventing overfitting [22] .

For real-time operation, GMR derived conditional probability distributions from the learned joint model, enabling continuous prediction of the manipulator's next target position based on streaming gaze data. The predicted target positions were streamed to the robotic controller, enabling smooth real-time trajectory execution.

B. Dataset Collection for Intent Inference Model

This study developed a hybrid teleoperation platform comprising: (1) a ROS-based system integrating Kinova collaborative manipulator control with Tobii Pro Glasses 3 eye-tracking, and (2) a ROS 2 interface for haptic force-feedback device data acquisition. The subsystems communicate via TCP protocol, with ROS handling robotic motion execution and gaze data processing while ROS 2 manages 6-DoF teleoperation data (position and orientation) from the haptic force-feedback device. All data streams - including gaze coordinates, end-effector states, and haptic inputs - are temporally synchronized during operation.

The teleoperation system acquires the stylus tip pose data from the haptic force-feedback device, including position coordinates and Euler angles, converting pose information from the haptic force-feedback device coordinate system to the manipulator's workspace coordinate system, as illustrated in Fig. 2 (a). These transformed pose commands are sent to the manipulator control system, enabling natural and intuitive control of the robotic manipulator through direct stylus manipulation.

Regarding the coordinate transformation, the three-dimensional coordinates of the haptic force feedback device are linearly assigned to the end effector of the manipulator using proportional coefficients k. The coefficient scale the input to align with the manipulator's workspace. As shown below:

$$R(t) = R(t_0) + k(H(t) - H(t_0)) \tag{1}$$

where $R(t) = (R_x(t), R_y(t), R_z(t))^T$ denotes the end-effector position in the coordinate system of the manipulator, $H(t) = (H_x(t), H_y(t), H_z(t))^T$ denotes the stylus tip position in the coordinate system of the haptic force-feedback device, and $R(t_0)$ and $H(t_0)$ represent their respective home positions.

Similarly, Euler angles are mapped using a direct one-to-one ratio, where the orientation of the control stylus tip from the haptic force-feedback device is directly transferred to the manipulator's end-effector without any scaling.

Five experimenters participated in the data collection process. Each experimenter used the haptic force-feedback device to teleoperate the robotic manipulator, guiding it through a series of predefined target locations, as shown in Fig. 3. The teleoperation process consisted of two stages. In stage 1 (S1), the manipulator moved from the home location to various locations with horizontal position variations. In stage 2 (S1), it transitioned from these horizontally varied positions to locations with vertical position variations.

Each experimenter repeated the procedure three times, with nine trials in each repetition, resulting in a total of 27 trials per experimenter, capturing gaze data and manipulator motion data across both planar and elevation movements. The predefined target locations were determined to ensure adequate coverage of the manipulator's reachable workspace, thereby capturing representative motions within its operational limits. All data were recorded synchronously at 50 Hz during both experimental stages, ensuring temporal alignment for model training.

C. Intent Inference Model Fitting

GMR-based Model In the data set construction process, based on the experimental task scenarios, we collected the gaze point sequences of the experimental subjects $g_t =$

$(g_{x_t}, g_{y_t}, g_{z_t})$ and the end-effector position sequences of the manipulator $\rho_t = (\rho_{x_t}, \rho_{y_t}, \rho_{z_t})$. A history window of size k was set to construct the observation feature vector [23]: $\Theta_t = [g_{t-k}, g_{t-k+1}, \ldots, g_t, \rho_{t-k}, \rho_{t-k+1}, \ldots, \rho_{t-1}]$.

For model construction, the observation variable $X_t = \Theta_t$ represents the historical gaze and position sequences, while the state variable $Y_t = \rho_t$ denotes the current end-effector position to be predicted. The collected data, with a total size of N, were fitted using a GMM to learn the mapping between observation variable and state variable [24].

Suppose the joint distribution of the observation variable and the state variable is a Gaussian distribution [25], denoted as:

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \sum_{k=1}^{K} \pi_k \mathcal{N}\left(\begin{pmatrix} \mu_X^k \\ \mu_Y^k \end{pmatrix}, \begin{pmatrix} \Sigma_{XX}^k & \Sigma_{XY}^k \\ \Sigma_{YX}^k & \Sigma_{YY}^k \end{pmatrix} \right) \tag{2}$$

where π_k is the weight of the k-th Gaussian component, μ_X^k and μ_Y^k are the mean vectors of the observation variable and the state variable of the k-th Gaussian component respectively, Σ_{XX}^k and Σ_{YY}^k are the covariance matrices of the observation variable and the state variable of the k-th Gaussian component respectively, $\Sigma_{YX}^k$and Σ_{XY}^k is the cross-covariance matrix of the two variables.

In the fitting process, the Expectation-Maximization (EM) algorithm is used for parameter estimation [26] . The EM algorithm iteratively updates the parameters until the log-likelihood function converges to a stable value.

After training the GMM, GMR is applied to compute the conditional distribution. This step refines the learned probabilistic relationship, allowing the model to infer the most likely state given the gaze point data.

For each Gaussian component, the conditional mean of the state data $Y|X$ is:

$$\mu_{Y|X}^k = \mu_Y^k + \Sigma_{YX}^k (\Sigma_{XX}^k)^{-1} (X - \mu_X^k) \tag{3}$$

The conditional covariance is:

$$\Sigma_{Y|X}^k = \Sigma_{YY}^k - \Sigma_{YX}^k (\Sigma_{XX}^k)^{-1} \Sigma_{XY}^k \tag{4}$$

Finally, the conditional distribution can be obtained, which is the weighted sum of each Gaussian component:

$$P(Y|X) = \sum_{k=1}^{K} \gamma_k \mathcal{N}(Y; \mu_{Y|X}^k, \Sigma_{Y|X}^k) \tag{5}$$

where γ_k is the posterior probability of the k-th Gaussian component:

$$\gamma_k = \frac{\pi_k \mathcal{N}(X; \mu_X^k, \Sigma_{XX}^k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(X; \mu_X^j, \Sigma_{XX}^j)} \tag{6}$$

After model training completion, real-time prediction is executed by computing the conditional probability distribution $P(\rho_t|\Theta_t)$ for any input observation feature vector. As illustrated in Fig. 4, the temporal dependencies in the system create a predictive relationship between consecutive time steps. Specifically, the end-effector position ρ_t and observation feature Θ_t at time t propagate forward as historical state

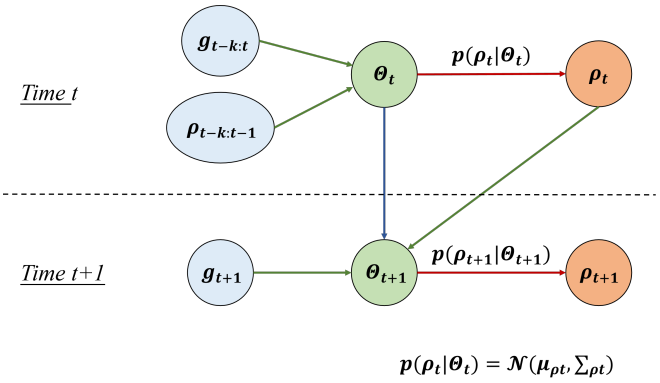

Fig. 4. Time-series prediction at time t and $t+1$. The blue arrows indicate time updates, the green arrows represent information fusion, and the red arrows denote the calculation of the predicted probability of the manipulator end-effector position based on the observed feature vector.

information, directly influencing the model's probabilistic prediction of ρ_{t+1} at time $t+1$. This explicit memory effect ensures smooth trajectory generation by preserving continuity across consecutive states. The distribution follows a Gaussian form, from which the predicted end-effector position is derived as its expectation:

$$\hat{\rho}_t = E[\rho_t|\Theta_t] \tag{7}$$

This expectation serves as the motion command for real-time control.

Hyperparameter Optimization The selection of hyperparameters was optimized to improve model performance, with key parameters derived from the trained GMM. Three key parameters influence the accuracy of the prediction: the number of Gaussian components (*numComponents*), the time window size and regularization coefficient.

The number of Gaussian components directly affects model precision and computational cost, where excessive components may reduce the real-time prediction frequency below 50 Hz under limited computational resources. The time window size determines the temporal context considered by the model, affecting its sensitivity to dynamic changes, while the regularization coefficient controls model complexity and helps prevent overfitting. All parameters were fine-tuned within a reasonable range to ensure reliable performance in real-time scenarios.

Dynamic Time Warping (DTW) is used to measure the similarity between two time series [27]. By performing local alignment, DTW minimizes distance variations, ensuring a more accurate assessment of overall trajectory trends while reducing the impact of local discrepancies. In this study, DTW evaluates the difference between the predicted trajectory generated by GMR and the recorded trajectory during the training phase.

Based on selecting DTW as the optimization function, the objective function is defined as:

$$Obj = DTW(\hat{S}, S) \tag{8}$$

Where $\hat{S}$ is the predicted trajectory, S is the recorded trajectory.

This study employs Bayesian optimization for hyperparameter tuning. Compared to traditional grid or random search methods, Bayesian optimization finds the global optimum with fewer iterations, making it suitable for the selection of the optimization algorithm in this study. The optimization objective function is:

$$Obj(k, \text{Regularize}) \rightarrow Obj \tag{9}$$

Where k represents the time window size , and Regularize represents the regularization coefficient. The optimization results are presented in Section IV-A.

III. Experiment and Evaluation

This section compares the performance between two gaze-based robotic teleoperation systems, each incorporating a different intent inference model: GMR and a Gaussian Hidden Markov Model (G-HMM). The G-HMM is chosen as a baseline due to its probabilistic formulation, allowing for a principled comparison between uncertainty-aware intent inference models. Three objective performance measures are defined to evaluate the smoothness and efficiency of the gaze-based teleoperation process: completion time, trajectory deviation, and noise-induced oscillation count. In addition, three subjective performance measures are included, focusing on the user's sense of involvement and control, and perceived temporal efficiency.

A. Experimental Protocol

The experiment compares two intent inference models: one is GMR-based intent inference model and the other is G-HMM-based model [28]. In both cases, the model processes incoming gaze point signals to generate the control commands for the robotic manipulator.

During the experiment, we recruited 12 participants (8 male, 4 female; aged 18–35 years) to perform gaze-based robotic teleoperation trials under two models. None of the participants were affiliated with the laboratory staff responsible for haptic teleoperation data collection, and three of them reported prior experience in robotic interaction. This study involved human subjects and was approved by the Institutional Review Board of Southeast University, Zhongda Hospital (Protocol No. 2025ZDSYLL377). All participants provided written informed consent in accordance with the Declaration of Helsinki.

Following the same two-stage protocol established in Section II-B. Each trial consisted of two coherent stages: S1 involved guiding the manipulator from the home position (O) to one of three horizontal target locations (A, B, C), while S2 required moving from these horizontal positions to one of three vertical target locations (a, b, c). The spatial coordinates of all target locations matched those defined in Fig. 3 for haptic teleoperation. This experimental design resulted in nine unique condition paths (O-A-a through O-C-c), yielding a total of 9 trials per participant.

To guarantee manipulator stability and prevent inverse kinematic singularities, we fixed the end-effector's Euler angles to a pre-programmed configuration optimized for both task requirements and environmental constraints.

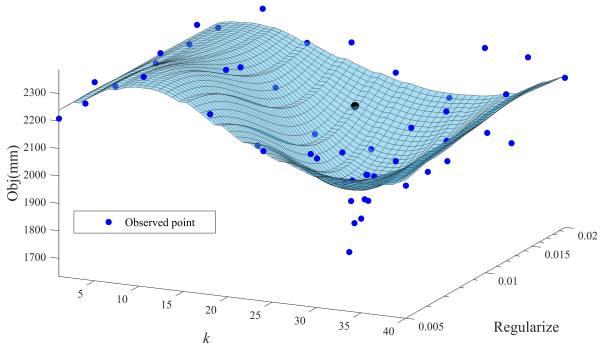

Fig. 5. Hyperparameter optimization results for the GMR model. Bayesian optimization is employed to determine the optimal hyperparameters for the GMR model, including the regularization coefficient and time window size. The DTW distance between the recorded and predicted end-effector trajectories is used as the optimization metric and objective function.

B. Evaluation Metrics

*1)Objective Measures:*To evaluate the smoothness and efficiency of the gaze-based teleoperation process, our objective performance measures focus on the following three key factors: completion time, trajectory deviation, noise-induced oscillation count.

Completion Time (CT) quantifies the duration required for the manipulator to execute both experimental stages, serving as a direct measure of teleoperation efficiency.

Trajectory Deviation (TD) quantifies the deviation between gaze-based control trajectories and the corresponding ground-truth teleoperation trajectories of the same participant using Dynamic Time Warping (DTW). For each participant in the final evaluation, nine trials were conducted, resulting in nine gaze-based trajectories and nine teleoperated reference trajectories. Each gaze-based trajectory is compared against its corresponding teleoperation trajectory, and the resulting DTW distance is used as the deviation measure.

Noise-Induced Oscillation Count (NIOC) quantifies the impact of saccadic eye movements and unintended fixations on the manipulator's trajectory by counting the number of local maxima in the movement path. Saccadic eye movements and unintended fixations cause rapid directional changes, creating peaks in the motion path.

*2)Subjective Measures:*Following each trial, participants completed a subjective evaluation based on a questionnaire [29]. Responses were collected using a 7-point Likert scale, with 1 indicating "strongly disagree" and 7 indicating "strongly agree." Participants were asked to rate their agreement with the following statements:

1) "I felt in control."
2) "The robot did what I wanted."
3) "I was able to accomplish the task quickly."

After all trials, participants answered two open-ended questions:

1) "Which intent Inference model do you prefer and why?"
2) "Do you have any general comments for the two

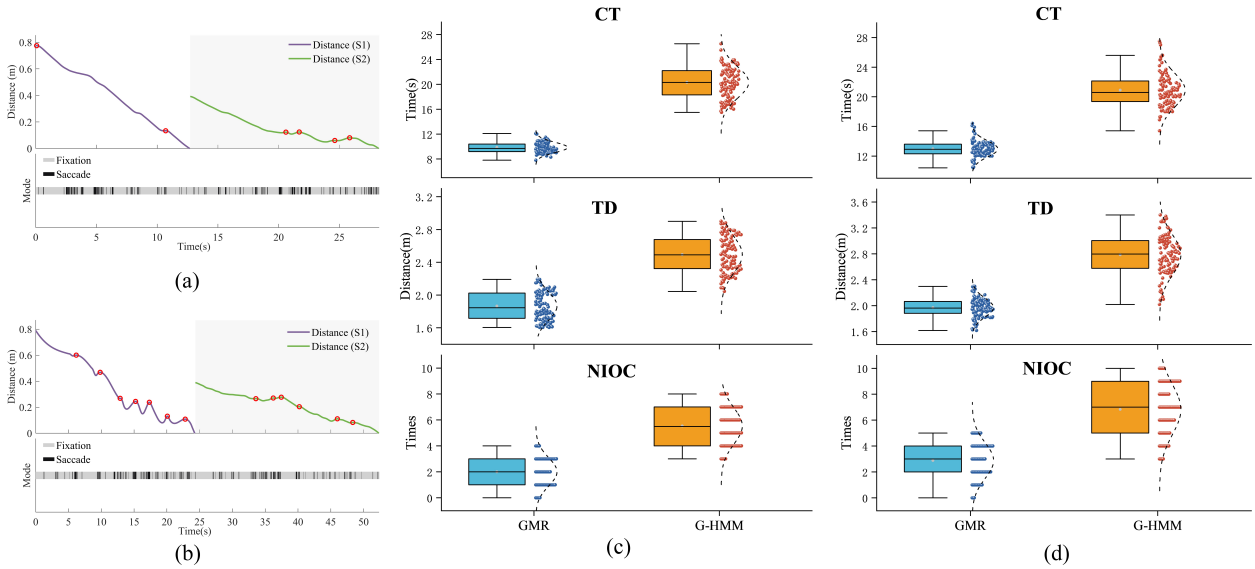

Fig. 6. Visualization of performance comparison between the two models. (a)(b):Representative trial comparison under identical task conditions with the same operator, showing the temporal evolution of end-effector-to-target distance and reflecting control smoothness. (c)(d):Aggregated performance in the S1 and S2 stages, respectively, visualized using box plots of three objective performance measures, providing a global view of robustness and consistency across all participants.

models?"

IV. Results and Analysis

A. Results of GMR-based Modeling

This study employed Bayesian optimization to tune the GMM hyperparameters, minimizing the DTW distance between actual and predicted motion trajectories [30]. As shown in Fig. 5, the algorithm executed 60 optimization iterations over the search space defined by the history window size (k, x-axis) and the regularization coefficient (z-axis), with the objective function value (y-axis) as the optimization target. The optimization converged to an optimal configuration with a window size of 26 and a regularization coefficient of 0.0023716. The surface distribution further indicated that model performance was more sensitive to variations in the history window size than to changes in the regularization term, highlighting the role of hyperparameter selection in ensuring robust trajectory prediction.

B. Objective Measures of Performance

Following the experimental protocol described in Section III-A, each participant performed 9 complete trial sets. For comparative analysis, two representative trials were selected under identical task parameters and operator, but with different Intent Inference modes: the gaze-based teleoperation with the GMR-based model and with G-HMM-based model, as illustrated in Fig. 6(a)(b). This pairing enables a detailed examination of the temporal evolution of end-effector-to-target distance under the two schemes. Fig. 6(a) shows that GMR effectively suppresses gaze noise arising from saccadic movements and unintended fixations, producing smoother manipulator trajectories, as evidenced by fewer NIOC (red circles). In contrast, G-HMM, shown in Fig. 6(b), struggles to consistently fit the temporal dynamics of gaze behaviors, leading to fragmented and less coherent motion inference, with noticeable trajectory fluctuations due to its limited ability to filter noisy gaze patterns.

TABLE I
Comparison of Intent Inference Models

Stage	mode	Objective Metrics of Performance		
		CT (s)	TD (m)	NIOC (times)
S1	GMR	9.84 ± 0.81	1.88 ± 0.15	1.97 ± 1.08
	G-HMM	20.10 ± 2.61	2.48 ± 0.22	5.57 ± 1.39
p value		$p < 0.01$	$p < 0.01$	$p < 0.01$
S2	GMR	13.01 ± 1.06	1.96 ± 0.14	2.67 ± 1.42
	G-HMM	21.03 ± 2.20	2.80 ± 0.31	6.79 ± 1.91
p value		$p < 0.01$	$p < 0.01$	$p < 0.01$

To further evaluate the two intent inference models beyond the representative trials in Fig. 6(a)(b), we further examined the performance distributions of all 12 participants under both models. Fig. 6(c) illustrates the statistical characteristics of three objective metrics during S1, whereas Fig. 6(d) presents the corresponding distributions for S2. This separation allows for a more intuitive decoupling of horizontal and vertical variables, facilitating clearer comparisons across different stages. Each box spans the interquartile range (IQR), with the top and bottom edges representing the 75th and 25th percentiles, respectively. The black horizontal line within each box denotes the median, while gray points indicate the mean values. The distributions for all objective measures were confirmed to follow approximate normality, which justified subsequent parametric tests. GMR consis-

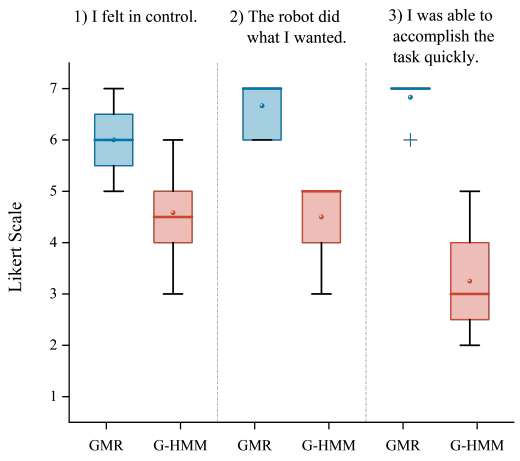

Fig. 7. Survey results via a Likert scale (1: strongly disagree; 7: strongly agree). Boxes depict the 25th, 50th, and 75th percentiles, whereas whiskers indicate the range excluding outliers; those beyond 1.5 interquartile ranges are plotted as "+". The GMR-based method (blue) consistently achieved significantly superior ratings compared to the G-HMM-based method (red) across all evaluated items ($p < 0.01$).

tently outperformed G-HMM across objective measures, demonstrating improvements in trajectory smoothness and operational efficiency. Although the performance gain in S2 was slightly reduced, likely due to the added difficulty of height variations, it still showed a clear advantage over G-HMM.

Table I summarizes the detailed average performance and standard deviations for each objective measure under the two intent inference models—averaged over all trials of 12 participants. In S1, GMR achieved markedly better performance, with a 53% reduction in CT, a 24% decrease in TD, and a 64% reduction in the number of NIOC. S2 exhibited similar trends across all objective measures, although the magnitude of improvement was slightly reduced due to the introduction of vertical variation, which increased task complexity. A paired-sample *t*-test was conducted for each metric, and all comparisons yielded statistically significant differences ($p < 0.01$), indicating consistent performance improvements when using GMR.

To conclude, GMR consistently achieves lower values across all objective measures, demonstrating the effectiveness in improving trajectory smoothness and operational efficiency. By filtering out both saccadic interference and unintended fixations, the model enhances motion stability and reduces unnecessary trajectory deviations. This results in more efficient and controlled manipulator movements, particularly in tasks involving frequent visual shifts.

C. Subjective Measures of Performance

Fig. 7 presents the mean post-trial questionnaire ratings for each intent inference model, as defined in Section III-B. Participants' responses under the GMR-based and G-HMM-based conditions were compared using paired t-tests. Significant differences were found for all three Likert-scale statements ($p < 0.01$), with consistently higher ratings obtained for the GMR-based model. For the statement "I felt in control," the GMR-based model yielded a mean (SD) score of 6.00 (0.74), compared to 4.58 (0.90) for the G-HMM-based model. For "The robot did what I wanted," the corresponding scores were 6.67 (0.43) and 4.50 (0.67), respectively. The largest performance gap was observed for "I was able to accomplish the task quickly," where the mean (SD) ratings were 6.84 (0.39) for the GMR-based model and 3.25 (1.06) for the G-HMM-based model.

In the post-experiment interview, 11 out of 12 participants expressed a preference for the GMR. A representative comment noted:

- *"The G-HMM sometimes gave me the impression that the arm could not reach the target, which created considerable pressure. I could clearly feel that as it struggled, my gaze interference increased, making the motion appear less smooth."*

Conversely, one participant reported preferring the G-HMM:

- *"Although the GMR was very fast and my gaze only introduced minor disturbances, its speed sometimes left me little time to react. In contrast, even though the G-HMM felt harder to use, it still allowed more time for adjustment during movement."*

Overall, the Likert scale and post-experiment surveys indicate that, from a subjective perspective, participants consistently perceived the GMR-based model as offering a stronger sense of involvement and control, along with better perceived temporal efficiency. These impressions were in line with the improvements observed in the objective metrics when compared to the G-HMM baseline.

V. Conclusion and Discussion

This study proposed a gaze-based teleoperation with intent inference model to enhance control in unstructured environments. Instead of directly mapping gaze points to robotic motions, the model infers user intent from gaze behaviors and manipulator states, thereby enabling accurate prediction of the next-step end-effector positions. This design allows the system to dynamically adjust the trajectory while effectively suppressing oscillations induced by rapid saccadic eye movements and unintended fixations. Consequently, the proposed approach improves both motion smoothness and operational efficiency, enhancing the user's sense of involvement and control in human–robot interaction.

To evaluate performance, we conducted comparative experiments using two models: the proposed intent inference model based on Gaussian Mixture Regression (GMR) and a benchmark model based on Gaussian Hidden Markov Models (G-HMM). Experimental results demonstrated that the GMR-based model consistently outperformed the G-HMM-based baseline. In terms of objective metrics, it achieved higher control efficiency and improved motion smoothness of the manipulator. For subjective evaluation, participants reported that the GMR-based model offered a better sense

of involvement and control, a better perceived temporal efficiency. While a structured grid was used for rigorous comparison and data acquisition, this controlled setup effectively isolates the performance of the intent inference mechanism. The proposed GMR model is intrinsically task-agnostic, providing a foundational step toward freeform teleoperation in unconstrained environments.

However, the study has several limitations. The dataset was relatively small, involving only five experimenters, which limits the model's generalizability and adaptability to a broader population. Additionally, the experimental scenarios were restricted to manipulator motion control without integration of gripper actuation, thus covering only part of a complete manipulation pipeline. Despite these limitations, the results indicate potential for real-world applications, such as assistive robots for medical care, rehabilitation training, and support for people with disabilities.

Future work will focus on expanding the dataset by incorporating a larger number of participants and a denser set of target configurations to improve statistical robustness and generalizability across the manipulator workspace, while integrating reaching with autonomous grasping and placement into a unified gaze-driven teleoperation pipeline.

VI. Acknowledgment

The authors would like to thank colleagues from the same laboratory for their assistance in dataset construction and for providing valuable suggestions on this work.

References

[1] A. Bilyea, N. Seth, S. Nesathurai, and H. Abdullah, "Robotic assistants in personal care: A scoping review," *Medical engineering & physics*, vol. 49, pp. 1–6, 2017.

[2] Y. Qin, W. Yang, B. Huang, K. Van Wyk, H. Su, X. Wang, Y.-W. Chao, and D. Fox, "Anyteleop: A general vision-based dexterous robot arm-hand teleoperation system," *arXiv preprint arXiv:2307.04577*, 2023.

[3] S. Musić and S. Hirche, "Control sharing in human-robot team interaction," *Annual Reviews in Control*, vol. 44, pp. 342–354, 2017.

[4] S. Luo, Q. Peng, J. Lv, K. Hong, K. R. Driggs-Campbell, C. Lu, and Y.-L. Li, "Human-agent joint learning for efficient robot manipulation skill acquisition," *arXiv preprint arXiv:2407.00299*, 2024.

[5] S. Falcone, G. Englebienne, J. Van Erp, and D. Heylen, "Toward standard guidelines to design the sense of embodiment in teleoperation applications: A review and toolbox," *Human–Computer Interaction*, vol. 38, no. 5-6, pp. 322–351, 2023.

[6] T. B. Pulikottil, M. Caimmi, M. G. D'Angelo, E. Biffi, S. Pellegrinelli, and L. M. Tosatti, "A voice control system for assistive robotic arms: preliminary usability tests on patients," in *2018 7th IEEE International Conference on Biomedical Robotics and Biomechatronics (Biorob)*. IEEE, 2018, pp. 167–172.

[7] O. Rogalla, M. Ehrenmann, R. Zollner, R. Becher, and R. Dillmann, "Using gesture and speech control for commanding a robot assistant," in *Proceedings. 11th IEEE International Workshop on Robot and Human Interactive Communication*. IEEE, 2002, pp. 454–459.

[8] L. Bi, C. Guan *et al.*, "A review on emg-based motor intention prediction of continuous human upper limb motion for human-robot collaboration," *Biomedical Signal Processing and Control*, vol. 51, pp. 113–127, 2019.

[9] A. F. Salazar-Gomez, J. DelPreto, S. Gil, F. H. Guenther, and D. Rus, "Correcting robot mistakes in real time using eeg signals," in *2017 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2017, pp. 6570–6577.

[10] L. R. Hochberg, D. Bacher, B. Jarosiewicz, N. Y. Masse, J. D. Simeral, J. Vogel, S. Haddadin, J. Liu, S. S. Cash, P. Van Der Smagt *et al.*, "Reach and grasp by people with tetraplegia using a neurally controlled robotic arm," *Nature*, vol. 485, no. 7398, pp. 372–375, 2012.

[11] H. Admoni and B. Scassellati, "Social eye gaze in human-robot interaction: a review," *Journal of Human-Robot Interaction*, vol. 6, no. 1, pp. 25–63, 2017.

[12] O. Palinko, F. Rea, G. Sandini, and A. Sciutti, "Robot reading human gaze: Why eye tracking is better than head tracking for human-robot collaboration," in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2016, pp. 5048–5054.

[13] M.-Y. Wang, A. A. Kogkas, A. Darzi, and G. P. Mylonas, "Free-view, 3d gaze-guided, assistive robotic system for activities of daily living," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018, pp. 2355–2361.

[14] A. Shafti, P. Orlov, and A. A. Faisal, "Gaze-based, context-aware robotic system for assisted reaching and grasping," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 863–869.

[15] H. Zeng, Y. Shen, X. Hu, A. Song, B. Xu, H. Li, Y. Wang, and P. Wen, "Semi-autonomous robotic arm reaching with hybrid gaze–brain machine interface," *Frontiers in neurorobotics*, vol. 13, p. 111, 2020.

[16] S. Li and X. Zhang, "Implicit intention communication in human–robot interaction through visual behavior studies," *IEEE Transactions on Human-Machine Systems*, vol. 47, no. 4, pp. 437–448, 2017.

[17] A. Haji Fathaliyan, X. Wang, and V. J. Santos, "Exploiting three-dimensional gaze tracking for action recognition during bimanual manipulation to enhance human–robot collaboration," *Frontiers in Robotics and AI*, vol. 5, p. 25, 2018.

[18] X. Wang, A. Haji Fathaliyan, and V. J. Santos, "Toward shared autonomy control schemes for human-robot systems: Action primitive recognition using eye gaze features," *Frontiers in Neurorobotics*, vol. 14, p. 567571, 2020.

[19] K. Hauser, "Recognition, prediction, and planning for assisted teleoperation of freeform tasks," *Autonomous Robots*, vol. 35, pp. 241–254, 2013.

[20] C. M. Bishop and N. M. Nasrabadi, *Pattern recognition and machine learning*. Springer, 2006, vol. 4, no. 4.

[21] X. Wang and V. J. Santos, "Gaze-based shared autonomy framework with real-time action primitive recognition for robot manipulators," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 31, pp. 4306–4317, 2023.

[22] A. Röfer, I. Nematollahi, T. Welschehold, W. Burgard, and A. Valada, "Bayesian optimization for sample-efficient policy improvement in robotic manipulation," 2024. [Online]. Available: https://arxiv.org/abs/2403.14305

[23] H. Ismail Fawaz, G. Forestier, J. Weber, L. Idoumghar, and P.-A. Muller, "Deep learning for time series classification: a review," *Data Mining and Knowledge Discovery*, vol. 33, no. 4, p. 917–963, Mar. 2019. [Online]. Available: http://dx.doi.org/10.1007/s10618-019-00619-1

[24] H. Zhang, X. Han, M. Fu, and W. Zhou, "Robot obstacle avoidance learning based on mixture models," *Journal of Robotics*, vol. 2016, no. 1, p. 7840580, 2016.

[25] C. Ye, J. Yang, and H. Ding, "Bagging for gaussian mixture regression in robot learning from demonstration," *Journal of Intelligent Manufacturing*, vol. 33, no. 3, pp. 867–879, 2022.

[26] W. Jannah and D. R. Saputro, "Parameter estimation of gaussian mixture models (gmm) with expectation maximization (em) algorithm," in *AIP Conference Proceedings*, vol. 2566, no. 1. AIP Publishing, 2022.

[27] R. J. Kate, "Using dynamic time warping distances as features for improved time series classification," *Data mining and knowledge discovery*, vol. 30, pp. 283–312, 2016.

[28] J. Manrique-Cordoba, C. Martorell-Llobregat, M. Á. de la Casa-Lillo, and J. M. Sabater-Navarro, "Trajectory learning using hmm: Towards surgical robotics implementation," *Sensors*, vol. 25, no. 11, p. 3487, 2025.

[29] S. Javdani, H. Admoni, S. Pellegrinelli, S. S. Srinivasa, and J. A. Bagnell, "Shared autonomy via hindsight optimization for teleoperation and teaming," *The International Journal of Robotics Research*, vol. 37, no. 7, pp. 717–742, 2018.

[30] F. Bjelonic, J. Lee, P. Arm, D. Sako, D. Tateo, J. Peters, and M. Hutter, "Learning-based design and control for quadrupedal robots with parallel-elastic actuators," *IEEE Robotics and Automation Letters*, vol. 8, no. 3, pp. 1611–1618, 2023.