IEEE Robotics and Automation Letters (RAL) paper presented at
2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
October 14-18, 2024. Abu Dhabi, UAE

Forward Prediction of Target Localization Failure Through Pose Estimation Artifact Modelling

Morgan Windsor[1], Alejandro Fontan[1], Peter Pivonka[2], and Michael J Milford[1]

***Abstract*—For safety critical applications the ability of localization systems to self-assess their own performance and know when they are failing is as important as absolute accuracy. Previous methods have self-assessed current system performance, identifying failure after it occurs. We propose to instead pre-emptively avoid failure by predicting likely localization performance at locations not yet explored by a robot. To achieve this, we propose an approach for supervising a target object localization system by modelling trends in internal pipeline artifacts that are predictive of localization accuracy. We use this model to predict where acceptable localization performance is possible and where failure is likely. We evaluate our approach with both off-line recorded datasets and live robot experiments in the context of an upper limb surgical task using human bone phantoms as localization targets. We demonstrate our approach implemented as both a *Long Range Predictor* for use in informing future planning, and a *Next-Step Predictor*, for ongoing task supervision to stop a robot before reaching localization failure. We show that our method provides significant improvement over a naive baseline achieving a mean increase in safe path length, or usable workspace, without localization failure of 84.1% for our long range predictor and 102.1% for our next-step predictor.**

***Index Terms*—Localization, Computer Vision for Medical Robotics, Vision-Based Navigation, Medical Robots and Systems.**

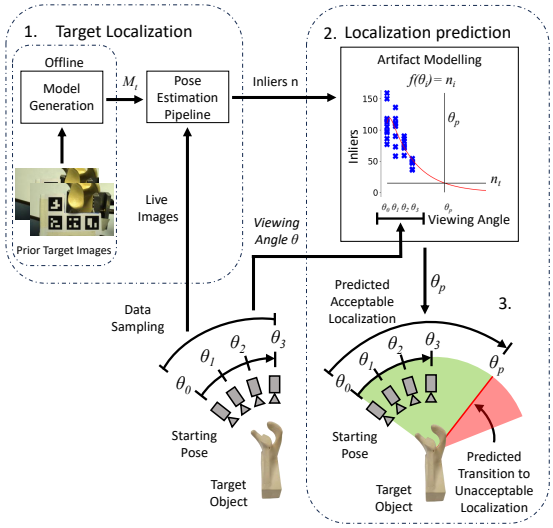

Fig. 1: Our approach extracts artifacts from a pose estimation pipeline (1) and builds a model of the artifact with respect to robot pose (2). This model is then used to predict where localization performance is likely to transition between acceptable and unacceptable performance (3).

I. INTRODUCTION

FOR safety critical domains such as autonomous vehicles and surgical robotics, the ability for their localization systems to self-identify poor performance is often as important as their absolute accuracy. While methods for localization self-assessment have been developed [1]–[4] these approaches detect failure after it has occurred. Recognizing that reaching the point of failure, even if detected, may be unacceptable, we propose a new approach that instead looks to predict where localization failures are likely to occur, enabling a robot to anticipate and avoid them.

Manuscript received: November 19, 2023; Revised February 18, 2024; Accepted March 11, 2024.

This paper was recommended for publication by Editor Pascal Vasseur upon evaluation of the Associate Editor and Reviewers' comments. This research was partially supported by the QUT Centre for Robotics and ARC Laureate Fellowship FL210100156 to MM. MW is supported by an ARC Industrial Transformation Training Centre (ITTC) for Joint Biomechanics grant IC190100020 and an Australian Government Research Training Program (RTP) Scholarship.

[1]MW, AF, and MM are with the QUT Centre for Robotics, School of Electrical Engineering and Robotics at the Queensland University of Technology. Email: morgan.windsor@hdr.qut.edu.au

[2]PP is with the QUT Centre for Biomedical Technologies, School of Mechanical, Medical, and Process Engineering at the Queensland University of Technology.

Digital Object Identifier (DOI): see top of this page.

In this work we look to predict the performance of a target localization system, based on 6 degree-of-freedom (6DOF) object pose recovery, at locations unexplored by a robot. We present an approach that uses data sampling during a robot task to build a model that predicts where target localization performance is likely to be acceptable and where localization is likely to fail. Specifically, we make the following contributions:

1) A new method for predicting localization performance at locations not yet visited by a robot through modelling internal pose estimation pipeline artifacts sampled at run time (Fig. 1).
2) Two implementations of our core prediction approach: a *Long Range Predictor* for robot planning that uses limited scene sampling to predict where localization performance is likely to be acceptable, and a *Next-Step Predictor* for ongoing task supervision that continually updates the performance model and predicts when the robot is close to the boundary of acceptable localization performance.
3) Extensive experiments, including live robot testing, within the context of robot assisted guide-wire placement, a critical step in the orthopaedic surgery of

shoulder arthroplasty (joint replacement) that requires constant accurate localization of the target bone.

In the remainder of this paper, we present the background of our work and relevant prior research. We describe our prediction approach and experimental methodology, and present results demonstrating the utility of our method in the context of a robotic upper limb surgical procedure. Finally, we conclude by discussing the key findings of our work and future research directions.

II. BACKGROUND

Here we present an overview of 6DOF pose estimation techniques, localization system integrity, and the use of prediction based approaches in robot perception.

A. Object Pose Estimation

6DOF object pose estimation is an important and widely studied problem in both robotics and computer vision. Techniques have been published using RGB images, depth information, and both. Approaches vary considerably in terms of mathematical model [5]. Traditional approaches include template matching, geometric, and algorithmic techniques [6]–[10]. Modern deep-learning methods include end-to-end learned approaches [11], [12] and hybrid approaches where deep learned stages inform traditional pose recovery or refinement [13]–[15]. State-of-the art performance on open challenges is now achieved by deep-learned systems combining color and depth data [16].

The ability of 6DOF pose estimation systems to return accurate results can be impacted by occlusion, clutter, texture, similar looking distractors, and variation in viewpoint and lighting [5], [17]. Some of these factors are intrinsic to the scene or target; others such as viewpoint and occlusion, can be influenced by sensor positioning within the scene.

B. Localization Integrity

Localization integrity is the trust we have that a localization estimate is within an acceptable tolerance [18]. This is especially important for safety critical applications such as robotic surgery and autonomous vehicles. Despite continuing improvement in visual localization system performance, localization integrity remains under explored [19]. Recent work has presented methods for self-evaluation of generalized autonomous systems [20], autonomous vehicles [18], [21], visual place recognition systems [1]–[3] and target localization in surgical robots [4]. These works however are reactive and detect failure as or after it occurs, rather than predicting future performance to avoid failure. Prediction of future perception system failure has been proposed for autonomous vehicles [22]. The need for data collection over repeated traversals however limits this approach in entirely new environments or for once-off tasks.

C. Prediction In Robot Applications

Prediction approaches have been applied to robotics in areas such as action planning in environments with other agents. Long short-term memory networks have become popular in prediction tasks including vehicle and pedestrian trajectory prediction [23], [24], a widely studied area applicable to autonomous vehicles [25]. Human action recognition and future action prediction have also been studied extensively [26].

The neuroscience inspired *PredNet* network performs next frame prediction in video sequences, implicitly learning object and scene structure and predicting changes in the scene from both camera and object movements [27].

Prediction is also applied directly to perception in next-best-view planning where a system looks to select sensor positions that maximize performance. In common *generate-and-test* approaches candidate views are generated and assessed based on their predicted information gain [28]. In object pose estimation tasks this assessment can be based on direct prediction of target feature visibility [29] or predicted entropy metrics [30].

III. APPROACH

Here we describe our approach to predicting the performance of a target localization system at locations not yet explored by a robot. We present an overview of our system, the underlying 6DOF pose estimation pipeline, and the pipeline artifacts used to predict performance. We describe our approach to artifact modelling, and our use of this model to predict where localization failure is likely to occur with our long range and next-step predictors.

A. System Overview

Our core approach consists of three major stages, localization pipeline artifact extraction, artifact modelling, and localization performance prediction. This process is common to both the long range and next-step predictor implementations of our approach, which differ as described in III-E.

In our approach, as the robot moves during a task, we extract artifacts from within an underlying localization pipeline that are predictive of target pose estimate accuracy. We generate a model of the artifact with respect to the robot pose and use this model to predict artifact values at unexplored locations. We use the predicted artifact values to estimate where localization accuracy will be within acceptable tolerance of ground truth and restrict motion of the robot to avoid localization failure. We call this area of acceptable localization performance the *localization envelope*. The architecture of our approach and its interactions with the target localization system is shown in Fig. 1.

B. Underlying Target Localization Pipeline

Our approach builds on top of an image based pose estimation pipeline which localizes a target object relative to the robot by recovering 6DOF pose estimates using keypoint feature correspondences.

Offline, we build a static sparse feature model of the target object, M_t, through stereo image reconstruction before the

This article has been accepted for publication in IEEE Robotics and Automation Letters. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/LRA.2024.3382528

robot task begins. We use two images of a static scene captured with different camera poses and reconstruct the scene as a sparse point cloud of keypoint features. We manually trim the reconstruction to retain only points belonging to our target and construct our model such that $M_t = [f_0, ..., f_n]$ where each feature, f_i, consists of a keypoint feature descriptor and a 3D coordinate in the desired target coordinate frame.

At run time, observed keypoint features from incoming camera frames are matched against the features contained in M_t and a list of 2D-3D correspondences is generated. 6DOF pose estimation is then performed using a perspective-n-point solver and RANdom SAmpling Consensus outlier rejection (PnP-RANSAC).

C. Localization Performance Prediction Artifacts

Our approach is based on the central idea of predicting values of pose estimation pipeline artifacts, that are themselves predictive of performance, at unexplored locations.

In previous work we identified multiple performance predictive artifacts for the underlying PnP-RANSAC pipeline [4]. In this work we focus on one, the number of keypoint features returned by the RANSAC loop as inliers, n_i. This was observed to be highly predictive of pose estimation error. While PnP algorithms can generate solutions with as few as 3 correspondences ($n_i = 3$), in practice, when n_i reaches a critical lower threshold, n_t, the RANSAC algorithm is unable to reliably reject incorrect correspondences as outliers resulting in a significant increase in pose estimation error.

D. Localization Prediction Artifact Modelling

To predict artifact values at unexplored locations we use a *sample-and-model* approach. As the robot moves during a task we record n_i for each target pose estimate returned by the localization pipeline along with P_r, the robot pose at that time. We then generate a model for the artifact value with respect to the robot pose, $f : P_r \rightarrow n_i$, which we call the *artifact model*.

In our application, robot assisted guide-wire placement, the surgical approach used means the viewing angle of the target bone when placing the guide-wire is typically different to the initial view available when generating M_t. As such the type of motion we are most interested in is change of viewing angle about the target object. We therefore formulate the generation of the artifact model as a regression problem. We set the dependent variable as n_i and the independent variable as θ_i, the change in target viewing angle, measured as the absolute magnitude of the difference in the camera orientations between the pose where M_t was created and robot pose at position i.

To determine suitable methods for fitting the artifact model we evaluated non-linear least squares regression, with a range of possible base functions, and a multilayer perceptron (MLP). Evaluations were performed using the viewing angle change datasets collected as described in IV-D. Each approach was assessed in terms of Root Mean Square Error (RMSE) averaged across the datasets (Fig. 2). The base functions assessed were selected having observed a decay relationship between n_i and θ_i.

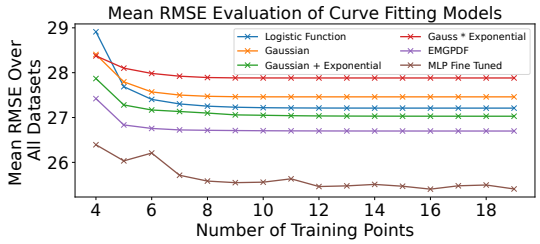

Fig. 2: Methods for fitting the artifact model were evaluated with mean RMSE measured across pre-recorded datasets. Curves were fit to n_i vs θ_i using data from the start of each dataset with different numbers of positions sampled. RMSE was calculated over the entire dataset.

The lowest mean RMSE for the least squares regression was returned by the Exponentially Modified Gaussian Probability Density Function (EMGPDF) as implemented in the SciPy library [31], scaled by a constant. The parametrization of this distribution is shown in (1) where α, K, and μ are fit parameters and *erfc* is the complementary error function defined in (2).

When trained using only sampled data points the MLP model failed to generalize beyond the training range returning very high RMSE. To achieve a mean RMSE competitive with least square regression, the MLP was first trained on an EMGPDF function matching the least squares regression initial parameter guess. The MLP was then fine tuned with actual sampled data further supplemented with artificial points at $(\theta_i, N_i) = (45, 0)$. Due to the added complexity of implementation for the MLP we determined to use least square regression with the EMGPDF base function for artifact model generation.

$$f(\theta, \alpha, K, \mu) = \frac{\alpha}{2K} \exp\left(\frac{1}{2K^2} - \frac{(\theta - \mu)}{K}\right) \operatorname{erfc}\left(-\frac{(\theta - \mu) - \frac{1}{K}}{\sqrt{2}}\right) \quad (1)$$

$$\operatorname{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-t^2} dt \quad (2)$$

E. Long Range and Next-Step Localization Performance Prediction

In our approach we estimate the localization envelope of our task using the artifact model to predict locations where $n_i > n_t$.

We assume the robot starts its task positioned inside the localization envelope. We build the artifact model and predict the boundary of the localization envelope, θ_p, using root finding to determine θ_p where $n_i = n_t$. In our approach we predict likely localization failure for robot poses where $\theta_i > \theta_p$.

We propose two implementations of our approach with different applications, a *long range predictor* for use in motion planning, and a *next-step predictor* to provide continuous task supervision.

For the long range predictor we use a limited amount of data sampled at the start of a robot motion to generate an artifact model and predict θ_p.

For next-step prediction we continually update our artifact model and stop the robot when it is within a fixed distance, θ_b, of the predicted localization envelope boundary. We generate an initial artifact model and localization envelope boundary prediction, θ_p, as sampling data becomes available. If $\theta_p > \theta_i + \theta_b$, where θ_i is the current viewing angle we predict acceptable localization performance at the 'next-step' and allow the robot to continue motion, updating the artifact model and θ_p at each step. If at any step $\theta_p < \theta_i + \theta_b$, we predict localization failure before the next step and stop the robot.

F. Application Specific Predictor Tuning

The ability of a localization prediction system to be tuned to meet application specific requirements is important to utility. Localization failure tolerance differs between tasks and even task stages. For example poor localization for a surgical robot is far more problematic while cutting a bone than exploring its working space away from the patient.

In our approach the inlier threshold, n_t, is used to decide if localization performance is predicted to be within acceptable bounds and can be modified to tune the system operating point to meet user requirements. Larger values of n_t tune the predictor to be more *conservative*, reducing the size of the predicted localization envelope, allowing less range of motion but decreasing the likelihood of localization failure. Smaller values of n_t tune the predictor to be more *aggressive*, predicting a larger localization envelope and allowing greater motion at the expense of a higher likelihood of localization failure.

IV. EXPERIMENTAL METHODOLOGY

Here we describe experiments conducted to evaluate our approach in three key areas: accuracy of localization envelope prediction, the effect of threshold tuning on localization envelope prediction, and the effect of predictor based supervision on a live robot task requiring constant target localization. We also describe our experimental environment, testing equipment, data collection processes and evaluation metrics.

A. Localization Envelope Prediction Accuracy and Threshold Tuning Experiments

We assess the accuracy of localization envelope predictions and the effect of inlier threshold tuning for our *Long Range* and *Next-Step* predictors using recorded datasets described in IV-D. We evaluate our approach using baselines and metrics described in IV-E.

For the *Long Range Predictor* we use the first 4 sampling positions from each dataset to generate an artifact model and predict the boundary of the localization envelope. We select 4 sampling positions to ensure that the sampled data is only drawn from within the localization envelope of each dataset while still providing sufficient data span to ensure reasonable curve fitting performance.

For the *Next-Step predictor* we again use the first 4 sampling positions from each dataset to generate initial artifact models and localization envelope predictions. We set the step size, θ_b, to the same sampling interval used during dataset collection and step through the remaining sampling positions of each dataset, updating the artifact model and localization envelope prediction at each step, repeating until localization failure is predicted before the next step.

To evaluate the effect of predictor threshold tuning we repeat each experiment for both predictors using inlier threshold values from 4 to 100.

B. Live Robot Experiments

We evaluate the effect of predictor based supervision on a live robot task with a series of live robot trials. We implement our next-step prediction approach to supervise a robot task inspired by shoulder arthroplasty guide-wire placement. We select a human scapula phantom (SAWBONES Model 1050), pre-drilled for implant placement, as a localization target. We position the tip of a robot mounted guide-wire into the oversized hole in the glenoid (articulating surface on the scapula) and generate a target model. We begin each individual trial with the robot and guide-wire re-positioned at the initial model generation pose, from where we move the robot in an arc about the scapula as shown in Fig. 3. We stop the robot to perform next step prediction every 1.5°, relying on target localization to maintain the position of the guide-wire tip relative to the scapula during motion. During each trial we record the predicted localization envelope as the viewing angle where the predictor would have stopped the robot but allow motion to continue until actual localization failure, defined as the guide-wire making contact with the scapula or being removed entirely from the hole.

We perform 10 trials using 5 target models generated from different initial viewing angles. For each target model we perform 2 trials with the robot moving in arcs along different azimuth angles. For all trials we set $n_t = 15$, selected based on the next-step predictor results described in IV-A.

C. Testing Equipment and Environment

Here we describe the experimental equipment, software implementation, and error measurement processes used during evaluation experiments.

Testing and data collection was performed using a Franka-Emika Panda robot equipped with an IDS U3-3060CP camera rigidly mounted to the final link with a custom adapter forming an eye-in-hand configuration (Fig. 3).

Our approach was implemented with SIFT [32] keypoint features and PnP-RANSAC pose estimation, using USAC, from the OpenCV libraries [33]. To improve feature detection on low texture bone phantoms, images were pre-processed using contrast limited adaptive histogram equalization before feature extraction in all experiments. Curve fitting was performed with non-linear least squares regression using the SciPy optimize library [31]. Robot software was implemented within RoboStack [34].

This article has been accepted for publication in IEEE Robotics and Automation Letters. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/LRA.2024.3382528

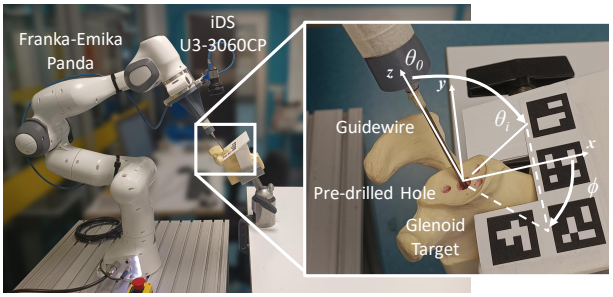

Fig. 3: During Live experiments we place the tip of a robot mounted guide-wire into a pre-drilled hole on the scapula phantom (inset). We move the robot in an arc away from the position where the target model was generated (θ_0) relying on target localization to maintain the guide-wire tip position relative to the phantom. The trial is stopped when the guide-wire makes contact with, or is entirely removed from the phantom. For repeated trials with the same target model the robot is moved in arcs along different azimuth angles (ϕ).

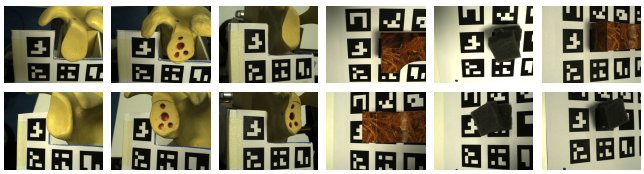

Fig. 4: Sample images from datasets collected in IV-D. Surgical targets were two human scapula phantoms (SAWBONES model 1050), one prepared for glenoid implant placement. Non-surgical targets, highly textured blocks, were included to assess our approach on more general objects.

For off-line testing, ground truth poses of target objects were determined with an ArUco marker board [35] that remained stationary relative to the target throughout data collection. During construction of M_t the origin of the model reference frame was transformed to be coincident with the origin of the ArUco board. Pose estimation error was then calculated as the Euclidean distance, E_t, and rotation magnitude, E_r between the estimated target object pose and the marker board pose.

D. Dataset Collection

For the offline experiments described in IV-A pre-recorded datasets were collected for two types of constrained robot motion. Our main experiment datasets, evaluating viewpoint change, consisted of two images for use in model generation and 200 sample images. During collection the robot was positioned such that the ArUco marker board and target object were within view of the camera. Model generation images were captured with 2cm of camera translation along the camera x-axis between frames. The robot was re-positioned to its initial pose and sample images captured as the robot moved in a 45° arc centred on the target object, stopping every 2.25° to capture 10 images. At each position the camera was oriented towards the target. 21 datasets were collected using 5 combinations of target objects arranged in 7 different scenes. Three datasets were collected for each scene, moving the robot along arcs with azimuth angles of 0°, 90°, and 270° relative to the robot x-axis. Sample images from the datasets are shown in Fig. 4.

Using the pre-drilled scapula phantom target 4 additional datasets were collected with straight line camera motions. 1 pair of model generation images was captured as previously described and used to process all 4 datasets. For each dataset the camera was moved away from the target object in a straight line along the negative camera z-axis. The camera was stopped at 15 sampling positions along the trajectory, separated by 10mm of camera translation. 10 sample images were collected at each position. Starting points for each trajectory were varied by adding viewing angle offsets of 0°, 5°, 10°, and 15° relative to the model generation view.

E. Evaluation Baseline and Metrics

To evaluate our method we compare our approach against two baselines, a naive static baseline and a variation of our predictor using an MLP in place of our proposed curve fitting approach. The naive baseline applies a static localization envelope size to each dataset without consideration to the target object or scene present. The MLP consists of 3 hidden layers of 10 neurons and a TanH activation function. In each experiment the MLP is trained using the fine tuning approach described in section III-D. The localization envelope is then predicted using the trained MLP model in place of our proposed artifact model. We measure envelope size in terms of the absolute magnitude of viewing angle change.

We compare our approach to the baselines using two metrics, *safe path length* (SPL) and *incursion rate* (R_i), which measure the utility and safety of a system respectively. SPL measures the maximum safe workspace permitted by a predictor and is calculated with (3) as the smaller of the two angles θ_p and θ_a, where θ_a is the actual localization envelope boundary. We define θ_a as the viewing angle change at which the mean localization error exceeds a specified error threshold, calculated using linear interpolation between discrete sampled points. A larger SPL indicates more 'usable' space being made available to the robot.

$$SPL = min\{\theta_p, \theta_a\} \tag{3}$$

R_i measures how often a predictor overestimates the size of the localization envelope and permits the robot to move to a position where target localization fails; this is calculated with (4) where I is the number of instances where $\theta_p > \theta_a$, and n is the number of trials assessed. A lower R_i indicates 'safer' operation.

$$R_i = I/n \tag{4}$$

While either metric can be individually optimized by massively over or under predicting the localization envelope, the key challenge is maximising SPL while minimizing R_i, allowing a robot to move right to the edge of localization system failure without actually reaching it.

This article has been accepted for publication in IEEE Robotics and Automation Letters. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/LRA.2024.3382528

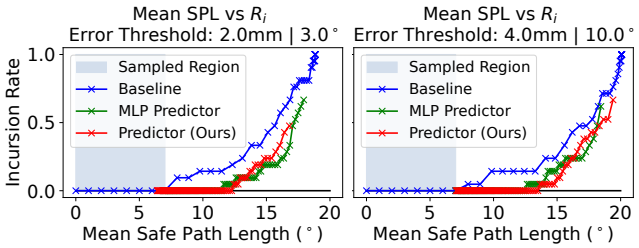

Fig. 5: Evaluation of our long-range predictor on the viewing angle change datasets using mean SPL vs R_i curves. Each point in the plot represents the R_i and mean SPL calculated across all datasets for a single system operating point. In both cases our predictor method shows improvements over the baselines for mean SPL with zero incursions.

V. RESULTS

Here we present results of experiments divided into three areas, long range prediction, next-step prediction, and prediction based task supervision on a live robot.

A. Long Range Predictor Localization Envelope Estimation

We evaluate the performance of our long range predictor on our viewing angle change datasets by comparison to both baselines using mean SPL vs R_i curves generated by modifying the operating point of each system. We modify the static baseline by sweeping the envelope size from 0° to 45°. Similarly, we modify our approach along with the MLP predictor by sweeping n_t. We report R_i and mean SPL calculated across all viewing angle change datasets at each system operating point in Fig. 5. We perform two sets of experiments, the first with 4mm/10° error tolerances selected as the limits that define clinically significant glenoid implant malpositioning in shoulder arthroplasty [36]. The second, using reprocessed datasets for higher precision, with 2mm/3° error tolerances which we select in line with the accuracy of shoulder arthroplasty surgical navigation systems [37], [38] and the accuracy limit of our pose estimation pipeline.

For both error thresholds our predictor out performs the baselines in terms of mean SPL at zero incursion. We achieve a mean relative improvement of 84.1% over the static baseline and 7.4% over the MLP predictor (Table I). Across all operating points assessed, our predictor maintains an equal or lower R_i than the static baseline for equal mean SPL. These results also show the effect of threshold tuning on predictor performance, as the threshold value decreases, prediction becomes more aggressive and the system operating point moves, achieving a larger mean SPL but a higher likelihood of incursion.

We assess the impact of different PnP algorithms on our prediction approach by repeating the viewing angle change dataset experiments at the 2mm/3° error tolerance. We reprocess each dataset through our complete localization and prediction pipelines with three PnP algorithms, EPnP [10], P3P [39], and AP3P [40]. We compare each against the static baseline using the metrics and operating point sweeps previously described (Fig. 6). Our predictor performs similarly

TABLE I: Long Range Predictor Maximum Mean SPL at Zero Incursion For Viewing Angle Change Datasets

System	Safe Path Length (°)		Mean	Relative Improvement (ours)
	2mm/3°	4mm/10°		
Static Baseline	7.0	7.0	7.0	84.1%
MLP	11.5	12.5	12.0	7.4%
Predictor (Ours)	**12.3**	**13.5**	**12.9**	-

across each test without modification. PnP algorithm selection did impact localization performance, and therefore the actual localization envelope, however the impact on our predictor was minimal.

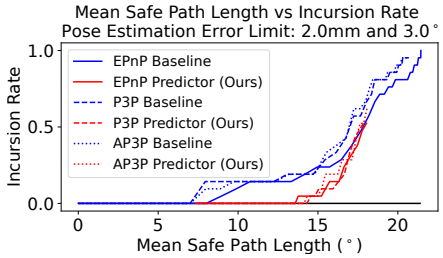

Fig. 6: Evaluation of the impact of different PnP algorithms on our predictor. While varying PnP algorithm did affect the actual localization envelope our predictor continued to outperform the baseline with equal or lower R_i for matched Mean SPL in all tests.

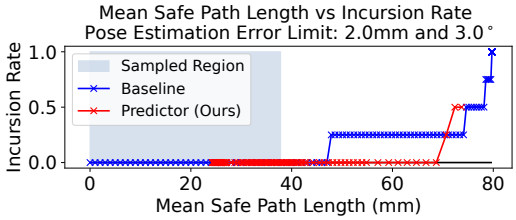

Fig. 7: Mean SPL vs R_i curves for our long-range predictor applied to straight line camera motions. Our predictor was able to adapt to new trajectories and motion types without significant modification.

We evaluate our approach on more diverse trajectories by repeating our long range predictor experiments using the straight line trajectory datasets at the 2mm/3° error tolerance. We adapt our approach to linear motion by modifying the artifact model, setting the independent variable as x_i, the absolute magnitude of the 3D camera translation between the first position of the trajectory and position i. We compare our predictor against the static baseline, swept from 0 to 150mm, using mean SPL vs R_i curves. We replace the angular measurements for SPL with equivalent camera translations measured relative to the trajectory starting point. Our predictor again out performs the naive baseline with a relative improvement of 46.2% for maximum mean SPL at zero incursion (Fig. 7). We note that 5 sampling positions were required to generate a suitable artifact model in these experiments due to the observed n_i vs x_i trend beginning to decay later than the n_i vs θ_i relationship in the

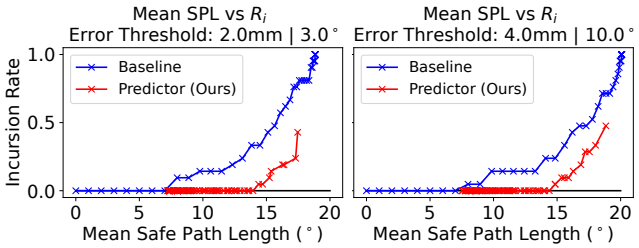

Fig. 8: Our next-step predictor compared to the static baseline for two different error thresholds using Mean SPL vs R_i curves. In both cases the next step predictor outperforms the baseline with lower R_i for equal mean SPL. The next step predictor also achieves larger SPL than the long range predictor.

TABLE II: Next-Step Predictor Maximum Mean SPL at Zero Incursion For Viewing Angle Change Datasets

System	Safe Path Length (°)		Mean	Relative Improvement (ours)
	2mm/3°	4mm/10°		
Baseline	7.0	7.0	7.0	102.1%
Predictor (Ours)	**13.9**	**14.4**	**14.2**	-

viewing angle change datasets. This highlights the importance of sampling data selection in prediction models.

B. Next-Step Predictor Localization Envelope Estimation

We evaluate the performance of our next-step predictor with the viewing angle change datasets. We use the same performance metrics and system operating points described in V-A and compare against the static baseline. As a more realistic measure of the available robot workspace when using this approach we report mean SPL calculated using θ_p as the viewing angle where the predictor would stop the robot rather than the predicted localization envelope boundary.

Fig. 8 shows the SPL vs R_i curves for the next-step predictor which achieves consistently lower R_i than the static baseline for equal mean SPL. Our next-step predictor outperforms the static baseline in terms of mean SPL at zero incursion with a relative improvement of 102.1% (Table II). This result also represents a relative improvement of 9.8% over the long-range predictor.

In Table III we show the inlier threshold values which achieve maximum mean SPL at zero incursion indicating that the next-step predictor, with ongoing performance model updates, does not need to be tuned as conservatively as the long range predictor to achieve the same incursion rate.

TABLE III: Predictor Inlier Thresholds for Maximum Mean SPL at Zero Incursion

System	Inlier Threshold (n_t)	
	2mm/3°	4mm/10°
Long Range Predictor	23	28
Next-Step Predictor	**12**	**15**

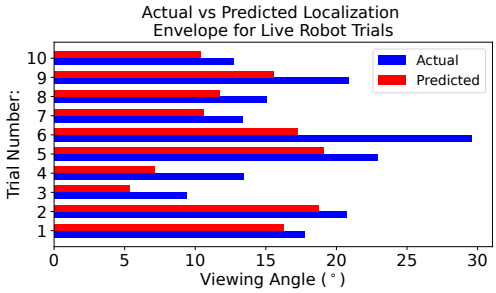

Fig. 9: Predicted stop locations and actual localization envelopes for live robot trials. Our method for task supervision prevented the robot from reaching localization failure while still allowing the robot to use most of the workspace where acceptable localization was possible.

C. Live Robot Experiments

We evaluate our prediction approach in the live robot experiments described in section V-C by comparing the predicted stop locations and actual localization envelopes for each trial. As shown in Fig. 9, across all trials our predictor stopped the robot before reaching localization failure, achieving an incursion rate of 0. The inlier threshold of $n_t = 15$ used for these trials was selected from the next-step predictor results in Table III for the 4mm/10° error threshold, showing that the system has some generalization capability in terms of transferring parameters between localization scenarios requiring similar error tolerance. Across all trials the predictor permitted the robot to use on average 74.7% of the actual localization envelope. This is an improvement over a system equivalent to our static baseline, which would have permitted only 60.4% usage, however there is still a considerable gap between this performance and ground truth. A contributing factor to this is the shape of the n_i vs θ_i trend; the inlier values that result in localization failure are typically on sections of the curve with shallow gradients resulting in prediction sensitivity. This sensitivity resulted in very early stopping in Trials 3, 4 and 6.

VI. DISCUSSION

We have presented a novel method to predict, at locations not yet visited by a robot, if acceptable target localization performance is possible. We have shown that our method can be used for both long range prediction and direct task supervision significantly reducing the likelihood of localization failure while maximizing the available robot working space. We have also demonstrated that our method translates well to live robot implementation.

We propose three key areas for future work to further develop our approach. In our current method, we select our prediction threshold experimentally, future work should look to develop methods to select prediction thresholds based on task specific localization accuracy requirements and failure tolerance. Where our current approach predicts localization performance for a specific constrained motion type using a two-dimensional artifact model, future work should look to increase generalization by building higher dimensional models

that account for additional motion types. Finally, we have demonstrated our approach with a static localization pipeline; future work could make use of multiple pipelines or pipeline configurations and select the most appropriate for deployment based on predicted performance.

REFERENCES

[1] J. Chen, J. Monica, W.-L. Chao, and M. Campbell, "Probabilistic uncertainty quantification of prediction models with application to visual localization," in *2023 IEEE Int. Conf. on Robot. and Automat. (ICRA)*, 2023, pp. 4178–4184.

[2] H. Carson, J. J. Ford, and M. Milford, "Predicting to Improve: Integrity Measures for Assessing Visual Localization Performance," *IEEE Robot. and Automat. Lett.*, pp. 1–8, 2022.

[3] ——, "Unsupervised quality prediction for improved single-frame and weighted sequential visual place recognition," in *2023 IEEE Int. Conf. on Robot. and Automat. (ICRA)*, 2023, pp. 3955–3961.

[4] M. Windsor, J. Peng, A. Gupta, P. Pivonka, and M. J. Milford, "Pose quality prediction for vision guided robotic shoulder arthroplasty," in *2023 IEEE Int. Conf. on Robot. and Automat. (ICRA)*, 2023, pp. 4797–4804.

[5] C. Sahin, G. Garcia-Hernando, J. Sock, and T.-K. Kim, "A review on object pose recovery: From 3D bounding box detectors to full 6D pose estimators," *Image and Vis. Comput.*, vol. 96, p. 103898, Apr 2020.

[6] S. Hinterstoisser, V. Lepetit, S. Ilic, P. Fua, and N. Navab, "Dominant orientation templates for real-time detection of texture-less objects," in *2010 IEEE Computer Society Conf. on Comput. Vision and Pattern Recognit.*, 2010, pp. 2257–2264.

[7] E. Muñoz, Y. Konishi, V. Murino, and A. Del Bue, "Fast 6d pose estimation for texture-less objects from a single rgb image," in *2016 IEEE Int. Conf. on Robot. and Automat. (ICRA)*. IEEE, 2016, pp. 5623–5630.

[8] A. S. Mian, M. Bennamoun, and R. A. Owens, "Automatic correspondence for 3D modeling: as extensive review," *Int. J. of Shape Model.*, vol. 11, no. 02, pp. 253–291, Dec 2005.

[9] J. Vidal, C.-Y. Lin, and R. Marti, "6D pose estimation using an improved method based on point pair features," in *2018 4th Int. Conf. on Control, Automat. and Robot. (ICCAR)*. IEEE, Apr 2018, pp. 405–409.

[10] V. Lepetit, F. Moreno-Noguer, and P. Fua, "EPnP: An accurate O(n) solution to the PnP problem," *Int. J. of Comput. Vis.*, vol. 81, no. 2, pp. 155–166, Feb 2009.

[11] Y. Xiang, T. Schmidt, V. Narayanan, and D. Fox, "Posecnn: A convolutional neural network for 6d object pose estimation in cluttered scenes," *arXiv preprint arXiv:1711.00199*, 2017.

[12] W. Kehl, F. Manhardt, F. Tombari, S. Ilic, and N. Navab, "Ssd-6d: Making rgb-based 3d detection and 6d pose estimation great again," in *Proc. of the IEEE Int. Conf. on Comput. Vis. (ICCV)*, Oct 2017.

[13] K. Park, T. Patten, and M. Vincze, "Pix2pose: Pixel-wise coordinate regression of objects for 6d pose estimation," in *Proc. of the IEEE/CVF Int. Conf. on Comput. Vis. (ICCV)*, October 2019.

[14] R. König and B. Drost, "A hybrid approach for 6dof pose estimation," in *European Conf. on Comput. Vis. (ECCV)*. Springer, 2020, pp. 700–706.

[15] X. Hu, A. Nguyen, and F. R. y. Baena, "Occlusion-robust visual markerless bone tracking for computer-assisted orthopedic surgery," *IEEE Tran. on Instrum. and Meas.*, vol. 71, pp. 1–11, 2022.

[16] M. Sundermeyer, T. Hodaň, Y. Labbe, G. Wang, E. Brachmann, B. Drost, C. Rother, and J. Matas, "Bop challenge 2022 on detection, segmentation and pose estimation of specific rigid objects," in *Proc. of the IEEE/CVF Conf. on Comput. Vision and Pattern Recognit. (CVPR)*, 2023, pp. 2784–2793.

[17] J. Chen, L. Zhang, Y. Liu, and C. Xu, "Survey on 6d pose estimation of rigid object," in *2020 39th Chinese Control Conf. (CCC)*, 2020, pp. 7440–7445.

[18] J. Al Hage, P. Xu, P. Bonnifait, and J. Ibanez-Guzman, "Localization Integrity for Intelligent Vehicles Through Fault Detection and Position Error Characterization," *IEEE Trans. on Intell. Transp. Syst.*, vol. 23, no. 4, pp. 2978–2990, Apr 2022.

[19] Y. D. Yasuda, L. E. G. Martins, and F. A. Cappabianco, "Autonomous visual navigation for mobile robots: A systematic literature review," *ACM Comput. Surv. (CSUR)*, vol. 53, no. 1, pp. 1–34, 2020.

[20] A. Gautam, T. Whiting, X. Cao, M. A. Goodrich, and J. W. Crandall, "A Method for Designing Autonomous Robots that Know Their Limits," in *2022 Int. Conf. on Robot. and Automat. (ICRA)*. IEEE, May 2022, pp. 121–127.

[21] G. D. Arana, O. A. Hafez, M. Joerger, and M. Spenko, "Localization safety validation for autonomous robots," in *2020 IEEE/RSJ Int. Conf. on Intell. Robots and Syst. (IROS)*. IEEE, 2020, pp. 6276–6281.

[22] C. Gurău, D. Rao, C. H. Tong, and I. Posner, "Learn from experience: Probabilistic prediction of perception performance to avoid failure," *The Int. J. of Robot. Res.*, vol. 37, no. 9, pp. 981–995, 2018. [Online]. Available: https://doi.org/10.1177/0278364917730603

[23] S. Dai, L. Li, and Z. Li, "Modeling vehicle interactions via modified lstm models for trajectory prediction," *IEEE Access*, vol. 7, pp. 38 287–38 296, 2019.

[24] H. Xue, D. Q. Huynh, and M. Reynolds, "Ss-lstm: A hierarchical lstm model for pedestrian trajectory prediction," in *2018 IEEE Winter Conf. on Appl. of Comput. Vis. (WACV)*, 2018, pp. 1186–1194.

[25] M. Gulzar, Y. Muhammad, and N. Muhammad, "A survey on motion prediction of pedestrians and vehicles for autonomous driving," *IEEE Access*, vol. 9, pp. 137 957–137 969, 2021.

[26] Y. Kong and Y. Fu, "Human action recognition and prediction: A survey," *Int. J. of Comput. Vis.*, vol. 130, no. 5, pp. 1366–1401, 2022.

[27] W. Lotter, G. Kreiman, and D. D. Cox, "Deep predictive coding networks for video prediction and unsupervised learning," *CoRR*, vol. abs/1605.08104, 2016. [Online]. Available: http://arxiv.org/abs/1605.08104

[28] R. Zeng, Y. Wen, W. Zhao, and Y.-J. Liu, "View planning in robot active vision: A survey of systems, algorithms, and applications," *Comput. Vis. Media*, vol. 6, no. 3, pp. 225–245, sep 2020.

[29] K. Wu, R. Ranasinghe, and G. Dissanayake, "Active recognition and pose estimation of household objects in clutter," in *2015 IEEE Int. Conf. on Robot. and Automat. (ICRA)*, 2015, pp. 4230–4237.

[30] J. Sock, S. H. Kasaei, L. S. Lopes, and T.-K. Kim, "Multi-view 6d object pose estimation and camera motion planning using rgbd images," in *2017 IEEE Int. Conf. on Comput. Vis. Workshops (ICCVW)*, 2017, pp. 2228–2235.

[31] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors, "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python," *Nature Methods*, vol. 17, pp. 261–272, 2020.

[32] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.

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

[34] T. Fischer, W. Vollprecht, S. Traversaro, S. Yen, C. Herrero, and M. Milford, "A robostack tutorial: Using the robot operating system alongside the conda and jupyter data science ecosystems," *IEEE Robot. and Automat. Mag.*, 2021.

[35] S. Garrido-Jurado, R. Muñoz-Salinas, F. Madrid-Cuevas, and M. Marín-Jiménez, "Automatic generation and detection of highly reliable fiducial markers under occlusion," *Pattern Recognit.*, vol. 47, no. 6, pp. 2280–2292, 2014.

[36] G. Villatte, A.-S. Muller, B. Pereira, A. Mulliez, P. Reilly, and R. Emery, "Use of Patient-Specific Instrumentation (PSI) for glenoid component positioning in shoulder arthroplasty. A systematic review and meta-analysis," *PLOS ONE*, vol. 13, no. 8, p. e0201759, Aug 2018.

[37] A. Greene, M. Hamilton, S. Polakovic, N. Mohajer, A. Youderian, T. Wright, I. Parsons, P. Saadi, E. Cheung, and R. Jones, "Navigated versus non-navigated results of a ct-based computer-assisted shoulder arthroplasty system in 30 cadavers," in *Orthopaedic Proc.*, vol. 101-B, no. SUPP_5, 2019, pp. 23–23.

[38] P. Kriechling, R. Loucas, M. Loucas, F. Casari, P. Fürnstahl, and K. Wieser, "Augmented reality through head-mounted display for navigation of baseplate component placement in reverse total shoulder arthroplasty: a cadaveric study," *Archives of Orthopaedic and Trauma Surgery*, no. 0123456789, Jul 2021.

[39] X.-S. Gao, X.-R. Hou, J. Tang, and H.-F. Cheng, "Complete solution classification for the perspective-three-point problem," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 8, pp. 930–943, 2003.

[40] T. Ke and S. I. Roumeliotis, "An efficient algebraic solution to the perspective-three-point problem," in *2017 IEEE Conf. on Comput. Vision and Pattern Recognit. (CVPR)*, 2017, pp. 4618–4626.