2024 IEEE International Conference on Robotics and Automation (ICRA 2024)
May 13-17, 2024. Yokohama, Japan

Iterative PnP and its application in 3D-2D vascular image registration for robot navigation

Jingwei Song, Keke Yang, Zheng Zhang, Meng Li, Tuoyu Cao, and Maani Ghaffari

Abstract— **This paper reports on a new real-time robot-centered 3D-2D vascular image alignment algorithm, which is robust to outliers and can align nonrigid shapes. Few works have managed to achieve both real-time and accurate performance for vascular intervention robots. This work bridges high-accuracy 3D-2D registration techniques and computational efficiency requirements in intervention robot applications. We categorize centerline-based vascular 3D-2D image registration problems as an iterative Perspective-n-Point (PnP) problem and propose using the Levenberg-Marquardt solver on the Lie manifold. Then, the recently developed Reproducing Kernel Hilbert Space (RKHS) algorithm is introduced to overcome the "big-to-small" problem in typical robotic scenarios. Finally, an iterative reweighted least squares is applied to solve RKHS-based formulation efficiently. Experiments indicate that the proposed algorithm processes registration over 50 Hz (rigid) and 20 Hz (nonrigid) and obtains competing registration accuracy similar to other works. Results indicate that our Iterative PnP is suitable for future vascular intervention robot applications.**

I. Introduction

Endovascular Image-Guided Interventions (EIGIs) is the process of inserting a guidewire/catheter into the femoral artery, which is then threaded under fluoroscopic and angiographic guidance through the vasculature to the site of the pathology to be treated [1]. Unlike conventional invasive surgeries, EIGIs play an important role in the modern minimally invasive treatment of heart disease, cancer, and stroke or neurovascular disease, thanks to their lower mortality rate and faster recovery. The minimally invasive EIGIs enable surgeons to deliver drugs, embolic material, or devices through the narrow endovascular. In typical EIGIs, rapid-sequential 2D fluoroscopic X-ray imaging is usually heavily relied on for guidance. In addition to the 2D fluoroscopic and angiographic images, pre-operative 3D Digitally Subtracted Angiograms (3D-DSA) or Computed Tomography Angiograms (3D-CTA) can be obtained for surgeons' 3D spatial perception.

Although EIGIs are gaining popularity, they cause serious 3D perception problems for surgeons. Surgeons are required to operate in the endovascular with the provided real-time 2D images and without direct 3D visualization. As [2] points out, surgeons mentally align the interventional tools and blood flow characteristics from real-time 2D images with the displayed 3D vascular morphology. Furthermore, deformation caused by breathing, heart beating, or even patient movement brings more difficulty for surgeons [3]. Thus, extensive research has been conducted in studying automatic alignment algorithms for 3D-2D vascular images by considering 6 Degrees-of-Freedom (DoF) rigid transformation and further nonrigid motion.

3D-2D image alignment methods can be categorized as model-driven and data-driven approaches. Model-driven approaches use hand-crafted similarity metrics. [2]'s comprehensive review categorizes model-driven approaches as intensity-based, feature-based, and gradient-based. These prior-free approaches adopt invariant features in both 3D and 2D images to define the metric, like intensity similarity-based, vascular centerline geometry, image gradient-based, or a hybrid. Data-driven approaches adopt Deep Neural Networks (DNN), which can encode prior information implicitly. Existing data-driven methods [4], [5], [6] model the entire procedure by predicting the rigid spatial transformation from 3D and 2D pairs in an end-to-end manner. Early works train DNN to predict the transformation in one time [7] and following works uses DNN to incrementally align 3D and 2D shapes [6], [8].

Although recent 3D-2D alignment approaches have progressed greatly, their applications in intervention robots remain challenging. First, outliers and incomplete vessel shapes pose great difficulty in alignment. Multi-modal 3D and 2D vascular images are converted to an invariant space for similarity measurement, and the conversion brings outliers. Furthermore, 2D images normally cover a small Field of View (FoV) in reference to the pre-operative 3D images. The missing/redundant features cause partial overlap, matching a big 3D shape to a small 2D shape (**"big-to-small"**). Existing model-driven approaches with Euclidean losses are especially sensitive to outliers. Secondly, most existing model-based approaches adopt derivative-free optimizers like Powell or Nelder-Mead algorithms. Even though derivative-free optimizers are less sensitive to local minima, they require far more iterations than derivative-based approaches. Recent deep Reinforcement Learning (RL)-based methods [4], [6] consume even more computation in their repetitive inference. Lastly, nonrigid vessels bring more obstacles to 3D-2D alignment. Model-driven approaches [9], [10], [11] formulate the alignment with much larger nonrigid state space and require heavy computation in the minimization process. Existing nonrigid data-driven approaches [12], [13], [14] are not efficient in real-time implementation.

In this paper, we aim to achieve the first robot-centered 3D-2D vascular image alignment algorithm, which **runs in**

J. Song is with United Imaging Research Institute of Intelligent Imaging, Beijing 100144, China jingweisong.eng@outlook.com

K. Yang, Z. Zhang, M. Li, and T. Cao are with United Imaging, Shanghai, China. {keke.yang, zheng.zhang, meng.li02, tuoyu.cao}@united-imaging.com.

M. Ghaffari is with the University of Michigan, Ann Arbor, MI 48109, USA. maanigj@umich.edu.

real-time, is robust to outliers, and can align shapes in a nonrigid manner. We first categorize the 3D-2D aligning problem as an Iterative Perspective-n-Point (PnP) [15], [16], then adopt 2D Reproducing Kernel Hilbert Space (RKHS) loss to handle outliers. Considering that a typical 2nd-order derivative solver cannot optimize RKHS loss, an Iterative Reweighted Least Squares (IRLS) [17] combined with Levenberg–Marquardt is applied in solving RKHS formulation. It substitutes conventional derivative-free optimizers with a 2nd-order optimizer, leading to a reduced number of iterations. Moreover, experiments indicate that RKHS loss is more robust to outliers than Euclidean loss, especially in the **"big-to-small"** case.

In particular, the contributions of this work are as follows.

1) 3D-2D vascular image alignment is formulated as an Iterative PnP with RKHS loss. The formulation shows increased robustness to outliers compared with conventional Euclidean loss.
2) This is the first work adopting an exact 2nd-order derivative-based optimizer for real-time alignment. IRLS is applied to minimize RKHS loss as the 2nd-order optimizer.
3) Moreover, this work combined the nonrigid and RKHS formulations into a joint framework.

II. Literature review

3D-2D alignment algorithms can be classified as model-driven and data-driven. Both are based on the assumption that the projected 3D image should be the most similar to the target 2D image with correct intrinsic projection matrix, pose, and optional nonrigid parameter [18].

Model-driven approaches formulate the problem by estimating the sensor pose and optional nonrigid parameter with the extracted handcrafted invariant features. According to [2], model-driven approaches consist of intensity-based, feature-based, and gradient-based approaches. Intensity-based algorithms project the 3D image with Digitally Reconstructed Radiographs (DRRs) [19] or Maximum Intensity Projection (MIP) [20] and measure the image similarity based on metrics like normalized cross-correlation. Intensity-based algorithms require heavy computation in the iterative 3D to 2D projection process in the full image domain [2]. Feature-based approaches extract invariant semantic features like centerline, point orientation, and bifurcation. [21] uses the Gaussian kernel function and estimates the 6 DoF transformation by maximizing the sum of kernels in 2D. [22] extends the rigid matching to nonrigid matching and defines the loss as a 2D Euclidean sum. [9] minimizes the projection error of 2D vessel tree skeleton in a two-step process (rigid and nonrigid). Gradient-based methods build the metric by exploiting the fact that the 2D gradient and 3D projected gradient are perpendicular on the aligned vessel. Experiments in [2] validate that intensity-based methods are the most accurate, while others are with similar accuracy. Three algorithms [9], [23], [24] achieve around 0.5s while the rest requires over 10s. [22] adopted the derivative-based algorithm Broyden–Fletcher–Goldfarb–Shanno (BFGS) as the solver. [23] used a derivative-free algorithm (Powell) but implemented on GPU. Although not reported, [24] is expected to be slow because it is based on *genetic algorithm*.

Data-driven algorithms learn the similarity matrix from the training data and predict the optimal transformation in an end-to-end manner. The network is trained on a simulated predicted 2D data set from the pre-operative 3D data set and applied to the real data set. Early works [4], [7] used multi-range networks for coarse-to-fine sensor pose estimation. Each network covers a certain range of the pose. [13] proposes a multi-channel convolution DNN that integrates multiple phases caused by the breathing and heartbeat of patients. Its inference step consumes 4 milliseconds (ms), thanks to the small size of the neural network. The following studies used RL to align shapes iteratively. A simplified Q-learning, based on the Markov decision process, was applied in a step-by-step manner [6], [8]. [12] adopted transfer learning to bridge the domain gap between training and testing data sets. Although DNN-based methods have been reported to be accurate, most of them require long training and inference time, and some are difficult to reproduce.

According to our knowledge, only [22] adopted a derivative-based BFGS solver in optimal pose searching procedure. Derivative-based searching, especially 2nd-order derivative algorithms, guarantees real-time searching in a fixed number of iterations. This research follows [9]'s roadmap by aligning 3D-2D images' vascular centerlines and adapts it to EIGIs scenario.

III. Methodology

This work first conducts 3D and 2D blood vessel centerline segmentation and then estimates the optimal rigid and nonrigid transformation for 3D and 2D vessel alignment. We will reveal that 3D-2D centerline matching is a special case of a well-researched topic, that is, PnP, in the computer vision community. Several state-of-the-art strategies are coupled to realize a real-time robust 3D to 2D vessel alignment. The proposed workflow ensures robust and prior-free methods for intervention robots.

A. Vessels segmentation

This research chose vascular-based data alignment for intervention robot applications for two reasons. First, blood vessel centerlines are invariant and observable tissues in both 3D and 2D images. The centerline matching procedure is always differentiable, while intensity-based and gradient-based alignments are not always differentiable if the initial alignment differs greatly. Second, 3D and 2D vessel centerline segmentation algorithms are heavily studied and are robust for robotic applications.

Widely used Otsu's method [25] was applied for fast 2D vessel segmentation. DNN-based method from commercial software provided by United Imaging of Health Co., Ltd is used for 3D vessel segmentation. Other prior-based DNNs [26], [27] are also applicable for 3D/2D image segmentation.

B. Iterative PnP formulation

After vessel segmentation, the 3D-2D registration algorithm can be summarized as an approximate Iterative PnP. PnP is the algorithm for estimating the optimal 6 DoF sensor pose of a set of 3D points given the calibrated projection matrix and 3D-2D correspondences that align projected 3D points and 2D points [28]. To realize the algorithm, correspondences should be established based on the closest projections search given the initial sensor pose. Thus, iterative PnP differs from authentic PnP, which associates points based on feature points similarity. Data association process in the iterative PnP is similar to the Iterative Closest Point (ICP) algorithm. Another difference from authentic PnP is that vessels may deform in surgeries like percutaneous coronary intervention (mainly from heartbeat) and transarterial chemoembolization (mainly from breath). Therefore, nonrigid parameterization is an option in addition to rigid pose estimation.

PnP is a fundamental tool [15], [16] for searching the optimal camera (sensor) pose $\mathbf{T}^* \in \mathrm{SE}(3)$ with know 3D point $\mathbf{p}_i \in \mathbb{R}^4$ (Ω is the set of source point index) and 2D point $\mathbf{q}_i \in \mathbb{R}^3$ (both $\mathbf{p}_i$ and $\mathbf{q}_i$ are in homogeneous coordinate) alignment. Thus, the correspondence $\mathbf{p}_i$ and $\mathbf{q}_i$ are aligned point pair. Authentic PnP is formulated as:

$$\mathbf{T}^* = \mathrm{argmin}_{\mathbf{T}} \sum_{i \in \Omega} ||\pi(\mathbf{T}\mathbf{p}_i, \mathbf{K}) - \mathbf{q}_i||_2^2, \tag{1}$$

where $\mathbf{K} \in \mathbb{R}^{3\times4}$ is the camera intrinsic matrix and $\pi(\mathbf{T}\mathbf{p}_i, \mathbf{K})$ is the 3D to 2D pin-hole camera projection function defined as:

$$\pi(\mathbf{T}\mathbf{p}_i, \mathbf{K}) = \mathbf{K}\mathbf{T}\mathbf{p}_i. \tag{2}$$

Similar to (1), [9] defines 2D points' Euclidean distance sum as the loss function and obtain optimal sensor pose as

$$\mathbf{T}^* = \mathrm{argmin}_{\mathbf{T}} \sum_{i \in \Omega} ||\pi(\mathbf{T}\mathbf{T}_{dsa\_cb}(\mathbf{p}_i - \mathbf{c}), \mathbf{K}) - \mathbf{q}_j||_2^2, \tag{3}$$

where $\mathbf{T}_{dsa\_cb} \in \mathrm{SE}(3)$ and $\mathbf{c} \in \mathbb{R}^4$ (homogeneous coordinate) are the pre-calibrated transformation matrices from pre-operative coordinate (CT for example) to intra-operative coordinate (DSA for example), $\mathbf{p}_i$ is the 3D source point in homogeneous coordinate and $\mathbf{q}_j$ is the corresponding 2D target point in homogeneous coordinate by closest distance searching. Instead of known correspondences, the temporal correspondences between groups $\{\mathbf{p}_i\}$ and $\{\mathbf{q}_j\}$ are determined by point-to-point closest Euclidean distance searching similar to ICP. [9] relaxed the rotation in $\mathbf{T}$ as an affine matrix for modelling simple deformation.

We define (1) with unknown $\mathbf{p}_i$ and $\mathbf{q}_i$ pairs (or $\mathbf{q}_j$ to be specific) as **Iterative PnP**. Robotic and computer vision communities seldom handle this scenario because the correspondences can be obtained with well-known image corner points matching algorithms. It only happens in special cases like vascular image registration. In contrast to PnP, Iterative PnP brings two issues: more outliers and more local minima. The following RKHS strategy handles outliers well.

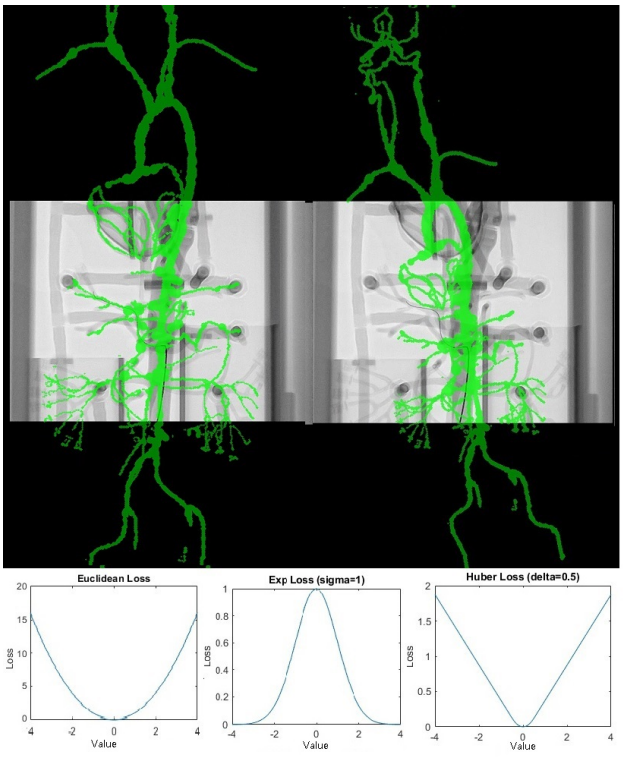

Fig. 1: The upper two figures show the typical **"big-to-small"** alignment. The green shape is the projection of the pre-operative 3D vessel, which is in large size. The upper left figure is the proposed Iterative PnP on RKHS space, which is robust to **"big-to-small"** alignment. The upper right figure is our implementation of [9], which uses the Huber loss. The bottom figures are three loss functions. Only the "Exp" (stands for exponential) loss manages to lower the impacts of "outlier" points.

Regarding local minima issue, our experiments show that obtaining the correct pose is difficult, but the registration task remains unaffected.

C. Outlier rejection and RKHS

Equation (3) is the widely used L2 norm loss for authentic PnP and ICP problems and works well in many situations. Although it can be solved with Gauss-Newton or quasi-Newton solver, it suffers from a serious issue widely known as *partially overlap*. Specifically, the pre-operative 3D blood vessel is much larger than the intra-operative 2D blood vessel, as **"big-to-small"** in this article. A massive number of unpaired source points $\mathbf{p}_i$ are wrongly matched to target points $\mathbf{q}_i$. Since the wrong pairs have large distances, the sum of errors is large and brings wrong derivative weighting in searching for optimal $\mathbf{T}^*$. Although M-estimators [29] reduce the influence of large distance pairs, the matching is still heavily influenced by outliers. Conventional 3D-3D ICP can circumvent this issue by reversing source and target point clouds (that is "small-to-big"). In 3D-2D matching, however, reversing is impossible because the key accelerating ingredient, the K-D tree-based correspondence searching, should only be built on the static 2D vessel. As Fig. 1 shows, many points outside of the FoV are outliers and Euclidean loss in (3) suffers from large outliers (upper right figure) and then wrong derivatives. Huber loss only manages to reduce the Binomial distance-to-loss mapping to linear mapping.

Gaussian kernels or Gaussian Mixture Model (GMM) [30], [31], [32] overcomes the partial overlapping problem (**"big-to-small"** in 3D-2D vessel registration case) based on its special distance-to-loss mapping manner. [21] is the first work applying 3D-2D vessel registration based on Gaussian kernel. This work uses a Gaussian kernel to measure alignment in 2D space and demonstrates its efficacy. During our tests, we also noticed that Gaussian kernel [21] brings better accuracy and robustness to 3D-2D blood vessel matching. Surprisingly, most centerline-based studies ignore it and use Euclidean loss instead. One possible explanation is that Gaussian kernel cannot be transformed into authentic least-square form, and real-time applications like [9] have difficulty feeding Gaussian kernel loss to BFGS solver.

Continuous Visual Odometry (CVO) [33] is the first work applying Reproducing Kernel Hilbert Space (RKHS) theory in ICP problem. Different from [30] using discrete Gaussian kernel space, CVO [33] enforces continuous GMM field and adopts 4th order Taylor expansion to retrieve an optimum step for the first-order derivative-based gradient descend. It achieves 2-5 Hz on modern CPU with parallel computation and for typical LiDAR-based point cloud registration. [34] further reduces CVO's time consumption with GPU acceleration. Following [34], this article modifies RKHS [33] and obtain optimal parameters $\mathbf{T}^*, \theta^*$ by jointly optimizing

$$
\begin{aligned}
&\max_{\mathbf{T},\theta} E_{data}, \text{ such that}, \min_{\mathbf{T},\theta} \lambda_1 E_{init} + \lambda_2 E_{reg} \\
&E_{data} = \\
&\sum_{i\in\Omega} \mathrm{w}_i \sum_{j\in\Delta(i)} \exp\left(-\frac{\|\pi(\mathbf{T}\Gamma(\mathbf{T}_{dsa\_cb}\mathbf{p}_i - \mathbf{c}, \theta), \mathbf{K}) - \mathbf{q}_j\|_2^2}{2\ell^2}\right) \\
&E_{init} = \|\log(\mathbf{T}_0^{-1}\mathbf{T})^\vee\|_2^2 \\
&E_{reg} = \mathrm{w}_1 \sum_{i\in\Omega} \|\mathbf{r}_i\|_2^2 + \mathrm{w}_2 \sum_{i\in\Omega} \sum_{j\in\{i-1,i+1\}} \|\mathbf{r}_i - \mathbf{r}_j\|_2^2 \\
&\qquad + \mathrm{w}_3 \sum_{i\in\Omega} \sum_{j\in\Omega_i} \|\mathbf{r}_i - \mathbf{r}_j\|_2^2,
\end{aligned}
\tag{4}
$$

where w_i is the weight and uniformly set as 1 in this research, $\Delta(i)$ is the set of correspondences to $\mathbf{p}_i$, $\mathbf{T}_0$ is the initial pose or pose in previous state and ℓ, λ_1, λ_2, w_1, w_2, w_3 are hyperparameters, and θ is the deformation parameter. Ω is the set of points' index set of $\mathbf{p}_i$. $(\cdot)^\vee$ convert Lie algebra $\mathfrak{se}(3)$ to 6-vector $\mathbb{R}^6$ and defines the distance on $\mathrm{SE}(3)$ manifold. It should be pointed out that E_{init} is adopted because $\mathbf{T}^*$'s z-direction may increases enormously, and the 3D shape's projection shrinks to a point and matches to an arbitrary target point. This phenomenon represents one scenario of outlier and thus should be handled with E_{init}. The scale parameter ℓ is defined as the maximum $||\pi(\mathbf{T}\Gamma(\mathbf{T}_{dsa\_cb}\mathbf{p}_i - \mathbf{c}, \theta)) - \mathbf{q}_j, \mathbf{K}||_2^2$ and shrinked by half every 5 iterations as coarse-to-fine registration. Please be noted that (4) suffers from constraint $\min_{\mathbf{T},\theta} \lambda_1 E_{init} + \lambda_2 E_{reg}$ and only has Pareto optimum. Next section adopted IRLS [17] for best gradient searching and energy-constraint balancing.

Moreover, the nonrigid function is defined as $\Gamma(\cdot, \theta) = \cdot$ in rigid cases. And λ_2 should be set to zero. For nonrigid Iterative PnP, we follow [9] by realizing $\Gamma(\mathbf{p}_i, \theta)$ as:

$$
\Gamma(\mathbf{p}_i, \theta) = \mathbf{p}_i + \mathbf{r}_i, \text{ such that } \mathbf{r}_i \in \theta, \tag{5}
$$

where $\theta = \{\mathbf{r}_i | i \in \Omega\}$ and $\mathbf{r}_i \in \mathbb{R}^4$. (4) differs from [21] as the nonrigid deformation field, 2D RKHS space supported by $\Delta(i)$, initial sensor pose $\mathbf{T}_0$ regularization and approximate 2nd-order derivative fastening covered in next section.

D. Approximate 2nd-order gradient for RKHS space

Real-time implementation is the key requirement for intervention robot applications. Most studies focus on achieving high accuracy instead of computational efficiency and thus adopt derivative-free easy-to-use solvers like the Powell or Nelder-Mead method. Only [22] adopts a 2nd-derivative-based solver BFGS for fast optimization. This work attempts to solve the original multi-objective function (4) by vectorizing the original problem in an IRLS manner. Specifically, considering that Gaussian kernel cannot be written in binomial form and thus least squares, [33] uses first-order gradient descent and retrieve the optimal step by approximating Gaussian kernel as 4th order Taylor expansion. However, their experiment indicates that their accelerated first-order gradient descent still requires a massive amount of iteration. This work follows the latest progress [33] and adopts the IRLS method in optimization. In the iterative optimization of (4), the searching direction and step size of $\mathbf{T}^n, \theta^n$ in step n is determined by minimizing

$$
\begin{aligned}
&\mathbf{T}^{n+1}, \theta^{n+1} = \min_{\mathbf{T},\theta} \sum_{i\in\Omega} \mathrm{w}_i \sum_{j\in\Delta(i)} \mathrm{w}_{ij}^n \|\pi(\mathbf{T} \\
&\cdot (\Gamma(\mathbf{T}_{dsa\_cb}\mathbf{p}_i - \mathbf{c}, \theta)) - \mathbf{q}_j\|_2^2 + \lambda_1 \mathrm{E}_{init} + \lambda_2 \mathrm{E}_{reg},
\end{aligned}
\tag{6}
$$

where $\mathrm{w}_{ij}^n = \exp\left(-\frac{\|\pi(\mathbf{T}^n\Gamma(\mathbf{T}_{dsa\_cb}\mathbf{p}_i - \mathbf{c}, \theta^n), \mathbf{K}) - \mathbf{q}_j\|_2^2}{2\ell^2}\right)$ and fixed as the weight in the iterative searching. $\mathbf{T}^n$ and θ^n are the state estimated in step n and indifferentiable in w_{ij}^n. $\mathbf{T}$ and θ in (6) are initialized with $\mathbf{T}^n$ and θ^n. **It should be emphasized that although the searching direction and step are determined by local descending of** (6) **in each step, the main objective is to maximize the sum of RKHS** E_{data} **in** (4)**.** In practice, the direction and step size in (6) is searched analytically in least square form.

E. Technical details

Following [9], 6 DoF rigid pose $\mathbf{T}^*$ in (4) is first estimated, and nonrigid field θ^* is then estimated for refinement. Moreover, rigid transformation $\mathbf{T}$ is incrementally optimized on $\mathrm{SE}(3)$ manifold [35] in contrast to Euler angles and $\mathbb{R}^3$ translation in existing reseaches. Optimization on continuous $\mathrm{SE}(3)$ manifold avoids gimbal lock degeneration on Euler angles space by following geodesics along the manifold for search. Levenberg-Marquardt is adopted as the minimizer.

IV. Experiments

A. Experimental Setup

The proposed algorithm was tested on a simulation data set, an *ex-vivo* data set, and an *in-vivo* data set[1]. The

[1]We strongly recommend readers to watch the attached video for the complete *in-vivo* and *ex-vivo* experiments.

simulation data set was generated by enforcing a simulated $\mathrm{SE}(3)$ transformation on segmented 3D vessel centerlines. The chosen algorithm aligned it with its original projection without the simulated $\mathrm{SE}(3)$ transformation. As [9] points out, testing on simulations characterizes the performance of the various algorithms in an ideal scenario with known sensor pose and without noises. The adopted *in-vivo* data sets of 6 patients are P1 (44), P2 (60), P3 (39), P4 (53), P5 (45), P6 (37), where the number in bracket records the size of images. An *ex-vivo* data set was also generated by scanning a phantom using United Imaging's uAngio 960 devices. 3D pre-operative vessel centerlines were obtained by segmenting a pre-operative CTA image. Then, live images were collected for real-time alignment.

The proposed algorithm was compared with widely accepted prior-free approaches AutoMask [36], DT-ICP [9], RGRB [37] and Normalized ICP (NICP) [21]. The original DT-ICP [9] uses BFGS as the 2nd-order solver, while our implementation adopts the Levenberg-Marquardt solver on the Lie manifold for better convergence and faster speed. Our experiments show faster speed and less iteration than BFGS. Pair-wise DNN [4] is implemented for comparison. For DNN-based approaches, we admit that other RL-related approaches [6], [8] are extremely difficult to implement and train. We fail to realize these approaches. As [38] pointed out, these complicated DNN-based approaches suffer from "hardness of obtaining the ground-truth transformation parameters" and "real intra-operative DSAs can be incomplete". [38] shifted the research direction back to two-step-based image-to-pose regression.

A commercial laptop ALIENWARE M17 R4 (i7-10870H and 32Gb RAM) is used. Its GPU, GeForce RTX 3060 (6Gb), is used for Pair-wise DNN only. As no open-source code is available, we implemented all these approaches ourselves. AutoMask was implemented on python3 because its derivative-free optimizer, Nelder-Mead, is slow and impossible for real-time performance. RGBR was implemented on Matlab, and its time consumption is meaningless. DT-ICP, NICP, and Iterative PnP were implemented in C++ and wrapped as a package in Robot Operating System (ROS) [39]. During all experiments, both 3D and 2D vessel sizes are within 1500 - 3000 points. The hyperparameters λ_1, λ_2, w_1, w_2 and w_3 are set as 100, 1, 0.1, 10 and 1.

Following previous research, all the algorithms were compared based on Projection Residual (PR), Gross Failure Rate (GFR), and time consumption. PR measures the Root Mean Square Error of all the projected source points. GFR accounts for test cases with PR greater than the selected threshold. Median, 75%th percentile, and 95%th percentile PRs are reported. In quantitative experiments, all points were selected as anchor points. Since the aim is limited to centerline registration as other works do, Otsu-based extraction's time consumption, which is around 400 ms for image size 512×512, is not counted.

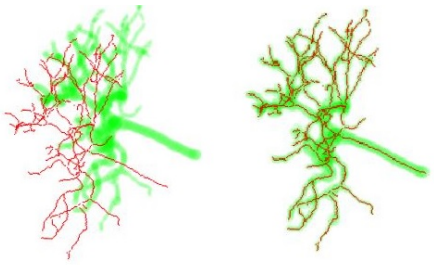

Fig. 2: Presented is a sample synthetic data set. The red curve is the synthetic 2D vascular centerline. The green curve is the transformed and projected 3D shape. The left shape is the projection with a fixed initial pose, and the right shape is projected with a fixed pose.

TABLE I: The table reports the average pose of the simulation data by enforcing noise with angular standard deviation 2° and translation standard deviation $5mm$. GFR refers meanPR greater than 5 pixels.

Method	GFR	MeanPR (mm)	Percentile (mm)		Run time (ms)
			95%	75%	
AutoMask	0%	0.85	1.50	1.06	97004.0
Pair-wise DNN	0%	1.74	2.54	2.02	311.3
RGBR	22%	2.21	3.92	1.71	-
DT-ICP (rigid)	54%	1.80	6.65	3.49	13.2
DT-ICP	0%	0.42	0.72	0.42	39.4
NICP	8%	0.50	3.11	1.14	385.0
Iterative PnP (rigid)	4%	0.39	0.84	0.56	16.9
Iterative PnP	0%	0.37	0.69	0.44	45.5

B. Accuracy and Robustness

The simulation experiment was conducted by sampling rigid pose 100 times with angular standard deviation 2° and translation standard deviation $5mm$. Simulation experiments guarantee one-to-one correspondence without noise. Fig. 2 shows one sample Iterative ICP result of the simulated 2D centerline (red) and 3D projection based on the selected pose. Table I shows that DT-ICP performs similarly as Iterative PnP (rigid) and Iterative PnP in the simulation data sets.

In-vivo and *ex-vivo* experiments were implemented. Fig. 3 visualizes sample comparisons of the registration of the 6 *in-vivo* data sets. Fig. 1 demonstrates that the proposed RKHS loss handles **"big-to-small"** issues well, while DT-ICP fails in the alignment. Fig. 4 shows a sample sequential registration in X-ray-based *ex-vivo* experiments. Visually, the proposed Iterative PnP achieves slightly better registration than competing algorithms. All qualitative results validate that Iterative PnP achieves slightly or even better results than competing methods.

C. Time consumption

Table I and Table II show that both DT-ICP and Iterative PnP manage over 50 Hz (rigid) and 20 Hz (nonrigid). The

TABLE II: The table shows the accuracy of the *in-vivo* data sets. Median PR is the median of PR of all source points. GFR refers meanPR greater than $5mm$.

Method	GFR	Median PR (mm)	Percentile (mm)		Run time (ms)
			95%	75%	
AutoMask	100%	17.01	29.75	23.26	150302.2
Pair-wise DNN	100%	153.25	437.24	168.60	529.0
RGRB	100%	5.99	38.85	12.16	-
DT-ICP	16%	4.68	31.22	9.613	13.2
NICP	56%	3.49	37.09	10.52	385.1
Iterative PnP	11%	3.74	29.92	10.34	16.9

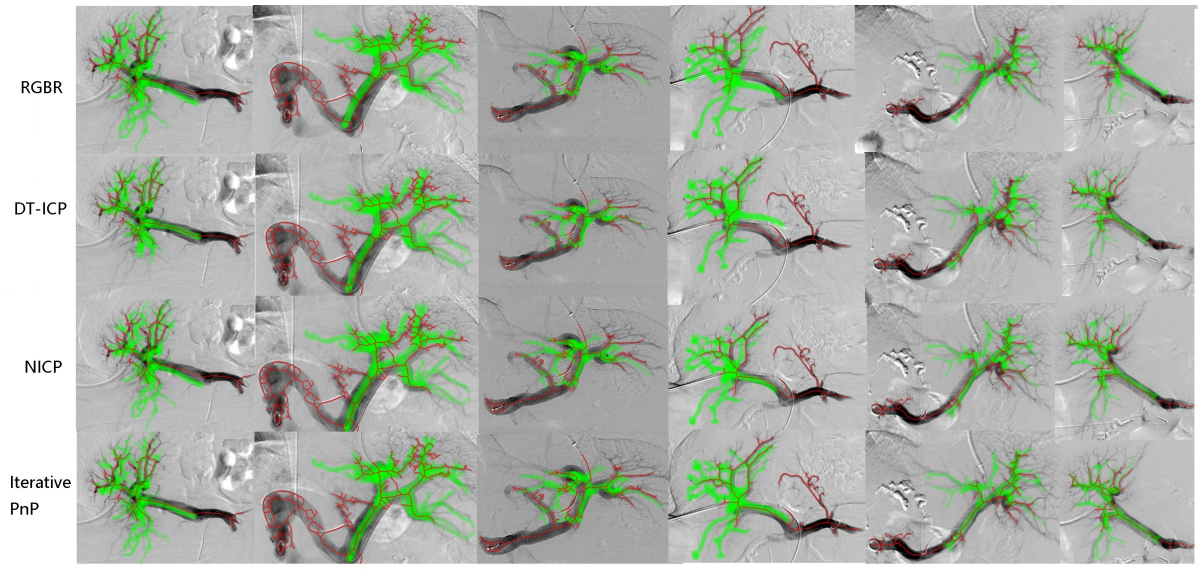

Fig. 3: Illustrated is the alignment comparisons of RGBR, DT-ICP, NICP and Iterative PnP. Green shape is the projection of the pre-operative 3D vessel and red shape is the 2D centerline of the vessel. The 6 columns represents the sample results of the 6 patients.

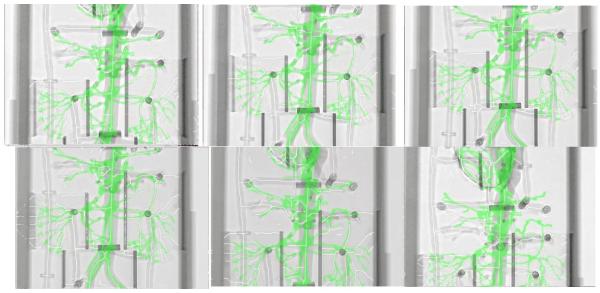

Fig. 4: The figure shows one sample of the *ex-vivo* experiments. Sequential registered 3D-2D phantom shapes masked on X-ray images are demonstrated. The green shapes are the projected 3D vessels. Other *ex-vivo* experiments can be found in the attached video.

original DT-ICP [9] reports around 2 Hz. Our implementation with Levenberg–Marquardt solver on Lie manifold requires much less iteration than their BFGS. Other reasons include computational devices and data set differences. Although the proposed Iterative PnP is around 25% slower than DT-ICP, it achieves higher accuracy and better robustness.

D. Local minima in Pose Estimation

It is interesting to reveal that Iterative PnP suffers from heavy local minima, making the estimated pose incorrect in most experiments. Table III shows the relationship between the sensor pose disturbance level and pose estimation accuracy of our proposed algorithm. It indicates that robot pose can be estimated correctly with fewer local minima if the initial pose is good (small pose disturbance in Table III). Table III also suggests that point-wise registration is still successful even though the estimated pose is wrong. A test on 3D-3D (with depth of target data provided) registration of RKHS formulation [33] shows that RKHS accurately estimates pose. Therefore, we hypothesize that Iterative PnP gains an extra amount of local minima and should be credited to the unobservable depth for 2D vascular centerlines compared with the notorious nonconvex ICP process. This conclusion should be further validated.

TABLE III: The table reports the average pose error of the simulation data with different disturbances, which is in different levels of Gaussian noises ("Std" refers to standard deviation). MeanPR and GFR shows that registration is still satisfying. Pose errors indicate that the algorithm fails. 1 pixel is equivalent to 0.30 mm. "Ang Std" refers angular standard deviation (in $^\circ$). "Trans Std" refers translation standard deviation (in mm). GFR refers MeanPR greater than 5 pixels.

Ang Std	Trans Std	MeanPR (mm)	GFR	Average Pose Error
0.50	1.00	0.10	0.0%	0.10°, $1.81mm$
1.00	3.00	0.28	0.0%	0.29°, $5.31mm$
2.00	5.00	0.43	0.0%	0.46°, $7.09mm$
3.00	8.00	0.82	0.0%	0.68°, $12.48mm$
5.00	10.00	1.51	21.0%	3.40°, $14.98mm$

E. Limitations and Future Works

The major issue of Iterative PnP is that its accuracy and robustness are heavily affected by the objective function's nonconvexity. Section IV-D reveals that Iterative PnP is not suitable for sensor pose estimation. Future works can be devoted to overcoming local minima with methods like Branch-and-Bound strategy [40] or gradual relaxation [41]. Better convergence not only estimates 6 DoF pose correctly but also provides better registration (median PR shown in Table III).

V. Conclusion

This research proposes an Iterative PnP algorithm for robot navigation tasks in EIGIs. Experiments show that the proposed algorithm achieves real-time alignment over 20 Hz (nonrigid) or 50 Hz (rigid) based on modern laptops. The proposed algorithm achieves similar accuracy and is robust to outliers compared with existing works, especially in typical **"big-to-small"** vascular centerline registration scenarios. Considering that our Iterative PnP is more robust, slightly more accurate, and computationally efficient, our proposed method is suitable for future EIGIs and intervention robot applications. This work is in the process of commercialization by the company United Imaging of Health Co., Ltd.

REFERENCES

[1] S. Rudin, D. R. Bednarek, and K. R. Hoffmann, "Endovascular image-guided interventions (EIGIs)," *Medical physics*, vol. 35, no. 1, pp. 301–309, 2008.

[2] U. Mitrović, B. Likar, F. Pernuš, and Ž. Špiclin, "3D–2D registration in endovascular image-guided surgery: Evaluation of state-of-the-art methods on cerebral angiograms," *Int. Jour. for Comp Assist. Rad. and Sur.*, vol. 13, no. 2, pp. 193–202, 2018.

[3] K. Yang, Z. Zhang, M. Li, T. Cao, M. Ghaffari, and J. Song, "Optical flow-based vascular respiratory motion compensation," *IEEE Robot. and Autom. Lett.*, pp. 1–8, 2023.

[4] J. Zheng, S. Miao, Z. Jane Wang, and R. Liao, "Pairwise domain adaptation module for CNN-based 2-D/3-D registration," *Journal of Medical Imaging*, vol. 5, no. 2, pp. 021204–021204, 2018.

[5] R. Liao, S. Miao, P. de Tournemire, S. Grbic, A. Kamen, T. Mansi, and D. Comaniciu, "An artificial agent for robust image registration," in *Proc. AAAI Nat. Conf. Artif. Intell.*, vol. 31, 2017.

[6] S. Miao, S. Piat, P. Fischer, A. Tuysuzoglu, P. Mewes, T. Mansi, and R. Liao, "Dilated FCN for multi-agent 2D/3D medical image registration," in *Proc. AAAI Nat. Conf. Artif. Intell.*, vol. 32, 2018.

[7] J. Zheng, S. Miao, and R. Liao, "Learning CNNS with pairwise domain adaption for real-time 6DoF ultrasound transducer detection and tracking from X-ray images," in *Medical Image Computing and Computer-Assisted Intervention- MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part II 20*, pp. 646–654, Springer, 2017.

[8] S. Miao and R. Liao, "Agent-based methods for medical image registration," *Deep learning and convolutional neural networks for medical imaging and clinical informatics*, pp. 323–345, 2019.

[9] D. Rivest-Henault, H. Sundar, and M. Cheriet, "Nonrigid 2D/3D registration of coronary artery models with live fluoroscopy for guidance of cardiac interventions," *IEEE Trans. Med. Imag.*, vol. 31, no. 8, pp. 1557–1572, 2012.

[10] C.-R. Chou and S. Pizer, "Real-time 2D/3D deformable registration using metric learning," in *International MICCAI Workshop on Medical Computer Vision*, pp. 1–10, Springer, 2012.

[11] Q. Zhang, A. Pevsner, A. Hertanto, Y.-C. Hu, K. E. Rosenzweig, C. C. Ling, and G. S. Mageras, "A patient-specific respiratory model of anatomical motion for radiation treatment planning," *Medical physics*, vol. 34, no. 12, pp. 4772–4781, 2007.

[12] S. Guan, T. Wang, K. Sun, and C. Meng, "Transfer learning for nonrigid 2D/3D cardiovascular images registration," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 9, pp. 3300–3309, 2020.

[13] S. Guan, C. Meng, Y. Xie, Q. Wang, K. Sun, and T. Wang, "Deformable cardiovascular image registration via multi-channel convolutional neural network," *IEEE Access*, vol. 7, pp. 17524–17534, 2019.

[14] M. Nakao, M. Nakamura, and T. Matsuda, "Image-to-graph convolutional network for 2D/3D deformable model registration of low-contrast organs," *IEEE Trans. Med. Imag.*, vol. 41, no. 12, pp. 3747–3761, 2022.

[15] B. M. Haralick, C.-N. Lee, K. Ottenberg, and M. Nölle, "Review and analysis of solutions of the three point perspective pose estimation problem," *International journal of computer vision*, vol. 13, pp. 331–356, 1994.

[16] V. Lepetit, F. Moreno-Noguer, and P. Fua, "Epnp: An accurate o(n) solution to the pnp problem," *International Journal Of Computer Vision*, vol. 81, pp. 155–166, 2009.

[17] R. Zhang, J. Song, X. Gao, J. Wu, T. Liu, J. Zhang, R. Eustice, and M. Ghaffari, "RKHS-BA: A semantic correspondence-free multi-view registration framework with global tracking," *arXiv preprint arXiv:2403.01254*, 2024.

[18] A. Schmidt, O. Mohareri, S. DiMaio, M. C. Yip, and S. E. Salcudean, "Tracking and mapping in medical computer vision: A review," *Medical Image Analysis*, p. 103131, 2024.

[19] J. H. Hipwell, G. P. Penney, R. A. McLaughlin, K. Rhode, P. Summers, T. C. Cox, J. V. Byrne, J. A. Noble, and D. J. Hawkes, "Intensity-based 2-D-3-D registration of cerebral angiograms," *IEEE Trans. Med. Imag.*, vol. 22, no. 11, pp. 1417–1426, 2003.

[20] E. Kerrien, M.-O. Berger, E. Maurincomme, L. Launay, R. Vaillant, and L. Picard, "Fully automatic 3D/2D subtracted angiography registration," in *Int. Conf. on Med. Image Comput. and Comput. Assist. Interv.*, pp. 664–671, Springer, 1999.

[21] S. R. Aylward, J. Jomier, S. Weeks, and E. Bullitt, "Registration and analysis of vascular images," *Int. J. Comput. Vis.*, vol. 55, no. 2, pp. 123–138, 2003.

[22] M. Groher, D. Zikic, and N. Navab, "Deformable 2D-3D registration of vascular structures in a one view scenario," *IEEE Trans. Med. Imag.*, vol. 28, no. 6, pp. 847–860, 2009.

[23] U. Mitrovic, Z. Spiclin, B. Likar, and F. Pernus, "3D-2D registration of cerebral angiograms: A method and evaluation on clinical images," *IEEE Trans. Med. Imag.*, vol. 8, no. 32, pp. 1550–1563, 2013.

[24] J. Jomier, E. Bullitt, M. V. Horn, C. Pathak, and S. R. Aylward, "3D/2D model-to-image registration applied to TIPS surgery," in *Int. Conf. on Med. Image Comput. and Comput. Assist. Interv.*, pp. 662–669, Springer, 2006.

[25] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE transactions on systems, man, and cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.

[26] Y. Li, Y. Zhang, W. Cui, B. Lei, X. Kuang, and T. Zhang, "Dual encoder-based dynamic-channel graph convolutional network with edge enhancement for retinal vessel segmentation," *IEEE Trans. Med. Imag.*, vol. 41, no. 8, pp. 1975–1989, 2022.

[27] S. Moccia, E. De Momi, S. El Hadji, and L. S. Mattos, "Blood vessel segmentation algorithms—review of methods, datasets and evaluation metrics," *Computer methods and programs in biomedicine*, vol. 158, pp. 71–91, 2018.

[28] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[29] R. A. Maronna, "Robust M-estimators of multivariate location and scatter," *The annals of statistics*, pp. 51–67, 1976.

[30] P. Biber and W. Straßer, "The normal distributions transform: A new approach to laser scan matching," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots and Syst.*, vol. 3, pp. 2743–2748, IEEE, 2003.

[31] M. Magnusson, A. Lilienthal, and T. Duckett, "Scan registration for autonomous mining vehicles using 3D-NDT," *J. Field Robot.*, vol. 24, no. 10, pp. 803–827, 2007.

[32] B. Jian and B. C. Vemuri, "Robust point set registration using gaussian mixture models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 8, pp. 1633–1645, 2010.

[33] W. Clark, M. Ghaffari, and A. Bloch, "Nonparametric continuous sensor registration," *J. Mach. Learning Res.*, vol. 22, no. 1, pp. 12412–12461, 2021.

[34] R. Zhang, T.-Y. Lin, C. E. Lin, S. A. Parkison, W. Clark, J. W. Grizzle, R. M. Eustice, and M. Ghaffari, "A new framework for registration of semantic point clouds from stereo and rgb-d cameras," in *Proc. IEEE Int. Conf. Robot. and Automation*, pp. 12214–12221, IEEE, 2021.

[35] N. Boumal, *An introduction to optimization on smooth manifolds*. Cambridge University Press, 2023.

[36] P. Steininger, M. Neuner, H. Weichenberger, G. Sharp, B. Winey, G. Kametriser, F. Sedlmayer, and H. Deutschmann, "Auto-masked 2D/3D image registration and its validation with clinical cone-beam computed tomography," *Physics in Medicine & Biology*, vol. 57, no. 13, p. 4277, 2012.

[37] P. Markelj, D. Tomazevic, F. Pernus, and B. Likar, "Robust gradient-based 3-D/2-D registration of CT and MR to X-ray images," *IEEE Trans. Med. Imag.*, vol. 27, no. 12, pp. 1704–1714, 2008.

[38] C. Meng, Y. Li, Y. Xu, N. Li, and K. Xia, "A weakly supervised framework for 2D/3D vascular registration oriented to incomplete 2D blood vessels," *IEEE Transactions on Medical Robotics and Bionics*, vol. 4, no. 2, pp. 381–390, 2022.

[39] M. Quigley, K. Conley, B. Gerkey, J. Faust, T. Foote, J. Leibs, R. Wheeler, A. Y. Ng, *et al.*, "ROS: an open-source robot operating system," in *ICRA workshop on open source software*, vol. 3, p. 5, Kobe, Japan, 2009.

[40] C. Olsson, F. Kahl, and M. Oskarsson, "Branch-and-bound methods for euclidean registration problems," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 5, pp. 783–794, 2008.

[41] H. Yang, P. Antonante, V. Tzoumas, and L. Carlone, "Graduated non-convexity for robust spatial perception: From non-minimal solvers to global outlier rejection," *IEEE Robot. and Autom. Lett.*, vol. 5, no. 2, pp. 1127–1134, 2020.