Uncertainty Quantification for Multi-Scan Registration

XIANGRU HUANG, The University of Texas at Austin, USA
ZHENXIAO LIANG, The University of Texas at Austin, USA
QIXING HUANG, The University of Texas at Austin, USA

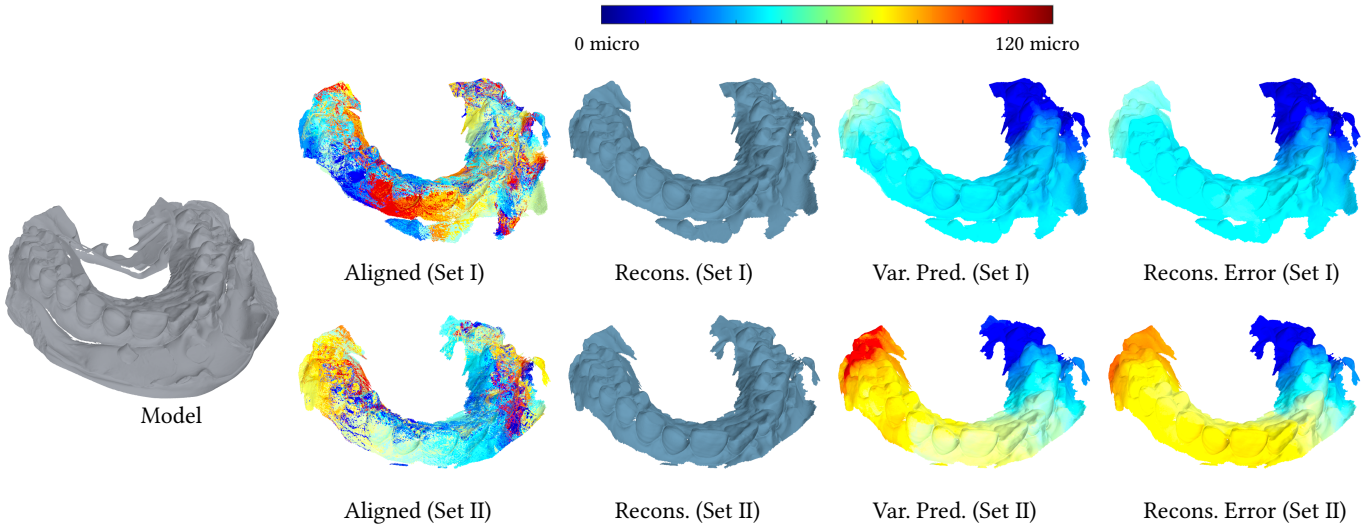

Fig. 1. Two reconstructions from two sets of scans of the same underlying teeth object. Both reconstructions nicely reconstruct local features, yet the bottom reconstruction possesses larger global drifts than the top reconstruction. Such global drifts are hard to detect manually. Our approach computes quantified uncertainties that accurately reveal such global drifts. The predictions are consistent with ground-truth reconstruction errors. From left to right: The ground-truth model obtained by an Artec Micro scanner, the two sets of scans obtained by an Artec Space Spider scanner, the corresponding reconstructions, the predicted uncertainties of the reconstructions, and the ground-truth reconstruction errors.

A fundamental problem in scan-based 3D reconstruction is to align the depth scans under different camera poses into the same coordinate system. While there are abundant algorithms on aligning depth scans, few methods have focused on assessing the quality of a solution. This quality checking problem is vital, as we need to determine whether the current scans are sufficient or not and where to install additional scans to improve the reconstruction. On the other hand, this problem is fundamentally challenging because the underlying ground-truth is generally unavailable, and it is challenging to predict alignment errors such as global drifts manually. In this paper, we introduce a local uncertainty framework for geometric alignment algorithms. Our approach enjoys several appealing properties, such as it does not require re-sampling the input, no need for the underlying ground-truth, informative, and high computational efficiency. We apply this framework to two multi-scan alignment formulations, one minimizes geometric distances between pairs of scans, and another simultaneously aligns the input scans with a deforming model. The output of our approach can be seamlessly integrated with view selection, enabling uncertainty-aware view planning. Experimental results and user studies justify the effectiveness of our approach on both synthetic and real datasets.

CCS Concepts: • **Computing methodologies** → **Point-based models**; **Shape analysis**;

Additional Key Words and Phrases: uncertainty quantification, multi-scan registration, approximation error, view planning

ACM Reference Format:
Xiangru Huang, Zhenxiao Liang, and Qixing Huang. 2020. Uncertainty Quantification for Multi-Scan Registration. *ACM Trans. Graph.* 39, 4, Article 130 (July 2020), 24 pages. https://doi.org/10.1145/3386569.3392402

Authors' addresses: Xiangru Huang, The University of Texas at Austin, 2317 Speedway, Austin, TX, 78712, USA; Zhenxiao Liang, The University of Texas at Austin, 2317 Speedway, Austin, TX, 78712, USA; Qixing Huang, The University of Texas at Austin, 2317 Speedway, Austin, TX, 78712, USA.

0730-0301/2020/7-ART130 $15.00
https://doi.org/10.1145/3386569.3392402

1 INTRODUCTION

Reconstructing 3D models from depth scans is a fundamental problem in geometry processing and beyond. A standard 3D reconstruction pipeline typically combines a geometry registration phase and a geometry reconstruction phase (c.f. [Huber 2002]). The goal of geometry registration is to bring depth scans obtained from different camera poses into the same coordinate system. The accuracy of this phase largely determines the quality of the reconstructed 3D model. Existing multi-scan registration techniques have predominantly

focused on developing suitable objective terms and novel optimization methods for computing a single registration result (c.f. [Tam et al. 2013]). In contrast, they place minimal efforts on studying an equally important problem of how to quantify the uncertainties of scan registration. Successful solutions to this task allow us to answer critical questions such as whether the current result is accurate, and if not, where to instill additional inputs to improve the quality of scan registration. A prominent example is how to quantify the uncertainties of global pose error, which is very hard to assess manually (particularly for models with non-planar structures).

In this paper, we introduce a local uncertainty quantification (or UQ) framework for multi-scan registration. The key idea is to treat multi-scan registration as solving an optimization problem. This optimization problem implicitly defines a map from the input scans to the output optimized poses. By assuming a noise model for generating the depth scans, this map then enables us to quantify local uncertainties of the optimized scans via propagating uncertainties associated with the input scans. We introduce new uncertainty measures derived from an approximation of the output covariance matrix, i.e., through the inverse Hessian evaluated at the current reconstruction. These uncertainty measures (e.g., diagonal blocks and leading eigenvectors of the approximated covariance matrix) enjoy multiple appealing properties, such as computational efficiency, quantified approximation errors, and information-rich. Compared to prior UQ approaches [Bosse and Zlot 2008; Manoj et al. 2015; Szeliski 1990; Szeliski and Kang 1997; Triggs et al. 1999; Wahba 1983; Zhang 1998] that utilize the inverse Hessian matrix to approximate the output covariance matrix, we provide a first analysis of the approximation errors, from which we develop informative and robust uncertainty measures. Such measures are critical for multi-scan registration, where it is infeasible and unnecessary to visualize the entire predicted covariance matrix.

We have applied this framework under two multi-scan registration settings, i.e., 1) geometric alignments between pairs of scans (or joint pairwise registration (JPR)) [Benjemaa and Schmitt 1998; Besl and McKay 1992; Brown and Rusinkiewicz 2007; Chen and Medioni 1992b; Krishnan et al. 2005; Pottmann et al. 2006; Raghuramu 2015], and 2) aligning the input scans to a deforming model (or simultaneous registration and reconstruction (SRAR)) [Huang et al. 2007a; Huang and Anguelov 2010; Jin et al. 1995; Liu et al. 2006; Tubic et al. 2003]. Under both settings, we develop geometric representations that lead to small discretization errors of our method on discrete objects (e.g., point clouds).

We have evaluated our approach on synthetic and real datasets. On synthetic datasets, experimental results justify that the predicted uncertainty measures match the simulated uncertainty measures obtained from repeated sampling the input depth-scans. On real datasets with ground-truth poses, the predicted covariance matrix is consistent with the pose errors. User studies demonstrate the effectiveness of our UQ approaches for enhancing user experiences in interactive scanning, and our approach accurately reveals the registration errors and significantly outperforms alternative methods.

As an application of our UQ approaches, we consider model-based view planning, which optimizes the camera poses on a similar 3D model to maximize the quality of the resulting 3D reconstruction [Scott et al. 2003]. We introduce a novel model-based view planner, where the objectives are to minimize the length of the scanning trajectory and the trace of the approximated covariance matrix. We present a practical greedy approach to solve the induced optimization problem efficiently. Compared to other model-based view planning approaches, our approach offers direct control of the reconstruction error and also generalizes better when the model and testing instances are different.

In summary, we present the following contributions:

- We define uncertainty measures for multi-scan registration that are computationally efficient, robust, and information-rich.
- We apply the proposed approach under two formulations, namely, joint pairwise registration and aligning scans to a deforming template model.
- We develop efficient approaches to visualize the quantified uncertainties for interactive scanning.
- We demonstrate the usefulness of our approach in interactive quality checking and model-based view planning.

2 RELATED WORKS

It is beyond the scope of this paper to provide a thorough overview of the literature on geometry registration. We refer to [Berger et al. 2014; Tam et al. 2013; Zollhöfer et al. 2018] for standard surveys on these topics. In the following, we discuss works on multi-scan registration and uncertainty quantification, which are most relevant to the context of this paper.

2.1 Multi-Scan Registration

Multi-scan registration seeks to bring depth scans stored in the individual camera coordinate systems into the same coordinate system. Depending on our assumptions about the input depth scans, existing approaches fall into the category of global methods [Arrigoni et al. 2016a,b; Govindu and Pooja 2014; Huang and Guibas 2013; Huang et al. 2017, 2019; Torsello et al. 2011; Wang and Singer 2013; Zhang et al. 2019], where we do not have any prior knowledge about the poses of the input scans, and the category of local methods [Benjemaa and Schmitt 1998; Brown and Rusinkiewicz 2007; Chen and Medioni 1992b; Huang et al. 2007a; Huang and Anguelov 2010; Krishnan et al. 2005; Pottmann et al. 2006; Raghuramu 2015], where we assume that we have some initial guesses about the underlying ground-truth poses. This paper studies uncertainty quantification for local methods. A standard strategy among local methods is to minimize the cumulative distances among the overlapping regions [Benjemaa and Schmitt 1998; Brown and Rusinkiewicz 2007; Chen and Medioni 1992b; Krishnan et al. 2005; Pottmann et al. 2006; Raghuramu 2015]. Although these methods work well on dense and clean depth scans, where the matches among overlapping regions are accurate, it becomes less reliable for sparse and noisy point clouds. In this case, depending on the structure of the overlapping areas, the alignment errors can accumulate.

Another category of methods unifies registration and reconstruction. The key idea is to optimize scan poses by minimizing the distances between the input scans and a deforming surface model. The representation of this dynamic model includes point clouds [Henry et al. 2012; Keller et al. 2013; Rusinkiewicz et al. 2002; Whelan et al.

2016] and implicit surfaces [Chen et al. 2013; Dai et al. 2017b; Huang et al. 2007b; Huang and Anguelov 2010; Izadi et al. 2011; Kähler et al. 2015; Whelan et al. 2015]. From the optimization point of view, these approaches fall into the category of incremental optimization [Chen et al. 2013; Dai et al. 2017b; Henry et al. 2012; Izadi et al. 2011; Kähler et al. 2015; Keller et al. 2013; Rusinkiewicz et al. 2002; Whelan et al. 2015, 2016] and global optimization [Huang et al. 2007b; Huang and Anguelov 2010]. A key advantage of these methods is that they align input scans with a densely sampled surface. As a result, they can take sparse point clouds as input and exhibit improved registration quality.

In contrast to developing new multi-scan registration techniques, this paper focuses on the new topic of quantifying the uncertainties of multi-scan registration techniques.

2.2 Uncertainty Analysis in Vision and Graphics

Uncertainty analysis is a long-standing topic in computer graphics and vision. In terms of methodology, present works fall into the category of simulation-based approaches and the group of algebraic methods. Simulation-based methods [Bengtsson and Baerveldt 2003; Iversen et al. 2017; Landry et al. 2019] apply sampling to generate inputs and then calculate statistics (e.g., the covariance matrix) among the induced outputs. These approaches can provide accurate estimations. However, they are computationally expensive, as the same algorithm has to run many times (typically in the order of thousands) on different inputs (c.f. [Landry et al. 2019]).

Algebraic methods [Bengtsson and Baerveldt 2003; Biber and Straßer 2003; Bosse and Zlot 2008; Censi 2007; Manoj et al. 2015; Nieto et al. 2005; Szeliski 1990; Szeliski and Kang 1997; Triggs et al. 1999; Wahba 1983; Zhang 1998] build on a noise model of the input and compute an approximation of the output covariance matrix via the inverse Hessian of the objective function. These approaches are efficient, as they only need to run each algorithm one time. However, they are less accurate than simulation-based methods due to approximation errors of the Hessian matrix and discretization errors (c.f. [Landry et al. 2019]). Our approach, which falls into this category, introduces innovative and rigorous means to address both issues. Specifically, we analyze the approximation errors of the inverse Hessian, from which we derive informative measures of the covariance matrix that are insensitive to these approximations. Besides, we carefully model the objective functions to minimize the effects of the discretization error.

In terms of applications, most existing works have focused on various pairwise alignment problems, e.g., structure-from-motion [Szeliski and Kang 1997; Triggs et al. 1999; Zhang 1998] and ICP registration [Bengtsson and Baerveldt 2003; Biber and Straßer 2003; Bosse and Zlot 2008; Censi 2007; Manoj et al. 2015; Nieto et al. 2005]. In contrast to these works, we focus on multi-scan registration. In this setting, the dimension of the covariance matrix is significantly higher than that of the pairwise scenario, and there are critical questions on what uncertainty measures are both informative and robust, how to compute them efficiently, and how to visualize them for interactive reconstruction.

Regarding multi-view reconstruction, several works [Kanatani and Morris 2006; Lhuillier and Perriollat 2006; Polic et al. 2018] studied uncertainty quantification under a feature-based structure-from-motion setting. Yet these approaches do not model the strong correlations among the location errors of image feature points, which depend on the camera poses and the underlying shape. In contrast, we follow a correspondence free approach, enabling us to model independent samples.

As the performance of algebraic methods depends on the discretization error, a technical focus of geometric uncertainty analysis is to develop suitable representations and error metrics that are insensitive to such errors. For example, in the application of pairwise geometric registration, the point-2-plane distance metric has proven to be much more accurate than the point-2-point distance metric (c.f. [Landry et al. 2019]). [Pauly et al. 2004] introduced a point-based surface representation to quantify the uncertainties of implicit surfaces derived from a point cloud. [Pöthkow et al. 2011] developed a grid-based representation for quantifying the uncertainties of converting an iso-surface into a triangular mesh. In contrast, we focus on modeling multi-scan registration under both the setting of minimizing pairwise distances between overlapping scans and the context of aligning input scans to a deforming template.

2.3 Geometrically Stable Sampling

A relevant task in geometry registration is to analyze the condition number of the Hessian matrix of the objective function. For example, [Gelfand et al. 2003] analyzed the condition number of the Hessian matrix for registering a pair of scans under the point-2-plane distance metric [Chen and Medioni 1992a]. They also introduced a greedy approach that selects a subset of the sample points so that the induced condition number is minimized. The fundamental difference between our UQ approach and the approach of optimizing the condition number is that we focus on the uncertainties of the optimal solution. In contrast, the method of minimizing the condition number can help the convergence to the optimal solution (c.f. [Nocedal and Wright 2006]).

2.4 Model-based View Planning

Besides maximizing the registration and reconstruction quality from a fixed set of scans, optimizing the set of input scans also dictates the reconstruction quality. Model-based view planning seeks to pre-compute a set of camera poses using a similar model, e.g., a coarse reconstruction or a model of the same category, and then uses the planned trajectory for scanning. The planned trajectory can be enforced exactly, e.g., using a camera rig, or showing it to the user as a suggestion for interactive scanning. Model-based view planning has been studied extensively in the literature. We refer to [Scott et al. 2003] for a literature review and to [Fan et al. 2016; Hepp et al. 2018] for recent advances.

In model-based view planning, designing a suitable 3D reconstruction quality score is crucial for both the reconstruction quality and the generalization behavior between training and testing models. This paper shows that when using the trace-norm of the pose covariance matrix, we obtain better reconstruction quality and generalization behavior than prior techniques.

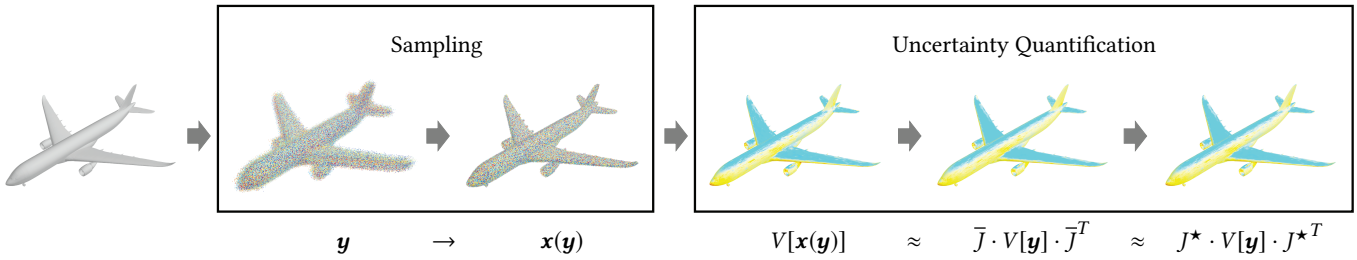

Fig. 2. Illustration of the proposed uncertainty quantification framework. (Left) We consider multi-scan registration as a mapping between a set of input scans and optimized scan poses using an off-the-shelf algorithm. The input scans are generated from a fixed set of unknown camera poses through a noise model characterized by the parameter $\boldsymbol{y}$. (Right) Our uncertainty quantification approach approximates the covariance of the output in two steps. The first step leverages an approximated linear map of the non-linear map at the underlying ground-truth. The second step uses the optimal solution to approximate the Jacobi matrix in the linear map, from which we develop informative uncertainty measures. This strategy allows us to perform uncertainty quantification from a single set of input scans. The sub-figures show norms of the diagonal blocks of the covariance matrices.

3 OVERVIEW

In this section, we provide a roadmap of the technical sections (Section 4 to Section 7) of this paper. As multi-scan registration algorithms typically solve optimization problems to compute the aligned scan poses, we describe in Section 4 a general framework for computing informative and robust uncertainty measures among the outputs of solving unconstrained optimization problems. In Section 5 and Section 6, we apply the proposed UQ framework to two popular formulations for multi-scan registration and reconstruction, namely, minimizing the distances between pairs of overlapping scans (Section 5) and minimizing the distances between the input scans and the deforming template (Section 6). In each setting, we present algorithms for uncertainty quantification and visualization schemes to interpret the quantified uncertainties. Finally, Section 7 presents a novel uncertainty-driven model-based view planner.

4 UNCERTAINTY QUANTIFICATION FOR UNCONSTRAINED OPTIMIZATION

In this section, we describe how to compute informative and robust uncertainty measures of the outputs that are defined by solving optimization problems. Our approach roots in an estimation of the output covariance matrix and an analysis of the approximation error. In the following, we first describe how to compute the approximated covariance matrix in Section 4.1. We then introduce robust and informative uncertainty measures in Section 4.2.

4.1 Implicitly Defined Optimal Solutions

We consider an unconstrained optimization problem, where the objective function admits the form $f(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{w}, \boldsymbol{p}^{gt})$. Here $\boldsymbol{p}^{gt}$ denotes the underlying ground-truth; $\boldsymbol{y}$ encodes the input noise model; $\boldsymbol{w}$ collects the hyper-parameters of f; $\boldsymbol{x}$ parameterizes the output variables. In the context of this paper (See Figure 2), $\boldsymbol{p}^{gt}$ denotes the underlying ground-truth scan points (i.e., when there is no noise) in a world coordinate system. Note that $\boldsymbol{p}^{gt}$ is generally unavailable. We will discuss how to address this issue in Section 4.2. $\boldsymbol{y}$ models the uncertainties when generating the input scans. Each sampling of $\boldsymbol{y}$ gives one set of input scans for geometry registration. $\boldsymbol{x}$ collects pose variables (joint pairwise registration and simultaneous registration and reconstruction) and parameters of the underlying 3D model (simultaneous registration and reconstruction). $\boldsymbol{w}$ encodes the scan indicators, i.e., whether each scan participates in multi-scan registration or not. Note that we treat $\boldsymbol{w}$ as a constant vector until Section 7, where we study scan selection. However, we keep $\boldsymbol{w}$ in our notations for the completeness of our approach.

Without losing generality, we place the additional assumption that $f(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{w}, \boldsymbol{p}^{gt}) \geq 0$ and $f(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{w}, \boldsymbol{p}^{gt}) = 0$ if and only if $\boldsymbol{x} = 0$ and $\boldsymbol{y} = 0$. In other words, $\boldsymbol{x}$ encodes deviations from the underlying ground-truth pose, $\boldsymbol{y}$ characterizes the noise-level, and the underlying ground-truth is recovered when there is no input noise. Under this setup, we formulate the map from $\boldsymbol{y}$ to $\boldsymbol{x}$ as solving the following optimization problem:

$$\boldsymbol{x}^\star(\boldsymbol{y}, \boldsymbol{w}, \boldsymbol{p}^{gt}) = \operatorname*{argmin}_{\boldsymbol{x}} f(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{w}, \boldsymbol{p}^{gt}). \tag{1}$$

Given a probability distribution over $\boldsymbol{y}$, the map defined in (1) induces a probability distribution over $\boldsymbol{x}$. Quantifying this distribution to visualize the uncertainties of multi-scan registration introduces several ground challenges. First, $\boldsymbol{x}^\star(\boldsymbol{y}, \boldsymbol{w}, \boldsymbol{p}^{gt})$ is a non-linear map of $\boldsymbol{y}$. Second, the underlying ground-truth $\boldsymbol{p}^{gt}$ is unknown. Finally, for high-dimensional problems such as multi-scan registration, we need concise and informative measures. Our approach is motivated from a line of works [Bengtsson and Baerveldt 2003; Bosse and Zlot 2008; Censi 2007; Manoj et al. 2015; Szeliski and Kang 1997; Wahba 1983; Zhang 1998] that approximates implicitly defined non-linear maps by their linear counterparts. The results lead to approximations of the underlying covariance matrix. A key novelty of this paper is that we analyze the approximation error to obtain accurate, informative, and concise uncertainty measures that are suitable for high-dimensional problems such as multi-scan registration. Another subtle difference is that we consider two approximated covariance matrices, one for computation, and another for the application of model-based view planning (See Section 7). In the following, we describe the approximated covariance matrices. In Section 4.2, we introduce our uncertainty quantification algorithm.

Specifically, we first approximate the non-linear map defined in (1) by its linear approximation, i.e.,

$$\boldsymbol{x}^\star(\boldsymbol{y}, \boldsymbol{w}, \boldsymbol{p}^{gt}) \approx \frac{\partial \boldsymbol{x}^\star}{\partial \boldsymbol{y}}(\boldsymbol{0}, \boldsymbol{w}, \boldsymbol{p}^{gt}) \cdot \boldsymbol{y}. \tag{2}$$

We can compute the partial derivative $\frac{\partial \boldsymbol{x}^\star}{\partial \boldsymbol{y}}(\mathbf{0}, \boldsymbol{w}, \boldsymbol{p}^{gt})$ via the implicit function theorem [Krantz and Parks 2013]:

$$\frac{\partial \boldsymbol{x}^\star}{\partial \boldsymbol{y}}(\mathbf{0}, \boldsymbol{w}, \boldsymbol{p}^{gt}) := -\big(\frac{\partial^2 f}{\partial \boldsymbol{x}^2}(\mathbf{0}, \mathbf{0}, \boldsymbol{w}, \boldsymbol{p}^{gt})\big)^{-1}\big(\frac{\partial^2 f}{\partial \boldsymbol{x}\partial \boldsymbol{y}}(\mathbf{0}, \mathbf{0}, \boldsymbol{w}, \boldsymbol{p}^{gt})\big) \quad (3)$$

To address the second challenge that $\boldsymbol{p}^{gt}$ is unavailable, we simply replace $\boldsymbol{p}^{gt}$ by a reconstruction $\boldsymbol{p}^\star$ derived from one set of input scans. This approach leads to the following approximation

$$\boldsymbol{x}^\star(\boldsymbol{y}, \boldsymbol{w}, \boldsymbol{p}^{gt}) \approx \frac{\partial \boldsymbol{x}^\star}{\partial \boldsymbol{y}}(0, \boldsymbol{w}, \boldsymbol{p}^\star) \cdot \boldsymbol{y}. \quad (4)$$

The approximations described in (2) and (4) offer two approximations of the output covariance matrix $V[\boldsymbol{x}^\star]$:

$$\overline{C} := \frac{\partial \boldsymbol{x}^\star}{\partial \boldsymbol{y}}(\mathbf{0}, \boldsymbol{w}, \boldsymbol{p}^{gt}) \cdot V[\boldsymbol{y}] \cdot \frac{\partial \boldsymbol{x}^\star}{\partial \boldsymbol{y}}(\mathbf{0}, \boldsymbol{w}, \boldsymbol{p}^{gt})^T, \quad (5)$$

$$C := \frac{\partial \boldsymbol{x}^\star}{\partial \boldsymbol{y}}(\mathbf{0}, \boldsymbol{w}, \boldsymbol{p}^\star) \cdot V[\boldsymbol{y}] \cdot \frac{\partial \boldsymbol{x}^\star}{\partial \boldsymbol{y}}(\mathbf{0}, \boldsymbol{w}, \boldsymbol{p}^\star)^T. \quad (6)$$

where $V[\boldsymbol{y}]$ denotes the input covariance matrix. Our UQ approach will leverage the second covariance matrix C, which only requires a single output [Szeliski and Kang 1997; Wahba 1983; Zhang 1998] (e.g., a single run of multi-scan registration). In contrast, we will use the first covariance matrix $\overline{C}$ to develop a model-based view planner (e.g., where the underlying ground-truth is known).

When the approximations in (2) and (4) are precise and the elements of $\boldsymbol{y}$ are independently random variables with zero mean, then $\boldsymbol{x}^\star$ is approximately a Gaussian distribution with zero mean (due to the central limit theorem [Van Der Vaart 1998]). Since the covariance matrix fully characterizes a Gaussian distribution with zero mean, both $\overline{C}$ and C are informative in this setting.

For multi-scan registration, both (2) and (4) are inexact. Therefore, it is critical to analyze the approximation errors. In the following, we use one such analysis to derive measures of C that are informative and insensitive to the approximation errors. Note that such uncertainty measures are also preferred for multi-scan registration, as it is infeasible to visualize any entire covariance matrix.

4.2 Informative Uncertainty Measurements

To understand the approximation errors of C and $\overline{C}$, we study the following model for generating the input noise. Specifically, we assume that the elements of $\boldsymbol{y}$ are independent, and each y_i satisfies

$$\text{Expectation}: \quad E[y_i] = 0 \quad (7)$$

$$\text{Variance}: \quad \mathrm{V}[y_i] = \sigma^2 \quad (8)$$

$$\text{Bounded}: \quad |y_i| \le \delta\sigma \quad (9)$$

for some $\delta > 1$. Note that this model does not require the distribution of each y_i to be a Gaussian. The following proposition states a bound on the difference between C ($\overline{C}$) and the output covariance matrix $V[\boldsymbol{x}^\star]$ with respect to the matrix spectral norm.

Proposition 1. ***(Informal)*** *Under the noise model described above and mild conditions on the derivatives of f, we have*

$$\frac{1}{\sigma^2}\max\left(\|C - V[\boldsymbol{x}^\star]\|, \|\overline{C} - V[\boldsymbol{x}^\star]\|\right) \le \sigma C(\kappa(f_{\boldsymbol{xx}}), \delta) \quad (10)$$

with high probability. Here $\|\cdot\|$ denotes the spectral norm; $f_{\boldsymbol{xx}} = \frac{\partial^2 f}{\partial^2 \boldsymbol{x}}$; C depends on δ; $\kappa(f_{\boldsymbol{xx}}) := \lambda_{\max}(f_{\boldsymbol{xx}})/\lambda_{\min}(f_{\boldsymbol{xx}})$ denotes the ratio between the maximum and the minimum eigenvalues of $f_{\boldsymbol{xx}}$, i.e., its the condition number. The smaller δ and $\kappa(f_{\boldsymbol{xx}})$, the bigger $C(\kappa(f_{\boldsymbol{xx}}), \delta)$.

The formal statement of Prop. 1 and the proof are derived from the Taylor expansions of $\boldsymbol{x}^\star(\boldsymbol{y})$ and f. We defer the details to the supplemental material.

Remark 1. *Prop. 1 states that the approximation error is dictated by the condition number of $f_{\boldsymbol{xx}}$. In the context of multi-scan registration, this condition number connects to the sparsity pattern of $f_{\boldsymbol{xx}}$, which depends on how the input scans overlap with each other. As we will see in Section 8, an implication is that when the overlapping relations among the input scans are strong, the approximation error of C is small. On the other hand, for thin objects or when input scans form strongly overlapping components, the approximation error of C can be large.*

Motivated from Prop 1, we introduce two groups of measures to access the uncertainties of $\boldsymbol{x}^\star$:

- **Diagonal blocks.** The first group consists of diagonal blocks of C, each of which corresponds to one group of variables, e.g., 6×6 blocks for scan poses. These measures are accurate, as the spectral norms of the deviations in diagonal blocks are upper bounded by those of the entire matrices. They are also easy to compute. On the other hand, they do not capture the correlations among uncertainties of multiple scans.
- **Leading eigenvectors.** The second group consists of leading eigenvectors of C, which capture structural correlations among groups of variables, e.g., clusters of scans that strongly overlap with each other.
  The motivation comes from the fact that leading eigenvectors are also stable under perturbations. Specifically, Given two symmetric matrices $A, E \in \mathbb{R}^{n\times n}$, a variant of the Davis-Kahan theorem [Davis and Kahan 1970] states that the distance between the k-th eigenvector $\boldsymbol{u}_k$ of A and the eigenspace $U_{k'} \in \mathbb{R}^{n\times k'}$ formed by the top $k' > k$ eigenvectors of $A + E$ is bounded as $\|\boldsymbol{u}_k - (U_{k'} \cdot U_{k'}^T)\boldsymbol{u}_k\|^2 \le \|E\|/(\lambda_k - \lambda_{k'} - \|E\|)$, where λ_k is the k-th eigenvalue of A. In other words, the top eigenvectors of $V[\boldsymbol{x}^\star]$ are well approximated by their projections in the leading eigenspace of C.

5 JOINT PAIR-WISE REGISTRATION

In this section, we apply the framework described in Section 4 to joint pairwise registration (or JPR), i.e., minimizing distances between all pairs of overlapping scans. We begin by describing the formulation of JPR in Section 5.1. We then discuss the uncertainty quantification setup in Section 5.2. Finally, we discuss how to visualize the quantified uncertainties in Section 5.3.

5.1 Formulation of JPR

This formulation considers the setting of aligning N scans $\mathcal{S} = \{S_1, \cdots, S_N\}$. Each scan $S_i = \{p_{i1}, \cdots, p_{iN_i}\}$ consists of a set of surfels. With $\boldsymbol{p}_{ij}$ and $\boldsymbol{n}_{ij}$ we denote the position and the normal of p_{ij}, respectively. Let $T_i = (R_i, \boldsymbol{t}_i)$ be the rigid pose associated

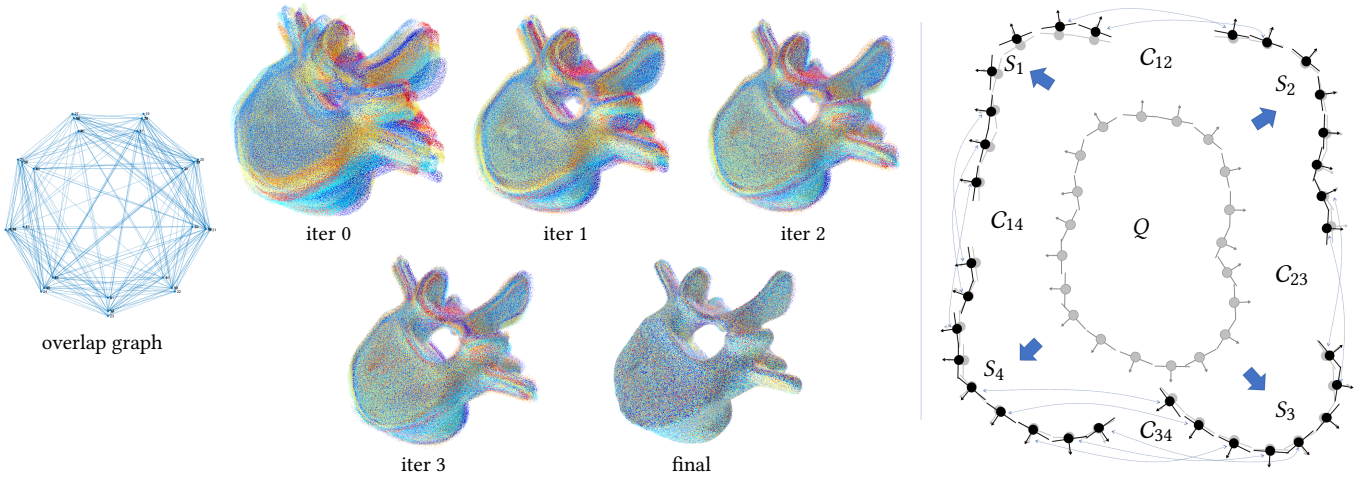

Fig. 3. (Left) Illustration of multi-scan registration by minimizing the pairwise distances between overlapping scans. We show the overlapping graph among the input scans, intermediate results when alternating between closest-point computation and pose optimization, and the final optimized scans. (Right) Illustration of the noisy model under this setting. Q: the underlying ground-truth model. S_i, $1 \le i \le 4$: Simulated scans. C_{ij}, $(i, j) \in \{(1, 2), (1, 4), (3, 4), (2, 3)\}$: Correspondences between overlapping scan pairs.

with scan S_i. Then multiple-scan registration amounts to solve the following optimization problem (c.f. [Tam et al. 2013]):

$$\begin{aligned} \underset{\{T_i\}}{\text{minimize}} \quad & \sum_{(i,i')\in\mathcal{E}} d^2(S_i, T_i, S_{i'}, T_{i'}) \\ \text{subject to} \quad & R_1 = I_3, \boldsymbol{t}_1 = \boldsymbol{0}. \end{aligned} \tag{11}$$

Here $\mathcal{E}$ encodes the graph of overlapping scans. $d(S_i, T_i, S_{i'}, T_{i'})$ denotes the distance between S_i and $S_{i'}$ along their overlapping region. Let $C_{ii'} \subset \mathcal{S}_i \times \mathcal{S}_{i'}$ collect the set of unsigned correspondences between $\mathcal{S}_i$ and $\mathcal{S}_{i'}$, i.e., $(p_{ij}, p_{i'j'}) \in C_{ii'}$ if and only if $(p_{i'j'}, p_{ij}) \in C_{i'i}$, we use the well-known point-2-plane distance metric [Chen and Medioni 1992a] to define $d^2(S_i, T_i, S_{i'}, T_{i'}) :=$

$$\sum_{(p_{ij}, p_{i'j'})\in C_{ii'}} \left((R_i \boldsymbol{p}_{ij} + \boldsymbol{t}_i - R_{i'}\boldsymbol{p}_{i'j'} - \boldsymbol{t}_{i'})^T (R_{i'}\boldsymbol{n}_{i'j'})\right)^2 \tag{12}$$

Substitute (12) into (11), we arrive at an optimization problem of non-linear least squares, where the variables are $T_i, 2 \le i \le n$. In this paper, we use the Gauss-Newton method for solving (11).

Remark 2. *For the purpose of uncertainty quantification, we assume the edge set $\mathcal{E}$ and the correspondence sets $C_{ii'}, (i, i') \in \mathcal{E}$ are fixed. During the course of registration, we follow the common strategy of updating both sets (c.f. [Tam et al. 2013]). Specifically, each correspondence set $C_{ii'}$ is iteratively updated by computing the closest points between S_i and $S_{i'}$ under the current scan poses. We trim each correspondence set if the point-wise distance is above two times the median of all point-wise distances. Since the particular registration algorithm is not the focus of this paper, we defer the technical details to ([Tam et al. 2013]). Figure 3(Left) illustrates the effects of multi-scan registration via minimizing pair-wise distances.*

5.2 Uncertainty Quantification

We first describe a noise model for generating the input scans. We then introduce our UQ algorithm (See Section 5.2.2).

5.2.1 Noise model. Figure 3(Right) illustrates the uncertainty quantification setup for joint pair-wise registration. We model the input scans as perturbations from a latent surface model. Specifically, consider a latent surface model represented as a set of surfels $\{s_k = (\boldsymbol{q}_k, \boldsymbol{n}_k), 1 \le k \le K\}$, where $\boldsymbol{q}_k$ and $\boldsymbol{n}_k$ denote the position and the normal of s_k, respectively. Let $k_{ij} \in \{1, \cdots, K\}$ be the index of the corresponding surfel for scan point $\boldsymbol{p}_{ij}$. With $\boldsymbol{p}_{ij}^{gt}$ we denote the ground-truth location for $\boldsymbol{p}_{ij}$, which is close to $\boldsymbol{q}_{k_{ij}}$ and lies on the plane specified by $(\boldsymbol{q}_{k_{ij}}, \boldsymbol{n}_{k_{ij}})$, i.e., $(\boldsymbol{p}_{ij}^{gt} - \boldsymbol{q}_{k_{ij}})^T \boldsymbol{n}_{k_{ij}} = 0$. We model $\boldsymbol{p}_{ij}$ and $\boldsymbol{n}_{ij}$ so that they perturb $\boldsymbol{p}_{ij}^{gt}$ and $\boldsymbol{n}_{k_{ij}}$, respectively:

$$\boldsymbol{p}_{ij} = \boldsymbol{p}_{ij}^{gt} + y_{ij}\boldsymbol{n}_{k_{ij}}, \quad \boldsymbol{n}_{ij} = \boldsymbol{n}_{k_{ij}} + \boldsymbol{z}_{ij} \tag{13}$$

where $y_{ij} \in \mathbb{R}$ and $\boldsymbol{z}_{ij} \in \mathbb{R}^3$ encode the noises in positions and normals, respectively. In our UQ approach, we assume $\{y_{ij}\} \cup \{\boldsymbol{z}_{ij}\}$ are independent with each other, and they satisfy

$$E[y_{ij}] = 0, \mathrm{V}[y_{ij}] = \sigma^2, \quad E[\boldsymbol{z}_{ij}] = 0, \mathrm{V}[\boldsymbol{z}_{ij}] = \Sigma_n. \tag{14}$$

where $\Sigma_n \in \mathbb{R}^{3\times 3}$ is a pre-defined covariance matrix.

In our experiments, we set σ and Σ_n as the variance of point-2-plane distance and the variance of the normal differences, respectively. We evaluate these variances among all correspondences in $\{C_{ii'}\}$. We them divide both variances by two to predict the variances from the latent surface model.

For uncertainty quantification, we place one additional assumption. Define the induced correspondences between S_i and $S_{i'}$ as

$$C_{ii'}^{gt} := \{(p_{ij}, p_{i'j'}) | 1 \le j \le N_i, 1 \le j \le N_{i'}, k_{ij} = k_{i'j'}\}.$$

We assume the correspondence set $C_{ii'}$ employed in joint pairwise registration is contained in $C_{ii'}^{gt}$, i.e., $C_{ii'} \subset C_{ii'}^{gt}$. In other words, if $(p_{ij}, p_{i'j'}) \in C_{ii'}$ then $k_{ij} = k_{i'j'}$. As analyzed in [Pottmann and Hofer 2002; Pottmann et al. 2004], this is a feasible assumption as the squared point-2-plane distance is a quadratic approximation of the squared distance function.

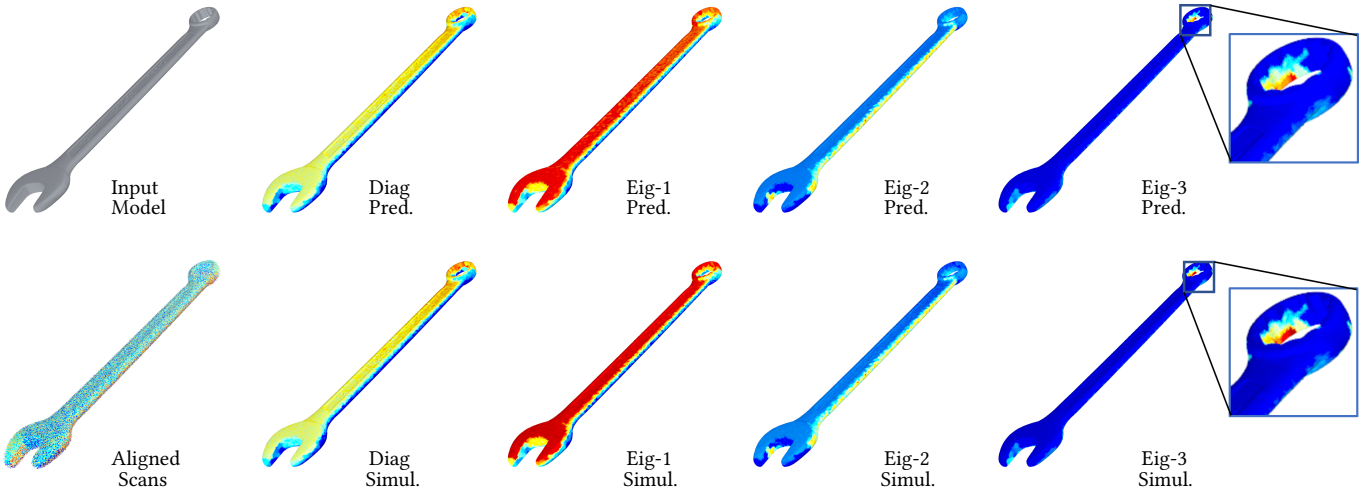

Fig. 4. Visualizing pose uncertainties of reconstructing a wrench model under joint pairwise registration (or JPR). We show the input model, the aligned scans, the diagonal blocks, and three leading eigenvectors. The top row shows the predicted results, and the bottom row shows the simulated results. We can see that the predicted uncertainties are consistent with the intuition that this model possesses planar faces that can glide when performing registration. Such planar face structures are revealed in the leading eigenvectors. Also, our approach reveals high-frequency signals of the simulated covariance matrix.

ALGORITHM 1: Uncertainty quantification algorithm for joint pairwise registration.

Input: An overlapping graph $\mathcal{E}$ and optimized correspondence sets $C_{ii'}$, $(i, i') \in \mathcal{E}$. Optimized scan point positions $\boldsymbol{p}_{ij}^{\star}$, $1 \le i \le N$, $1 \le j \le N_i$ and scan point normals $\boldsymbol{n}_{ij}^{\star}$, $1 \le i \le N$, $1 \le j \le N_i$.
Output: Diagonal blocks and leading eigenvectors of C_{JPR}
Step 1: Compute $\{\boldsymbol{h}_{ij}\}$ and $\{v_{ij,i'j'}\}$ using (15);
Step 2: Compute C_m, A_{dis}, and C_{nor} using (16);
Step 3: Use (17) to compute the diagonal blocks and leading eigenvectors of C_{JPR};

5.2.2 Algorithm. We proceed to introduce an algorithm for quantifying the uncertainties of joint pairwise registration (See Algorithm 1 for the pseudo-code). First, we parameterize the rotation of each scan as $R_i = \exp(-\boldsymbol{c}_i\times)$, where $\boldsymbol{c}_i \in \mathbb{R}^3$. This parameterization turns (11) into a unconstrained optimization problem.

Applying the methodology described in Section 4, we obtain the following expression of the estimated covariance matrix $C_{\text{JPR}} \in \mathbb{R}^{6(N-1)\times 6(N-1)}$ from one single input.

Definition 1. *Given a matrix $A \in \mathbb{R}^{n_1\times n_2}$. Define the linear operator $V_{i:i'}(A) \in \mathbb{R}^{(i'-i+1)\times n_2}$ so that it takes the sub-matrix of A that collects its i-th row to the i'-th row. With $V_i(A)$ we denote the operator that takes the i-th row of A. This notation also applies to the case where A is a column vector.*

Proposition 2. *Denote $\boldsymbol{p}_{ij}^{\star}$ and $\boldsymbol{n}_{ij}^{\star}$ as the optimized position and the optimized normal for surfel p_{ij}, respectively. Introduce sparse coefficient vectors $\boldsymbol{h}_{ij} \in \mathbb{R}^{6(N-1)}$, $1 \le i \le N, 1 \le j \le N_i$, where the non-zero elements are given by*

$$\begin{aligned} V_{(6i-11):(6i-9)}(\boldsymbol{h}_{ij}) &:= \boldsymbol{n}_{ij}^{\star}, & 2 \le i \le N, 1 \le j \le N_i \\ V_{(6i-8):(6i-6)}(\boldsymbol{h}_{ij}) &:= \boldsymbol{p}_{ij}^{\star} \times \boldsymbol{n}_{ij}^{\star}, & 2 \le i \le N, 1 \le j \le N_i \end{aligned} \quad (15)$$

Let $C = \cup_{(i,i')\in\mathcal{E}} C_{ii'}$ collect all point-wise correspondences. Introduce

$$\begin{aligned} C_m &:= \sum_{(p_{ij},p_{i'j'})\in C} (\boldsymbol{h}_{ij} - \boldsymbol{h}_{i'j'})(\boldsymbol{h}_{ij} - \boldsymbol{h}_{i'j'})^T \\ A_{dis} &:= \sigma \cdot \sum_{(p_{ij},p_{i'j'})\in C} (\boldsymbol{h}_{ij} - \boldsymbol{h}_{i'j'})\boldsymbol{n}_{i'j'}^T \big((\boldsymbol{e}_{ij} - \boldsymbol{e}_{i'j'})^T \oplus I\big) \end{aligned} \quad (16)$$

where $\boldsymbol{e}_{ij} \in \mathbb{R}^{\sum N_i}$ is the canonical basis with the ij-th element being one and $\boldsymbol{e}_{ij} \otimes I \in \mathbb{R}^{(3\sum N_i)\times 3}$. Then we have

$$C_{\text{JPR}} = C_m^{-1} A_{dis} A_{dis}^T C_m^{-1}. \quad (17)$$

Proof. Please refer to the supplemental material. □

5.3 Uncertainty Visualization

As described in Section 4, we interpret C_{JPR} using its diagonal blocks and leading eigenvectors. This section discusses how to visualize these quantities. Let $C_{\text{JPR}}^{(i)} \in \mathbb{R}^{6\times 6}, 2 \le i \le N$ denote the diagonal block that corresponds to scan S_i, we convert it into an uncertainty value for scan S_i. To this end, we first compute its leading eigenvalue $\lambda_i^{\max} \in \mathbb{R}$ and leading eigenvector $\boldsymbol{u}_i^{\max} \in \mathbb{R}^6$, respectively. Decompose $\sqrt{\lambda_i^{\max}}\boldsymbol{u}_i^{\max} = (\bar{\boldsymbol{c}}_i; \boldsymbol{c}_i)$, where $\bar{\boldsymbol{c}}_i$ and $\boldsymbol{c}_i$ correspond the translation and rotation components, respectively. We then define the uncertainty value for S_i as

$$\sigma_i = \frac{1}{N_i}\sum_{j=1}^{N_i} \|\bar{\boldsymbol{c}}_i + \boldsymbol{c}_i \times \boldsymbol{p}_{ij}\| \quad (18)$$

which is the average displacement among S_i incurred by $(\bar{\boldsymbol{c}}_i; \boldsymbol{c}_i)$. As shown in Figure 4, we reconstruct a mesh from the aligned scans and visualize $\sigma_i, 1 \le i \le N$ as a scalar field on the reconstructed mesh. This is done by performing a range query on the aligned scans from each mesh vertex and then computing the mean among

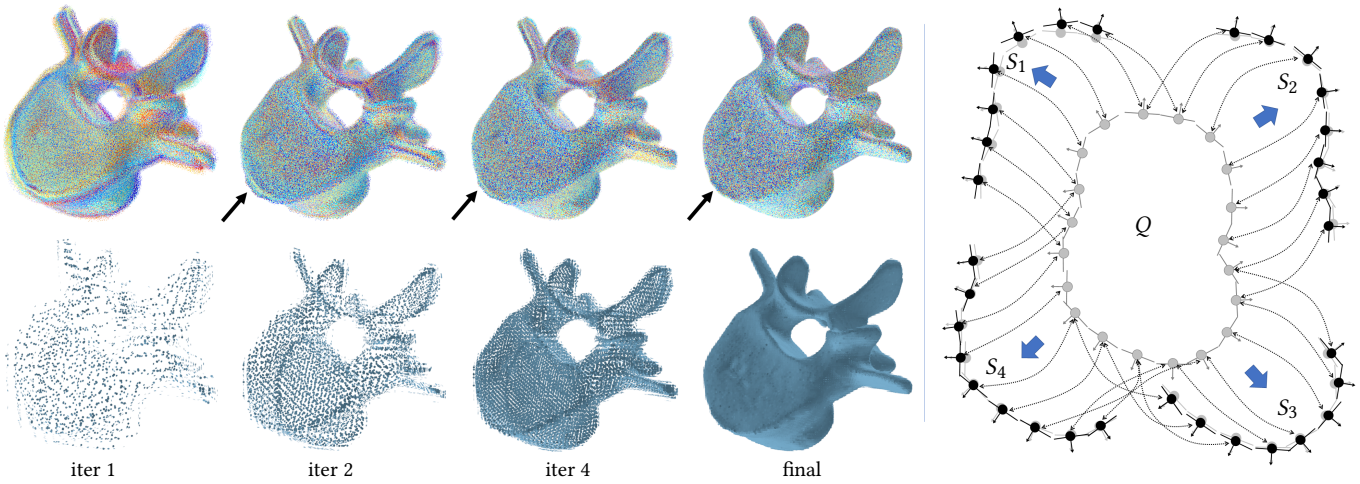

Fig. 5. The second approach studied in this paper applies simultaneous registration and reconstruction (or SRAR) to jointly align a set of scans and reconstruct a 3D model from the aligned scans. (Left) The procedure of SRAR for reconstructing the same model in Figure 3. Each iteration optimizes the scan poses (top) and a collection of surfels (bottom) by minimizing the distances from the scan points to the surfels. These surfels are initialized by performing principal component analysis (or PCA) among scan points that fall into cells of a grid. Note that while the procedure involves registration and reconstruction at multiple levels, uncertainty quantification is performed at the finest level. (Right) An illustration of the uncertainty quantification setup for SRAR.

the associated values of the neighboring points. We use a similar scheme to visualize each leading eigenvector, i.e., by feeding the corresponding elements of each scan into (18).

Figure 4 shows visualizations of a wrench model that possesses planar structures. Intuitively, scans of these planar faces can drift (c.f. [Gelfand and Guibas 2004]). We can see that the diagonal blocks and the leading eigenvectors nicely reveal such structural patterns.

6 SIMULTANEOUS REGISTRATION AND RECONSTRUCTION

In this section, we apply the proposed framework to the setting of simultaneous registration and reconstruction (or SRAR) [Huang et al. 2007a; Huang and Anguelov 2010]. A standard reconstruction pipeline usually consists of two sequential steps, where the first step performs joint pairwise registration, and where the second step reconstructs a 3D model from the aligned scans. In contrast, SRAR unifies registration and reconstruction by solving a single optimization problem. The key idea is to align the input scans to a latent surface model, which is also jointly optimized with the scan poses. In the following, we first describe the formulation of SRAR in Section 6.1. We then introduce the uncertainty quantification formulation for SRAR in Section 6.2.

6.1 Formulation of SRAR

The input to SRAR consists of N depth scans $S_i = \{\boldsymbol{p}_{ij}, 1 \le j \le N_i\}, 1 \le i \le N$. Note that unlike joint pairwise registration (or JPR) which relies on the point normals to formulate the registration potential, SRAR does not require point normals as input.

The key difference between SRAR and JPR lies in how to represent the latent surface model and how to define the distance metric for aligning each scan with the latent surface. We adapt the setup in Section 5 to represent the latent surface as a collection of surfels $s_k = (\boldsymbol{q}_k, \boldsymbol{n}_k), 1 \le k \le K$. As we will use the point-2-plane distance metric [Chen and Medioni 1992a] to align the input scans with the latent surface, each surfel possesses three independent variables, i.e., the corresponding plane equation $(d_k, \boldsymbol{n}_k)$, where d_k denotes the distance from the origin to s_k.

Let $k_{ij} \in \{1, \cdots, K\}$ denote the corresponding plane of point $\boldsymbol{p}_{ij}$ (i.e., the one with closest surfel position $\boldsymbol{q}_{k_{ij}}$). SRAR amounts to optimize the pose $(R_i, \boldsymbol{t}_i)$ of each scan S_i and the plane equations $\{(d_k, \boldsymbol{n}_k)\}$ by solving the following optimization problem:

$$\begin{aligned} \underset{\{R_i, \boldsymbol{t}_i\}, \{(d_k, \boldsymbol{n}_k)\}}{\text{argmin}} \quad & \sum_{i=1}^{N} \sum_{j=1}^{N_i} \left((R_i \boldsymbol{p}_{ij} + \boldsymbol{t}_i)^T \boldsymbol{n}_{k_{ij}} - d_{k_{ij}} \right)^2 \\ \text{subject to} \quad & R_1 = I_3, \boldsymbol{t}_1 = \boldsymbol{0}. \end{aligned} \tag{19}$$

In other words, (19) utilizes the point-2-plane distance metric to align the data points with the latent surface model. As discussed in [Huang and Anguelov 2010], (19) admits a simple and effective minimization strategy, which alternates between optimizing $\{(R_i, \boldsymbol{t}_i)\}$ and $\{(d_k, \boldsymbol{n}_k)\}$.

Remark 3. *In the same spirit as ICP registration [Besl and McKay 1992], we also update the associations between the scan points and the latent planes by recursively finding the closest surfel positions. Such operations are particularly useful when the input scans are not aligned. To achieve fast convergence, we can also use a coarse latent surface and gradually refine it when scans become more aligned. Figure 5(Left) shows a typical optimization procedure for SRAR on the same dataset shown in Figure 3(Left). However, when performing uncertainty quantification, we assume these associations are fixed.*

6.2 Uncertainty Quantification Formulation

We proceed to apply the proposed framework described in Section 4 to the objective function of SRAR.

6.2.1 Noise model. We employ a similar noise model as the one used in JPG (See Figure 5(Right)). Specifically, consider a set of ground-truth planes $(d_k^{gt}, \boldsymbol{n}_k^{gt})$. Let $\boldsymbol{p}_{ij}^{gt}$ be the ground-truth location for scan point $\boldsymbol{p}_{ij}$. With k_{ij} we denote the index of the corresponding plane of $\boldsymbol{p}_{ij}$. Then

$$\boldsymbol{n}_{k_{ij}}^{gt\ T} \boldsymbol{p}_{ij}^{gt} - d_{k_{ij}}^{gt} = 0. \tag{20}$$

The input positions $\{\boldsymbol{p}_{ij}\}$ are generated by following (13). Unlike JPR, we compute σ as the variance of the point-2-plane distance between the optimized scan points and their corresponding planes.

6.2.2 Formulation. In the same spirit as JPG, we assume that for each sampled input described above, SRAR converges to the underlying point-plane associations $\{k_{ij}\}$. In practice, we found that the discretization error of the proposed SRAR formulation is small. Intuitively, point-2-plane distances are insensitive to wrong point-plane associations when the planes are compatible with each other locally.

To apply (4), the remaining task is to convert (19) into an unconstrained optimization problem by using the following local parameterizations for rotations and scan poses:

$$R_i = \exp(\boldsymbol{c}_i\times) \quad 1 \le i \le N$$
$$\boldsymbol{n}_k = \frac{\boldsymbol{n}_k^{gt} + x_{k,1}\boldsymbol{t}_{k,1}^{gt} + x_{k,2}\boldsymbol{t}_{k,2}^{gt}}{\|\boldsymbol{n}_k^{gt} + x_{k,1}\boldsymbol{t}_{k,1}^{gt} + x_{k,2}\boldsymbol{t}_{k,2}^{gt}\|},\ d_k = d_k^{gt} + x_{k,3} \quad 1 \le k \le K \tag{21}$$

where $(\boldsymbol{n}_l^{gt}, \boldsymbol{t}_{k,1}^{gt}, \boldsymbol{t}_{k,2}^{gt})$ forms a local coordinate system.

Applying (5) to the objective function (19) of SRAR, we arrive at the following proposition for computing the approximated covariance matrix.

Proposition 3. *Introduce sparse coefficient vectors $\boldsymbol{g}_{ij} \in \mathbb{R}^{6(N-1)+3K}$, $1 \le i \le N, 1 \le j \le N_i$, whose non-zero elements are given by*

$$\begin{aligned}
V_{(6i-11):(6i-9)}(\boldsymbol{g}_{ij}) &:= \boldsymbol{n}_{k_{ij}}^{gt}, && 2 \le i \le N, 1 \le j \le N_i\\
V_{(6i-8):(6i-6)}(\boldsymbol{g}_{ij}) &:= \boldsymbol{p}_{ij}^{gt} \times \boldsymbol{n}_{k_{ij}}^{gt}, && 2 \le i \le N, 1 \le j \le N_i\\
V_{6(N-1)+3k_{ij}}(\boldsymbol{g}_{ij}) &:= \boldsymbol{t}_{k_{ij},1}^{T} \cdot \boldsymbol{p}_{ij}^{gt}, && 1 \le i \le N, 1 \le j \le N_i\\
V_{6(N-1)+3k_{ij}+1}(\boldsymbol{g}_{ij}) &:= \boldsymbol{t}_{k_{ij},2}^{T} \cdot \boldsymbol{p}_{ij}^{gt}, && 1 \le i \le N, 1 \le j \le N_i\\
V_{6(N-1)+3k_{ij}+2}(\boldsymbol{g}_{ij}) &:= -1, && 1 \le i \le N, 1 \le j \le N_i
\end{aligned}$$

Denote

$$C_{XX} := \sum_{i=1}^{N}\sum_{j=1}^{N_i} \boldsymbol{g}_{ij}\boldsymbol{g}_{ij}^T$$

Then the uncertainty quantification approach described above outputs the following approximated covariance matrix:

$$C_{\text{SRAR}} = \sigma^2 C_{XX}^{-1}. \tag{22}$$

Proof. Please refer to the supplemental material. □

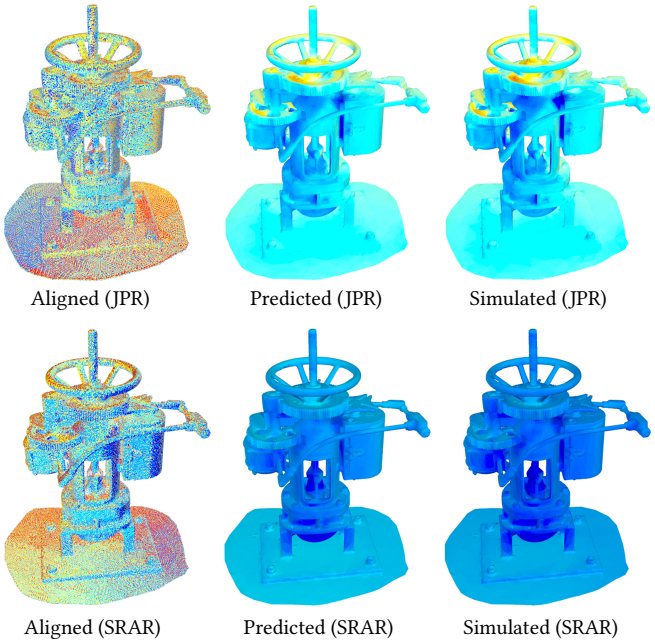

Fig. 6. Comparison between joint pairwise registration (or JPR) and simultaneous registration and reconstruction (or SRAR) on the same set of scans of an engine model. (Top) From left to right: the aligned scans under JPR, a visualization of the diagonal blocks of the predicted covariance matrix under JPR, and a visualization of the diagonal blocks of the simulated covariance matrix under JPR. (Bottom) Visualizations of the corresponding quantities under SRAR. We can see that SRAR is superior to JPR.

Decompose $C_{XX} = (A, B; B^T, \Sigma)$, where $A \in \mathbb{R}^{6(N-1)\times 6(N-1)}$, $B \in \mathbb{R}^{6(N-1)\times 3K}$, and $\Sigma \in \mathbb{R}^{3K\times 3K}$. Due to the structure in $\boldsymbol{g}_{ij}$, it is easy to check that Σ is a block-diagonal matrix, whose inverse can be efficiently computed. As a result, one can compute C_{XX}^{-1} effectively using Schur complement (c.f. [Zhang 2005]). We left the technical details to the supplemental material. Algorithm 2 shows the pseudo-code of our uncertainty quantification approach under SRAR.

ALGORITHM 2: Uncertainty quantification algorithm for simultaneous registration and reconstruction.

Input: Optimized scan point positions $\boldsymbol{p}_{ij}^{\star}$, $1 \le i \le N, 1 \le j \le N_i$ and latent planes $(\boldsymbol{n}_k^{\star}, d_k^{\star})$, $1 \le k \le K$.
Output: Diagonal blocks and leading eigenvectors of C_{SRAR}
Step 1: Compute C_{XX};
Step 2: Use (22) to compute the diagonal blocks and leading eigenvectors of C_{SRAR};

6.2.3 Uncertainty Visualization. Uncertainty visualization employs the same visualization schemes described in Section 5.3. As shown in Figure 6(Left), our predicted covariance matrices are consistent with the simulated covariance matrices. Another observation is that the predicted diagonal blocks of SRAR have smaller norms than those of JPR (See Figure 6(Right)). An interpretation is that

JPR is sensitive to the accuracy of point normals. In contrast, SRAR does not require point normals.

7 UNCERTAINTY-DRIVEN VIEW-PLANNING

In this section, we introduce a model-based view-planning approach. Generally speaking, a model-based view-planner takes a 3D model as input and outputs a set of camera poses. These camera poses typically optimizes a two-term objective function (c.f. [Scott et al. 2003]), where the first term characterizes a coverage score of the camera poses and where the second term measures the length of the shortest path that connects the camera poses. Our model-based view planner replaces the coverage score by an uncertainty score derived from the approach described in Section 6. This formulation allows us to plan views to minimize the reconstruction error of the resulting scans directly. In the following, we first describe the uncertainty score in Section 7.1. We then describe our model-based view planner in Section 7.2.

7.1 An Uncertainty Score

This section defines an uncertainty score among scans $S_i, 1 \le i \le n$ of a model M, where S_i are generated from pre-defined camera poses C_i using the noisy model described in Section 6. To model the process of selecting a subset of scans for reconstruction, we consider a reweighted total objective function under SRAR:

$$f_{\text{weighted}} := \sum_{i=1}^{n} w_i \sum_{j=1}^{n_i} \left((R_i \boldsymbol{p}_{ij} + \boldsymbol{t}_i)^T \boldsymbol{n}_{k_{ij}} - d_{k_{ij}}\right)^2 \tag{23}$$

where w_i is an indicator that specifies whether S_i is selected or not. We seek to set w_i as either 0 or 1.

Applying (5) to (23), we obtain the following approximated covariance matrix among both the scan poses and the reconstruction:

$$C_{\text{SRAR}}(\boldsymbol{w}, \boldsymbol{p}^{gt}) := \left(\sum_{i=1}^{n} w_i G_i G_i^T\right)^{-1} \left(\sum_{i=1}^{n} w_i^2 G_i G_i^T\right) \left(\sum_{i=1}^{n} w_i G_i G_i^T\right)^{-1} \tag{24}$$

where $G_i = (\boldsymbol{g}_{i1}, \cdots, \boldsymbol{g}_{iN_i})$ collects vectors g_{ij} associated with S_i in its columns. Note that since the ground-truth model M is present, we define $C_{\text{SRAR}}(\boldsymbol{w}, \boldsymbol{p}^{gt})$ with respect to the ground-truth scan points, i.e., each S_i stays on M.

The following theorem states an important property of $C_{\text{SRAR}}(\boldsymbol{w}, \boldsymbol{p}^{gt})$, which sheds lights on the proposed uncertainty score that we will introduce.

Theorem 1. *Suppose*

$$\sum_{i=1}^{n} w_i G_i G_i^T \succ 0, \quad \sum_{i=1}^{n} G_i G_i^T \succ 0.$$

Then

$$C_{\text{SRAR}}(\boldsymbol{w}, \boldsymbol{p}^{gt}) \succeq \left(\sum_{i=1}^{n} G_i G_i^T\right)^{-1}. \tag{25}$$

In particular, equality holds if $w_1 = w_2 = \cdots = w_n$.

Proof: Please refer to the supplemental material. □

Remark 4. *Theorem 1 states that adding more scans always reduces variance, i.e., when scan weights of active scans are identical. This property motivates us to use $C(\boldsymbol{w}, \boldsymbol{p}^{gt})$ to define the quality score.*

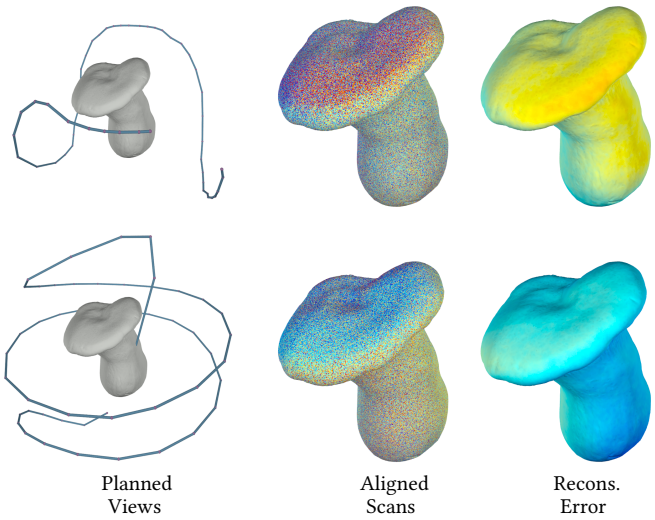

Fig. 7. This figure shows the results of our model-based view planner on a mushroom model. We placed 1000 camera poses among a viewing sphere centered at the barycenter of the input model. The planned trajectory optimizes 50 views. For each block, we show the planned trajectory, the resulting aligned scans, and the color-coded reconstruction error. (Top) The result obtained from setting $\lambda = 100$ in (28). (Bottom) The result obtained from setting $\lambda = 1$ in (28).

Remark 5. *On the other hand, one should interpret this result from the perspective that assuming global convergence, adding more scans leads to more accurate alignments and reconstructions. The issue of convergence radii of SRAR, which is related to the condition number of $C(\boldsymbol{w}, \boldsymbol{p}^{gt})$ (c.f. [Nocedal and Wright 2006]), appears to be quite different. Similar to geometrically stable sampling for pair-wise registration [Gelfand et al. 2003], one may apply sampling to adjust the performance of SRAR. However, we found using uniform sample weights does not incur any convergence issue for all the datasets tested in this paper.*

We proceed to define the quality score as a suitable function of the covariance matrix.

Corollary 1. *Consider a broad range of matrix functions $\mathcal{H}$ that include polynomials of eigenvalues of principal sub-matrices with non-negative weights, e.g., trace or determinant of a principal sub-matrix. Then for each $h \in \mathcal{H}$, we have*

$$h(C_{\text{SRAR}}(\boldsymbol{w}, \boldsymbol{p}^{gt})) \ge h((\sum_{i=1}^{n} G_i G_i^T)^{-1}). \tag{26}$$

Following Corollary 1, we define the uncertainty score for view-planning as the trace of the sub-matrix $C_{\text{Pose}}(\boldsymbol{w}, \boldsymbol{p}^{gt}) \in \mathbb{R}^{6(N-1)\times 6(N-1)}$ of $C_{\text{SRAR}}(\boldsymbol{w}, \boldsymbol{p}^{gt})$ that corresponds to scan poses:

$$f_{\text{uncertainty}}(\boldsymbol{w}) := \text{Trace}(C_{\text{Pose}}(\boldsymbol{w}, \boldsymbol{p}^{gt})). \tag{27}$$

7.2 A Greedy Uncertainty-Driven View Planner

Besides $f_{\text{uncertainty}}(\boldsymbol{w})$, which evaluates the reconstruction quality from a subset of scans specified by $\boldsymbol{w}$, another essential goal in

model-based view planning is to minimize the length of the path that connects the selected views. While this quantity, i.e., a solution to the travel salesman problem (or TSP), admits integer programming relaxations. We found that it is challenging to solve the induced global optimization problem. Fundamental challenges are 1) the linear programming relaxations are loose, and 2) rounding fractional solutions into integer solutions effectively. To address these issues, we utilize a greedy approach in this paper.

Specifically, let $f_{\text{length}}(\boldsymbol{w})$ denote the length of the trajectory among selected cameras specified by $\boldsymbol{w}$. We use a greedy approach to determine the best binary indicator vector $\boldsymbol{w}$ that minimizes the following cumulative objective function:

$$\min_{\boldsymbol{w}\in\{0,1\}^n} f_{\text{uncertainty}(\boldsymbol{w})} + \lambda f_{\text{length}}(\boldsymbol{w}) \tag{28}$$

where λ is a tradeoff-parameter that balances the reconstruction error and the length of the TSP path. Starting from $w_i = 1, 1 \leq i \leq n$, we calculate the value of the objective function in (28) when removing each scan. We then remove the scan that leads to the smallest objective value after removing that scan. This procedure is iterated when we cannot reduce the objective function further. In our implementation, we use the Christofides's algorithm [Christofides 1976] for computing $f_{\text{length}}(\boldsymbol{w})$.

In practice, we found that this simple strategy leads to high-quality solutions. The total running time for a model with 100K latent planes and 500 camera poses is between 110 seconds to 210 seconds on a desktop with a dual-core 3.4G Hz GPU and 128G main memory. As shown in Figure 7, increasing the value of λ leads to smoother paths but with increased reconstruction errors.

8 EXPERIMENTAL EVALUATION

This section presents experimental evaluations of the proposed UQ and view-planning methods. In Section 8.1 and Section 8.2, we describe experimental evaluations of our UQ approaches on synthetic examples and real examples, respectively. The goal of evaluating on synthetic examples is to assess the approximation errors of the proposed UQ techniques empirically. In contrast, experimental evaluations on real examples study whether the quantified uncertainties match the pose and reconstruction errors compared to the underlying ground-truth. In Section 8.3, we present a user study to gauge the effectiveness of the proposed UQ approaches on model checking. In Section 8.4, we present a user study on the model-based view-planning approach described in Section 7.

8.1 Experimental Evaluations on Synthetic Datasets

In this section, we first describe the experimental setup. We then analyze the experimental results on approximation errors of the proposed UQ approaches. We also quantitatively compare the accuracies of the two multi-scan registration strategies discussed in this paper, i.e., SRAR and JPR.

8.1.1 Experimental Setup. We have collected a synthetic dataset that consists of four categories of models (two models per category), namely, Plant, Engine, Airplane, and Bridge (See Figure 8). These models cover a diverse range of model scales. 3D reconstructions of these models have direct scientific and industrial impacts. For example, reconstructing plants over time helps us to understand the internal mechanisms of biological forms (c.f. [Li et al. 2013]). In this scenario, quantified uncertainties of the reconstructions are critical for analyzing growing patterns. The Engine, Airplane, and Bridge categories collect diverse man-made objects, where uncertainty quantification can significantly enhance all phases of interactive scannings, such as view planning and quality checking. The resulting accurate reconstructions offer various downstream applications, e.g., damage detection for engines, airplanes, and bridges.

The camera locations of the synthetic scans are uniformly among the viewing sphere centered at the origin of each model. We set the viewing sphere's radius as a random number between two times and four times the diameter of each model. Moreover, the upright direction of the camera always passes through the upright orientation of each model. In total, we place 50 scans for Plant and Engine models, 200 scans for Airplane models, and 500 scans for Bridge models (the number of scans increases as the model size increases).

When generating the input scans, we choose three noise ratios, i.e., $\epsilon = \frac{L}{1000}, \frac{L}{400}, \frac{L}{200}$, where L is the diameter of that model in the Euclidean distance. At each level, we place 10000 sets of depth scans independently. Each scan point is independently perturbed along the viewing direction by a random variable in $[-\epsilon, \epsilon]$. We use these sets to compute both the simulated covariance matrix and the predicted covariance matrices. Note that although the scans deviate from one set to another, their overlapping patterns are fixed.

When generating the input to JPR and SRAR, we randomly perturb the pose of each input scan by a random rotation specified by $\exp(-\boldsymbol{c}\times)$ (where $\boldsymbol{c}$ is a random vector in $[-0.02, 0.02]^3$) and a random translation in $[-4\epsilon, 4\epsilon]^3$. Since the ground-truth correspondences are not given in these experiments, they also evaluate the discretization errors of the proposed approaches.

8.1.2 Analysis of Approximation Errors. We first analyze the approximation errors of our UQ approaches by comparing the diagonal blocks and the leading eigenvectors of the simulated covariance matrix with those predicted by our approach. For each predicted diagonal block A_{pred} and that obtained from the simulated co-variance matrix A_{simul}, we report the relative error as $\frac{\|A_{\text{simul}} - A_{\text{pred}}\|}{\|A_{\text{simul}}\|}$. We then plot the cumulative distribution of relative errors for each approach per model.

The evaluation protocol for the leading eigenspace is similar. For each eigenvector $\boldsymbol{v}$ of a simulated covariance matrix $\overline{C}$, we first compute $\boldsymbol{v}^{\text{perp}} = \text{argmin}_{\boldsymbol{v}^\star \in \mathcal{U}_C} \|\boldsymbol{v}^\star - \boldsymbol{v}\|^2$, where $\mathcal{U}_C$ is the leading eigenspace of the corresponding predicted covariance matrix. Then for all the leading eigenvectors $\boldsymbol{v}$, we collect statistics on the relative errors for all the corresponding sub-vectors of $\boldsymbol{v}$ and $\boldsymbol{v}^{\text{perp}}$, i.e., $\|\boldsymbol{v}_i^{\text{perp}} - \boldsymbol{v}_i\|/\|\boldsymbol{v}_i\|$. Intuitively, these relative errors characterize the coverage of the leading eigenspace of the predicted covariance matrix C on that of the simulated covariance matrix $\overline{C}$. They are zero if and only if these two eigenspaces coincide. The same as diagonal blocks, we plot the cumulative distribution of relative errors for all eigenvectors of the simulated covariance matrix.

The sixth and the seventh columns of Figure 8 show cumulative distributions of the approximation errors of the predicted diagonal blocks and the predicted eigenspace. Note that we conducted the experiments under both JPR and SRAR while varying the noise levels

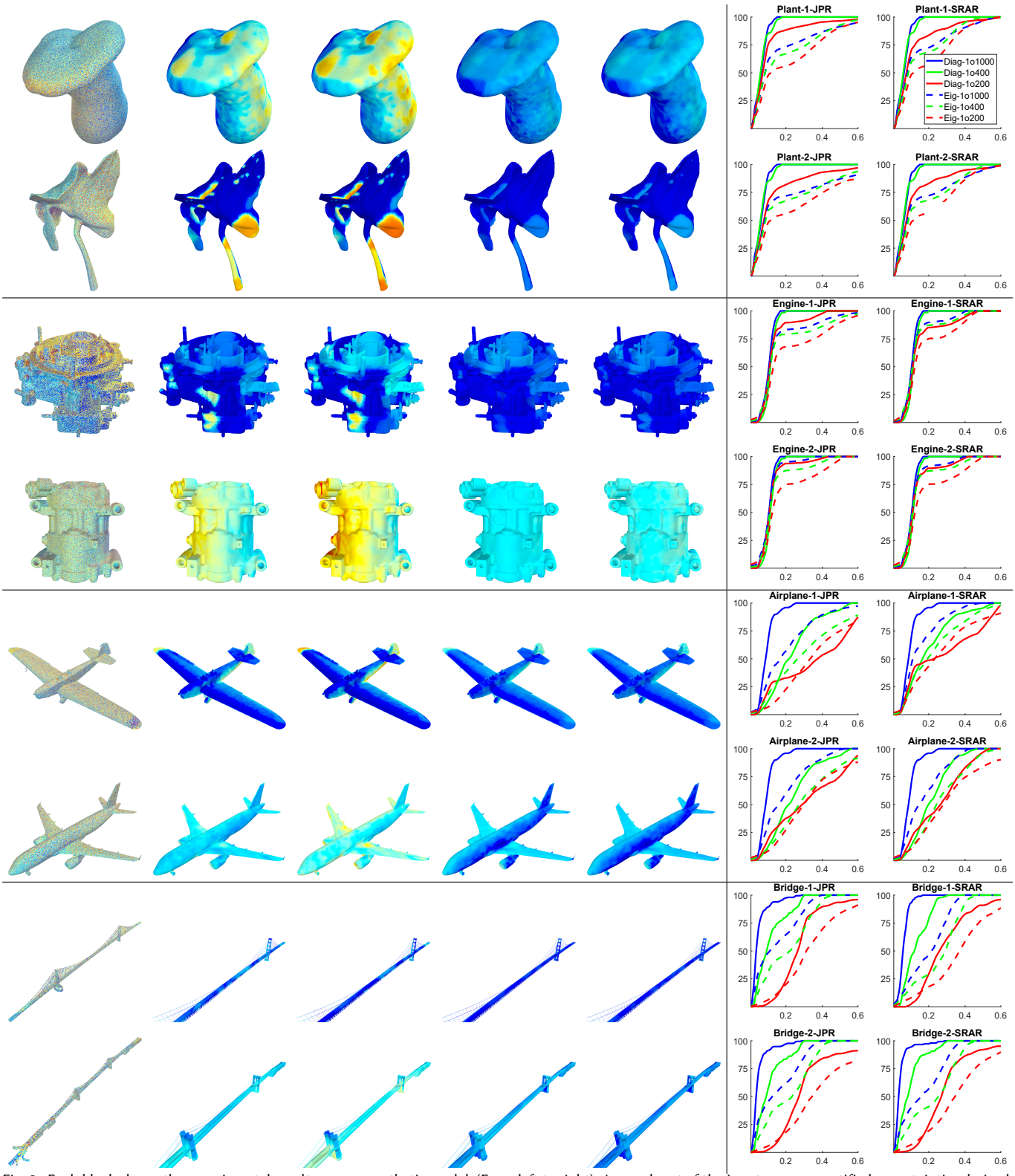

Fig. 8. Each block shows the experimental results on one synthetic model. (From left to right): A sample set of the input scans, quantified uncertainties derived from the diagonal blocks of the predicted covariance matrix under JPG, corresponding quantified uncertainties of the simulated covariance matrix under JPG, quantified uncertainties derived from the diagonal blocks of the predicted covariance matrix under SRAR, corresponding quantified uncertainties of the simulated covariance matrix under SRAR, approximation errors of our approach under JPR, and approximation errors of our method under SRAR.

of the input scans. Overall, our UQ approaches offer accurate predictions of the underlying uncertainty measures. For JPR, the mean relative errors range from 0.108 to 0.171 for diagonal blocks and from 0.158 to 0.216 for the leading eigenspace, respectively. For SRAR, the mean relative errors range from 0.117 to 0.182 for diagonal blocks and from 0.149 to 0.228 for the leading eigenspace, respectively. These quantitative results are consistent with the qualitative results (from the second to the fifth columns of Figure 8), which demonstrate that the predicted results offer meaningful approximations of the corresponding simulated ones.

We can also observe that the relative errors increase when the noise level increases. Moreover, the relative errors increase when the shapes become thin and long. For example, the approximation errors on Airplane and Bridge are more significant than those on Plant and Engine. Intuition is that uncertainties on Airplane and Bridge are more prominent, and the linear map approximation is less accurate. However, even at the highest noise level, the relative errors of the diagonal blocks and the leading eigenspace are below 0.250 under both settings. Such results justify that uncertainty measures derived from diagonal blocks and leading eigenvectors of the predicted covariance matrix are robust, informative, and efficient.

8.1.3 Multi-Scan Registration Comparison. Another interesting question is: how do the two registration methodologies compare against each other in terms of the registration error? As compared in the third and the fifth columns of Figure 8, the approach of simultaneous registration and reconstruction (or SRAR) yields better registration results than joint pairwise registration (or JPR). The simulated uncertainties of SRAR are consistently smaller than those of JPR. On Airplane and Bridge, the performance of SRAR is considerably better than JPR. Again, the advantage of SRAR is that it does not utilize point normals, which may be noisy.

8.2 Experimental Evaluations on Real Datasets

In this section, we evaluate the proposed UQ approaches on real datasets. Similar to Section 8.1, we first describe the experimental setup. We then analyze the experimental results of UQ and compare the two registration approaches.

8.2.1 Experimental Setup. We have collected a real benchmark dataset that consists of four categories of models, namely, Teeth, MechPart, ScanNet [Dai et al. 2017a], and Waymo [Sun et al. 2019] (See Figure 9). Similar to the synthetic dataset, 3D reconstructions of these models enjoy rich medical and industrial applications. Applications of 3D teeth reconstructions include surgery planning for implanting and orthodontics [Yau et al. 2014]. Both applications benefit from certified 3D reconstructions as input. 3D reconstructions of mechanical parts are used for model inspection, in which quantified uncertainties of 3D reconstructions enable us to evaluate the manufacturing error. ScanNet and Waymo are used to assess the performance of our approach to large-scale scenes. Such reconstructions are essential for applications of virtual reality and 3D mapping.

Each model in this dataset is associated with a set of input scans and ground-truth poses and reconstructions. The scans of Teeth and MechPart come from an Artec Space Spider scanner. The number of scans per model ranges from 300 to 500. For both Teeth and MechPart, the underlying ground-truth models come from an Artec Micro high precision industrial scanner. We obtain the ground-truth pose of each input scan by aligning it with the underlying ground-truth model. For ScanNet and Waymo, we use the scans provided by the original datasets. Similar to the synthetic dataset, the overlapping graph among the input scans of each model exhibits great diversity, e.g., from sparse graphs to dense graphs.

For Teeth and MechPart, we use the ground-truth poses and reconstructions described above for experimental evaluation. Since scenes of ScanNet do not possess ground-truth reconstructions, we utilize planar areas of each scene and evaluate deviations of the data points in each planar area from the corresponding fitted plane. For models in the Waymo dataset, we directly assess the pose error using the provided ground-truth poses.

8.2.2 Analysis of UQ Results. The setting of real scans differ from that of synthetic scans in two ways: 1) we do not know the noise model for generating the input scans, and 2) we do not have an exact underlying ground-truth. Therefore, we employ a different evaluation protocol for quantitative evaluation. Specifically, for each reconstruction, we compare it against the approximated underlying ground truth (i.e., obtained using the high-precision industrial scanner) to pose errors and shape errors. We then evaluate whether the quantified uncertainties are consistent with these errors, as detailed below.

We first consider the cumulative distribution of the relative difference between the ground-truth pose error described above and the predicted variance vector associated with each scan pose. The statistics collect the results of 10000 experiments, i.e., we pick 200 subsets of scans per model, where each subset randomly picks 90% of the input scans 50 times. Note that given two vectors $\boldsymbol{a}, \boldsymbol{b}$, their relative difference is given by $\|\boldsymbol{a} - \boldsymbol{b}\|/\max(\|\boldsymbol{a}\|, \|\boldsymbol{b}\|)$. As shown in the third column of Figure 9, most normalized differences fall between 0.042 to 0.225. In other words, the predicted pose variance per scan nicely indicates the underlying ground-truth pose error per scan.

We proceed to analyze whether our UQ approaches capture the structural patterns among the ground-truth pose errors. To this end, we compute the L2 projection of the ground-truth pose error vector onto the leading eigenspace computed using our method. We then calculate the relative projection error between each pose error vector and its projection. As shown in the fourth column of Figure 9, most projection errors fall between 0.052 to 0.238. In other words, the predicted leading eigenspace provides a meaningful approximation of the simulated eigenspace.

8.2.3 Multi-Scan Registration Comparison. As shown in Figure 9, the pose error of SRAR is usually smaller than that of JPR. For example, the mean rotation/translation errors when using JPR are $0.29^\circ/59micron$, $0.32^\circ/87micron$, and $0.51^\circ/32mm$ on Teeth, MechPart, and Waymo, respectively. Likewise, the corresponding errors when using SRAR are $0.19^\circ/27micron$, $0.27^\circ/39micron$, and $0.31^\circ/13mm$, respectively. The relative error reductions range from 12.1% to 52.3% on these datasets.

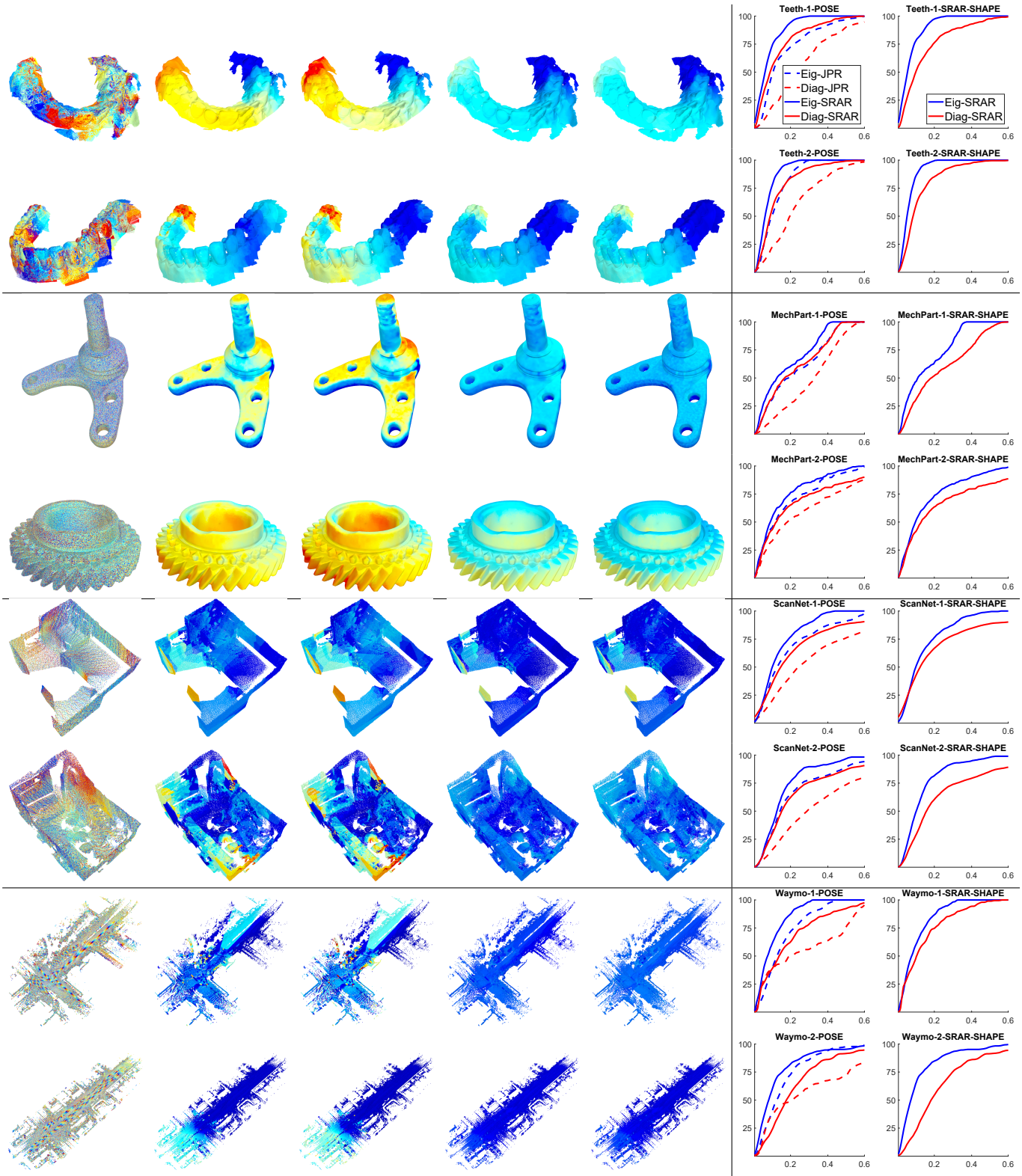

Fig. 9. Each block shows the experimental results on one real model. (From left to right): A sample of the input scans, quantified uncertainties derived from the diagonal blocks of the predicted covariance matrix under JPG, the ground-truth reconstruction error under JPG, quantified uncertainties under SRAR, the ground-truth reconstruction error under SRAR, approximation errors of the pose components of the covariance matrices (JPR and SRAR), and approximation errors of the shape component of the covariance matrix (SRAR).

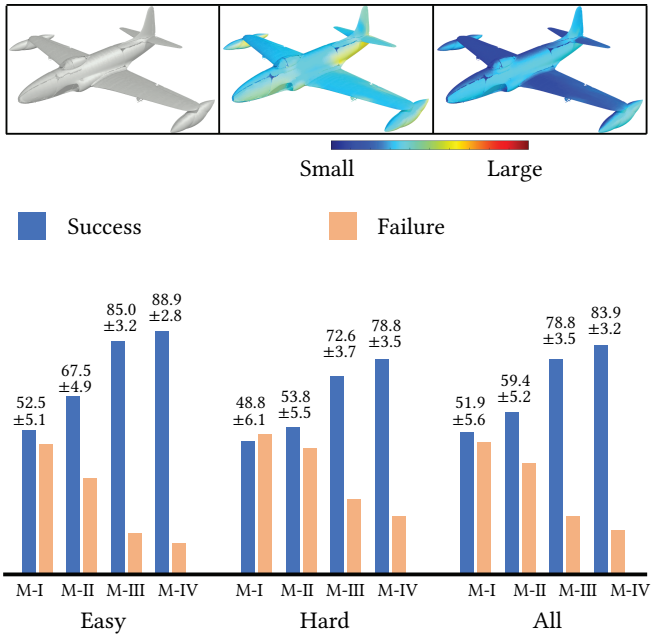

Fig. 10. Quantitative evaluations of the proposed uncertainty quantification approach for assessing reconstruction quality. (Top) Illustration of the user interface, which shows from the original model (left) and two reconstructions (right). The reconstructions are chosen so that one is better than the other in terms of the mean reconstruction error. (Bottom) Statistics of the user study results. We show the average result among an Easy subset (the mean error of one reconstruction is smaller than half of the other one), among a Hard subset (remaining ones), and across the entire dataset (i.e., Overall). The reconstructions are shown under four modes (from left to right in each block), namely, (M-I) raw reconstruction, (M-II) reconstruction color-coded by a local registration error, (M-III) reconstruction color-coded by quantified uncertainties, and (M-IV) reconstruction color-coded by the ground-truth reconstruction error. For each instance, a user is asked to choose the reconstruction with the smaller mean error. Statistics collect the mean and variance among all the users.

8.3 User Study on UQ for Model Checking

We proceed to evaluate the effects of the proposed UQ approaches for model checking, i.e., whether they help users to understand the quality of a 3D reconstruction in interactive scanning.

8.3.1 Experimental Setup. For quantitative evaluation, we have designed a user interface that consists of three sub-windows (See Figure 10). The left sub-window shows the ground-truth model, which mimics images of the underlying physical model. The middle and right sub-windows show two reconstructions of the underlying model, where the mean error of one reconstruction is smaller than that of the other one. We employed Poisson Surface Reconstruction [Kazhdan et al. 2006] to obtain the reconstructions. For JPR, the reconstruction applies to the aligned data points. For SRAR, the reconstruction applies to the optimized latent surface.

Each pair of reconstructions is displayed in four modes:

- *M-I: Raw reconstruction.* The first mode merely displays raw reconstructions without additional information. The user judges which one is better via visual comparisons with the underlying ground-truth.
- *M-II: Local alignment error.* The second mode shows the mean distances between the aligned point clouds and the reconstructed model. We collect statistics on the reconstructed model using the approach described in Section 5.3.
- *M-III: Results of our approach.* The third mode visualizes diagonal blocks (the default option) and leading eigenvectors of the predicted covariance matrix using the method described in Section 5.3.
- *M-IV: Ground-truth reconstruction error.* The last mode shows the ground-truth reconstruction error by measuring the closest distance between each point of the reconstruction and the underlying ground-truth reconstruction.

The reconstructions conducted in this study are generated by running SRAR and JPR on the synthetic dataset. We sample pairs of 3D reconstructions, compute mean reconstruction errors, and sample reconstruction pairs. In total, we collected 400 reconstruction pairs for the user study (200 using JPR and the other 200 using SRAR). Half of them form an Easy subset where the mean error of one reconstruction is smaller than half of another reconstruction. The remaining pairs form the Hard subset.

8.3.2 Analysis of Results. Figure 10 shows the results of a user study among 20 AMT users. We can see that when only displaying the rough reconstructions, picking the reconstruction with the smaller error is close to a random guess. This behavior is consistent among both the Easy subset and the Hard subset. In other words, it is critical to provide users with additional information to make accurate judgments. The local registration error is useful to some extent, yielding 67.5%/32.5% (±4.9%), 53.8%/46.2% (±5.5%), and 59.4%/40.6% (±5.2%) success/failure rates on Easy, Hard, and Overall, respectively. In other words, the local registration error may indicate the reconstruction error when one reconstruction is significantly better than the other reconstruction. On the other hand, it is ineffective for instances in the Hard subset.

In contrast, our approach yields far better results. The success/failure rates on settings of Easy, Hard, and Overall are 85.0%/15.0% (±3.2%), 72.6%/27.4% (±3.7%), and 78.8%/21.2% (±3.5%), respectively. The performance matches the results of using ground-truth reconstruction errors, i.e., 88.9%/11.1% (±2.8%), 78.8%/21.2% (±3.5%), and 83.9%/16.1% (±3.2%), respectively. Note that using the ground-truth reconstruction errors do not offer perfect results, as salient error regions of some reconstruction pairs are different, making it difficult to judge which one has a smaller mean error.

Question	Percentages
Effective indication of global drifts?	94%
Effective indication of reconstruction errors?	88%
Eigenvectors indicate uncertainty structures?	85%

Table 1. A questionnaire for assessing our UQ approach.

We have conducted an additional questionnaire to gauge the effectiveness of our visualization schemes (See Table 1). For example, 94% of the users indicated that our UQ approach is useful for revealing

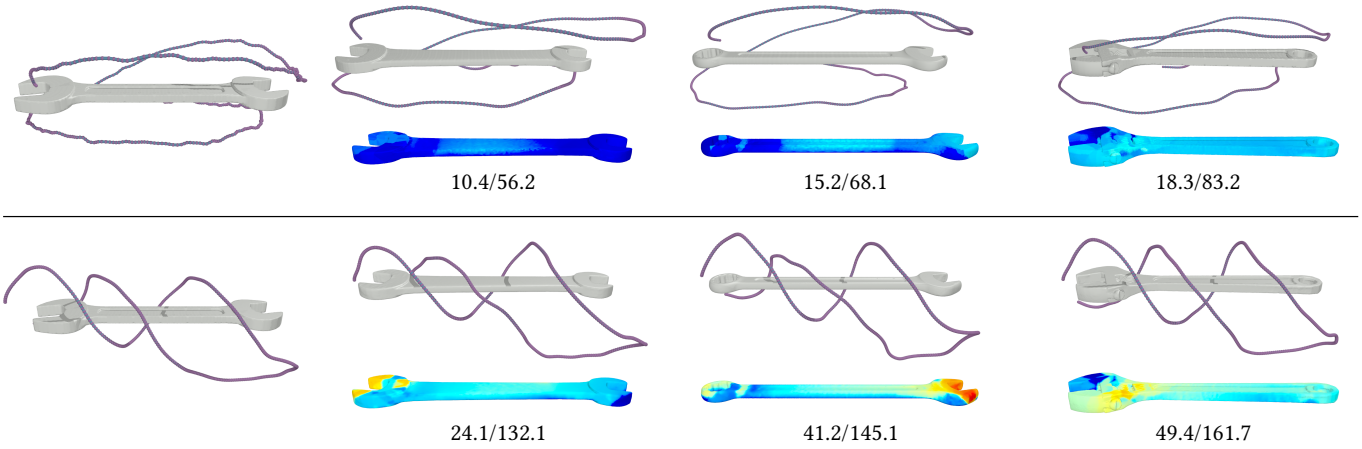

Fig. 11. This figure shows model-based view planning on a Wrench dataset that consists of a reference model (Left) and three testing objects. The ground-truth models of these objects are reconstructed using an Artec Micro scanner. We perform view-planning on the reference model, and the results guide users to scan similar objects using an Artec Space Spider scanner. The top row shows the view planning results of our approach. The bottom row shows the baseline approach that solves a set cover problem [Scott et al. 2003]. The scanning trajectory is displayed concerning the ground-truth model of each object. For each model, we show the scanning trajectory on the top and the color-coded reconstruction errors on the bottom (mean/max in micros).

global drifts. Moreover, 88% of the users reported that the output of our approach reflects magnitudes and patterns of reconstruction errors. Finally, 85% of the users commented that visualizations of eigenvectors indicate structures in the quantified uncertainties. All percentages are above 85%, and they justify the effectiveness of our approach.

8.4 Uncertainty-Driven View Planning

In this section, we present an experimental analysis of the proposed uncertainty-driven view planning approach under the framework of SRAR.

8.4.1 Experimental Setup. We perform experimental evaluations under the setting of category-specific view planning. Specifically, consider the task of scanning a physical object whose category label is known. We use a similar 3D model that belongs to the same category to compute the scanning trajectory. We then display the planned path to a user for interactive scanning. Note that the planned trajectory only suggests how to scan a new object. We find that users do not precisely follow the intended trajectory.

We have performed an experimental study on a Wrench category that consists of one reference model and three testing objects. Note that the differences between the testing objects and the reference model vary, and such differences help assess the generalization ability of our approach. Interactive scanning employed an Artec Space Spider scanner. To generate the planned trajectory, we placed 1000 camera poses around the reference model. The distance from each camera origin to the reference model is 2cm, the viewing direction of each camera targets at the closest point. We run the algorithm described in Section 7 to generate the planned trajectory.

8.4.2 Analysis of View-Planning. For a baseline comparison, we pick a state-of-the-art model-based view planning approach [Scott 2009], which applies a set cover method to optimize a visibility-driven score among the selected views. As shown in Figure 11, our UQ-driven view planner delivers results that are better than the baseline approach. The mean/max reconstruction errors (in micros) on the three testing objects are 10.4/56.2, 15.2/68.1, and 18.3/83.2, respectively. In contrast, the mean/max reconstruction errors (in micros) of the baseline approach are 24.1/132.1, 41.2/145.1, and 49.4/161.7, respectively. Visually, our approach offers much evenly distributed reconstruction errors than the baseline approach, which exhibits certain global drifts. These results show the advantage of modeling uncertainties of geometry reconstruction for view planning.

For both approaches, the reconstruction errors increase when the differences between the testing model and the reference model escalate. However, our method offers a smaller variance among the three testing models than the baseline approach. These statistics show that the uncertainty score, which considers how scans overlap, is more stable than the visibility score for characterizing the reconstruction quality.

9 CONCLUSIONS AND FUTURE WORK

In this paper, we have studied the problem of uncertainty quantification for multi-scan registration. Our approach examines the approximation errors of utilizing the inverse of the Hessian matrix of the objective function to estimate the covariance matrix among the output derived from minimizing this objective function. This analysis results in robust and informative uncertainty measures. We have applied this approach under two settings of multi-scan registration, namely, minimizing pairwise distances between overlapping scans and aligning the input scans with a deforming template. Under both settings, we showed how to formulate objective functions that possess small discretization errors. Our uncertainty quantification methods also allowed us to compare these two formulations, both empirically and theoretically. Experimental results

on a variety of synthetic and real datasets and user studies on real datasets demonstrate the usefulness of our approaches. Finally, we showed the effectiveness of our uncertainty measures for the task of model-based view planning.

There are ample opportunities for future research. For example, it would be interesting to explore how to extend the linear approximation scheme employed in this paper to non-linear schemes, e.g., by introducing high-order components of the Taylor expansion. Besides, we would also like to incorporate non-rigid scan registration (that can address scan warping) into our formulation. Finally, it would be interesting to investigate uncertainty quantification for other geometry processing tasks such as shape editing, shape modeling, and reverse engineering.

ACKNOWLEDGMENTS

We thank the anonymous reviewers for their valuable comments. Qixing Huang would like to acknowledge support from NSF DMS-1700234, NSF CIP-1729486, NSF IIS-1618648, a gift from Snap Research and a GPU donation from Nvidia Inc.

REFERENCES

Federica Arrigoni, Beatrice Rossi, and Andrea Fusiello. 2016a. Global Registration of 3D Point Sets via LRS Decomposition. In *Computer Vision - ECCV 2016 - 14th European Conference, Proceedings, Part IV*. Springer Verlag, Amsterdam, The Netherlands, 489–504. https://doi.org/10.1007/978-3-319-46493-0_30

Federica Arrigoni, Beatrice Rossi, and Andrea Fusiello. 2016b. Spectral Synchronization of Multiple Views in SE(3). *SIAM J. Imaging Sciences* 9, 4 (2016), 1963–1990. https://doi.org/10.1137/16M1060248

Ola Bengtsson and Albert-Jan Baerveldt. 2003. Robot localization based on scan-matching - estimating the covariance matrix for the IDC algorithm. *Robotics Auton. Syst.* 44, 1 (2003), 29–40. https://doi.org/10.1016/S0921-8890(03)00008-3

Raouf Benjemaa and Francis Schmitt. 1998. A Solution for the Registration of Multiple 3D Point Sets Using Unit Quaternions. In *Proceedings of the 5th European Conference on Computer Vision-Volume II - Volume II (ECCV âĂŹ98)*. Springer-Verlag, Berlin, Heidelberg, 34âĂŞ50.

Matthew Berger, Andrea Tagliasacchi, Lee M. Seversky, Pierre Alliez, Joshua A. Levine, Andrei Sharf, and Cláudio T. Silva. 2014. State of the Art in Surface Reconstruction from Point Clouds. In *Eurographics 2014 - State of the Art Reports*. The Eurographics Association, Strasbourg, France, 161–185. https://doi.org/10.2312/egst.20141040

Paul J. Besl and Neil D. McKay. 1992. A Method for Registration of 3-D Shapes. *IEEE Trans. Pattern Anal. Mach. Intell.* 14, 2 (Feb. 1992), 239âĂŞ256. https://doi.org/10.1109/34.121791

Peter Biber and Wolfgang Straßer. 2003. The normal distributions transform: a new approach to laser scan matching. In *2003 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE/RSJ, Las Vegas, Nevada, USA, 2743–2748. https://doi.org/10.1109/IROS.2003.1249285

Michael Bosse and Robert Zlot. 2008. Map Matching and Data Association for Large-Scale Two-dimensional Laser Scan-based SLAM. *I. J. Robotics Res.* 27, 6 (2008), 667–691. https://doi.org/10.1177/0278364908091366

Benedict J. Brown and Szymon Rusinkiewicz. 2007. Global Non-Rigid Alignment of 3-D Scans. *ACM Trans. Graph.* 26, 3 (July 2007), 21âĂŞes. https://doi.org/10.1145/1276377.1276404

Andrea Censi. 2007. An accurate closed-form estimate of ICP's covariance. In *2007 IEEE International Conference on Robotics and Automation, ICRA 2007*. IEEE, Roma, Italy, 3167–3172. https://doi.org/10.1109/ROBOT.2007.363961

Jiawen Chen, Dennis Bautembach, and Shahram Izadi. 2013. Scalable Real-Time Volumetric Surface Reconstruction. *ACM Trans. Graph.* 32, 4, Article Article 113 (July 2013), 16 pages. https://doi.org/10.1145/2461912.2461940

Yang Chen and Gérard Medioni. 1992a. Object Modelling by Registration of Multiple Range Images. *Image Vision Comput.* 10, 3 (April 1992), 145–155. https://doi.org/10.1016/0262-8856(92)90066-C

Yang Chen and Gérard G. Medioni. 1992b. Object modelling by registration of multiple range images. *Image Vision Comput.* 10, 3 (1992), 145–155. https://doi.org/10.1016/0262-8856(92)90066-C

Nicos Christofides. 1976. *Worst-case analysis of a new heuristic for the travelling salesman problem.* Technical Report 388. Graduate School of Industrial Administration, Carnegie Mellon University.

Angela Dai, Angel X. Chang, Manolis Savva, Maciej Halber, Thomas A. Funkhouser, and Matthias Nießner. 2017a. ScanNet: Richly-Annotated 3D Reconstructions of Indoor Scenes. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*. IEEE, Honolulu, HI, USA, 2432–2443. https://doi.org/10.1109/CVPR.2017.261

Angela Dai, Matthias Nieundefinedner, Michael Zollhöfer, Shahram Izadi, and Christian Theobalt. 2017b. BundleFusion: Real-Time Globally Consistent 3D Reconstruction Using On-the-Fly Surface Reintegration. *ACM Trans. Graph.* 36, 3, Article Article 24 (May 2017), 18 pages. https://doi.org/10.1145/3054739

Chandler Davis and W. M. Kahan. 1970. The Rotation of Eigenvectors by a Perturbation. III. *SIAM J. Numer. Anal.* 7, 1 (1970), 1âĂŞ46.

Xinyi Fan, Linguang Zhang, Benedict Brown, and Szymon Rusinkiewicz. 2016. Automated View and Path Planning for Scalable Multi-Object 3D Scanning. *ACM Trans. Graph.* 35, 6, Article Article 239 (Nov. 2016), 13 pages. https://doi.org/10.1145/2980179.2980225

Natasha Gelfand and Leonidas J. Guibas. 2004. Shape Segmentation Using Local Slippage Analysis. In *Second Eurographics Symposium on Geometry Processing, July 8-10, 2004*. The Eurographics Association, Nice, France, 214–223. https://doi.org/10.2312/SGP/SGP04/219-228

Natasha Gelfand, Szymon Rusinkiewicz, Leslie Ikemoto, and Marc Levoy. 2003. Geometrically Stable Sampling for the ICP Algorithm. In *4th International Conference on 3D Digital Imaging and Modeling (3DIM 2003), 6-10 October 2003, Banff, Canada*. IEEE Computer Society, Banff, Canada, 260–267. https://doi.org/10.1109/IM.2003.1240258

Venu Madhav Govindu and A. Pooja. 2014. On Averaging Multiview Relations for 3D Scan Registration. *IEEE Trans. Image Processing* 23, 3 (2014), 1289–1302.

Peter Henry, Michael Krainin, Evan Herbst, Xiaofeng Ren, and Dieter Fox. 2012. RGB-D mapping: Using Kinect-style depth cameras for dense 3D modeling of indoor environments. *I. J. Robotics Res.* 31, 5 (2012), 647–663. https://doi.org/10.1177/0278364911434148

Benjamin Hepp, Matthias Nieundefinedner, and Otmar Hilliges. 2018. Plan3D: Viewpoint and Trajectory Optimization for Aerial Multi-View Stereo Reconstruction. *ACM Trans. Graph.* 38, 1, Article Article 4 (Dec. 2018), 17 pages. https://doi.org/10.1145/3233794

Qixing Huang, Bart Adams, and Michael Wand. 2007a. Bayesian surface reconstruction via iterative scan alignment to an optimized prototype. In *Proceedings of the Fifth Eurographics Symposium on Geometry Processing, 2007*. The Eurographics Association, Barcelona, Spain, 213–223. https://doi.org/10.2312/SGP/SGP07/213-223

Qixing Huang, Bart Adams, and Michael Wand. 2007b. Bayesian Surface Reconstruction via Iterative Scan Alignment to an Optimized Prototype. In *Proceedings of the Fifth Eurographics Symposium on Geometry Processing (SGP '07)*. Eurographics Association, Aire-la-Ville, Switzerland, Switzerland, 213–223. http://dl.acm.org/citation.cfm?id=1281991.1282021

Qi-Xing Huang and Dragomir Anguelov. 2010. High quality pose estimation by aligning multiple scans to a latent map. In *IEEE International Conference on Robotics and Automation, ICRA 2010*. IEEE, Anchorage, Alaska, USA, 1353–1360.

Qi-Xing Huang and Leonidas Guibas. 2013. Consistent Shape Maps via Semidefinite Programming. In *Proceedings of the Eleventh Eurographics/ACMSIGGRAPH Symposium on Geometry Processing (SGP âĂŹ13)*. Eurographics Association, Goslar, DEU, 177âĂŞ186.

Xiangru Huang, Zhenxiao Liang, Chandrajit Bajaj, and Qixing Huang. 2017. Translation Synchronization via Truncated Least Squares. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, 4-9 December 2017*. Curran Associates, Long Beach, CA, USA, 1459–1468. http://papers.nips.cc/paper/6744-translation-synchronization-via-truncated-least-squares

Xiangru Huang, Zhenxiao Liang, Xiaowei Zhou, Yao Xie, Leonidas J. Guibas, and Qixing Huang. 2019. Learning Transformation Synchronization. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019*. IEEE, New York City, NY, USA, 8082–8091. https://doi.org/10.1109/CVPR.2019.00827

Daniel Huber. 2002. *Automatic Three-dimensional Modeling from Reality*. Ph.D. Dissertation. Carnegie Mellon University, Pittsburgh, PA.

Thorbjørn Mosekjær Iversen, Anders Glent Buch, and Dirk Kraft. 2017. Prediction of ICP pose uncertainties using Monte Carlo simulation with synthetic depth images. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2017, September 24-28, 2017*. IEEE/RSJ, Vancouver, BC, Canada, 4640–4647. https://doi.org/10.1109/IROS.2017.8206335

Shahram Izadi, David Kim, Otmar Hilliges, David Molyneaux, Richard Newcombe, Pushmeet Kohli, Jamie Shotton, Steve Hodges, Dustin Freeman, Andrew Davison, and Andrew Fitzgibbon. 2011. KinectFusion: Real-time 3D Reconstruction and Interaction Using a Moving Depth Camera. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology (UIST '11)*. ACM, New York, NY, USA, 559–568. https://doi.org/10.1145/2047196.2047270

Hubert Jin, Thomas Duchamp, Hugues Hoppe, John Alan McDonald, Kari Pulli, and Werner Stuetzle. 1995. Surface reconstruction from misregistered data. In *Vision Geometry IV*, Robert A. Melter, Angela Y. Wu, Fred L. Bookstein, and William D. K. Green (Eds.), Vol. 2573. International Society for Optics and Photonics, SPIE, San Diego, CA, United States, 324 – 328. https://doi.org/10.1117/12.216424

Olaf Kähler, Victor Adrian Prisacariu, Carl Yuheng Ren, Xin Sun, Philip H. S. Torr, and David William Murray. 2015. Very High Frame Rate Volumetric Integration of Depth Images on Mobile Devices. *IEEE Trans. Vis. Comput. Graph.* 21, 11 (2015), 1241–1250. https://doi.org/10.1109/TVCG.2015.2459891

K. Kanatani and D. D. Morris. 2006. Gauges and Gauge Transformations for Uncertainty Description of Geometric Structure with Indeterminacy. *IEEE Trans. Inf. Theor.* 47, 5 (Sept. 2006), 2017–2028. https://doi.org/10.1109/18.930934

Michael Kazhdan, Matthew Bolitho, and Hugues Hoppe. 2006. Poisson Surface Reconstruction. In *Proceedings of the Fourth Eurographics Symposium on Geometry Processing (SGP '06)*. Eurographics Association, Aire-la-Ville, Switzerland, Switzerland, 61–70. http://dl.acm.org/citation.cfm?id=1281957.1281965

Maik Keller, Damien Lefloch, Martin Lambers, Shahram Izadi, Tim Weyrich, and Andreas Kolb. 2013. Real-Time 3D Reconstruction in Dynamic Scenes Using Point-Based Fusion. In *Proceedings of the 2013 International Conference on 3D Vision (3DV âĂŹ13)*. IEEE Computer Society, USA, 1âĂŞ8. https://doi.org/10.1109/3DV.2013.9

Steven G. G. Krantz and Harold R. Parks. 2013. *The Implicit Function Theorem: History, Theory, and Applications*. Springer, New York City.

Shankar Krishnan, Pei Yean Lee, John B. Moore, and Suresh Venkatasubramanian. 2005. Global Registration of Multiple 3D Point Sets via Optimization-on-a-Manifold. In *Proceedings of the Third Eurographics Symposium on Geometry Processing (SGP âĂŹ05)*. Eurographics Association, Goslar, DEU, 187âĂŞes.

David Landry, François Pomerleau, and Philippe Giguère. 2019. CELLO-3D: Estimating the Covariance of ICP in the Real World. In *International Conference on Robotics and Automation, ICRA 2019*. IEEE, Montreal, QC, Canada, 8190–8196. https://doi.org/10.1109/ICRA.2019.8793516

Maxime Lhuillier and Mathieu Perriollat. 2006. Uncertainty Ellipsoids Calculations for Complex 3D Reconstructions. In *Proceedings of the 2006 IEEE International Conference on Robotics and Automation*. IEEE, Orlando, Florida, USA, 3062–3069. https://doi.org/10.1109/ROBOT.2006.1642167

Yangyan Li, Xiaochen Fan, Niloy J. Mitra, Daniel Chamovitz, Daniel Cohen-Or, and Baoquan Chen. 2013. Analyzing Growing Plants from 4D Point Cloud Data. *ACM Trans. Graph.* 32, 6, Article Article 157 (Nov. 2013), 10 pages. https://doi.org/10.1145/2508363.2508368

Yang Liu, Helmut Pottmann, and Wenping Wang. 2006. Constrained 3D shape reconstruction using a combination of surface fitting and registration. *Computer-Aided Design* 38, 6 (2006), 572–583. https://doi.org/10.1016/j.cad.2006.01.014

Prakhya Sai Manoj, Bingbing Liu, Yan Rui, and Weisi Lin. 2015. A closed-form estimate of 3D ICP covariance. In *14th IAPR International Conference on Machine Vision Applications, MVA 2015*. IAPR, Miraikan, Tokyo, Japan, 526–529. https://doi.org/10.1109/MVA.2015.7153246

Juan I. Nieto, Tim Bailey, and Eduardo Mario Nebot. 2005. Scan-SLAM: Combining EKF-SLAM and Scan Correlation. In *Field and Service Robotics, Results of the 5th International Conference, FSR 2005*. Springer, Port Douglas, QLD, Australia, 167–178. https://doi.org/10.1007/978-3-540-33453-8_15

Jorge Nocedal and Stephen J. Wright. 2006. *Numerical Optimization* (second ed.). Springer, New York, NY, USA.

Mark Pauly, Niloy J. Mitra, and Leonidas J. Guibas. 2004. Uncertainty and Variability in Point Cloud Surface Data. In *Proceedings of the First Eurographics Conference on Point-Based Graphics (SPBGâĂŹ04)*. Eurographics Association, Goslar, DEU, 77âĂŞ84.

Michal Polic, Wolfgang Förstner, and Tomás Pajdla. 2018. Fast and Accurate Camera Covariance Computation for Large 3D Reconstruction. In *Computer Vision - ECCV 2018 - 15th European Conference, September 8-14, 2018, Proceedings, Part II*. Springer Verlag, Munich, Germany, 697–712. https://doi.org/10.1007/978-3-030-01216-8_42

Kai Pöthkow, Britta Weber, and Hans-Christian Hege. 2011. Probabilistic Marching Cubes. *Comput. Graph. Forum* 30, 3 (2011), 931–940. https://doi.org/10.1111/j.1467-8659.2011.01942.x

Helmut Pottmann and Michael Hofer. 2002. Geometry of the Squared Distance Function to Curves and Surfaces. In *Third International Workshop "Visualization and Mathematics", VisMath 2002*. Springer Verlag, Berlin, Germany, 221–242. https://doi.org/10.1007/978-3-662-05105-4_12

Helmut Pottmann, Qi-Xing Huang, Yong-Liang Yang, and Shi-Min Hu. 2006. Geometry and Convergence Analysis of Algorithms for Registration of 3D Shapes. *International Journal of Computer Vision* 67, 3 (2006), 277–296. https://doi.org/10.1007/s11263-006-5167-2

Helmut Pottmann, Stefan Leopoldseder, and Michael Hofer. 2004. Registration without ICP. *Computer Vision and Image Understanding* 95, 1 (2004), 54–71. https://doi.org/10.1016/j.cviu.2004.04.002

Anil C. Raghuramu. 2015. Robust Multiview Registration of 3D Surfaces via L_1-norm Minimization. In *Proceedings of the British Machine Vision Conference (BMVC)*, Mark W. Jones Xianghua Xie and Gary K. L. Tam (Eds.). BMVA Press, Swansea, UK, Article 62, 11 pages. https://doi.org/10.5244/C.29.62

Szymon Rusinkiewicz, Olaf Hall-Holt, and Marc Levoy. 2002. Real-time 3D Model Acquisition. *ACM Trans. Graph.* 21, 3 (July 2002), 438–446. https://doi.org/10.1145/566654.566600

William R. Scott. 2009. Model-Based View Planning. *Mach. Vision Appl.* 20, 1 (Jan. 2009), 47âĂŞ69.

William R. Scott, Gerhard Roth, and Jean-François Rivest. 2003. View Planning for Automated Three-Dimensional Object Reconstruction and Inspection. *ACM Comput. Surv.* 35, 1 (March 2003), 64âĂŞ96. https://doi.org/10.1145/641865.641868

Pei Sun, Henrik Kretzschmar, Xerxes Dotiwalla, Aurelien Chouard, Vijaysai Patnaik, Paul Tsui, James Guo, Yin Zhou, Yuning Chai, Benjamin Caine, Vijay Vasudevan, Wei Han, Jiquan Ngiam, Hang Zhao, Aleksei Timofeev, Scott Ettinger, Maxim Krivokon, Amy Gao, Aditya Joshi, Yu Zhang, Jonathon Shlens, Zhifeng Chen, and Dragomir Anguelov. 2019. Scalability in Perception for Autonomous Driving: Waymo Open Dataset. arXiv:cs.CV/1912.04838

Richard Szeliski. 1990. Bayesian modeling of uncertainty in low-level vision. *Int. J. Comput. Vis.* 5, 3 (1990), 271–301. https://doi.org/10.1007/BF00126502

Richard Szeliski and Sing Bing Kang. 1997. Shape Ambiguities in Structure From Motion. *IEEE Trans. Pattern Anal. Mach. Intell.* 19, 5 (1997), 506–512. https://doi.org/10.1109/34.589211

Gary K. Tam, Zhi-Quan Cheng, Yu-Kun Lai, Frank Langbein, Yonghuai Liu, A. David Marshall, Ralph Martin, Xianfang Sun, and Paul Rosin. 2013. Registration of 3D Point Clouds and Meshes: A Survey from Rigid to Nonrigid. *IEEE Transactions on Visualization and Computer Graphics* 19, 7 (July 2013), 1199–1217. https://doi.org/10.1109/TVCG.2012.310

Andrea Torsello, Emanuele Rodolà, and Andrea Albarelli. 2011. Multiview registration via graph diffusion of dual quaternions. In *The 24th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2011*. IEEE, Colorado Springs, CO, USA, 2441–2448. https://doi.org/10.1109/CVPR.2011.5995565

Bill Triggs, Philip F. McLauchlan, Richard I. Hartley, and Andrew W. Fitzgibbon. 1999. Bundle Adjustment - A Modern Synthesis. In *Vision Algorithms: Theory and Practice, International Workshop on Vision Algorithms, held during ICCV '99*. IEEE, Corfu, Greece, 298–372. https://doi.org/10.1007/3-540-44480-7_21

Dragan Tubic, Patrick Hébert, and Denis Laurendeau. 2003. A volumetric approach for interactive 3D modeling. *Comput. Vis. Image Underst.* 92, 1 (2003), 56–77. https://doi.org/10.1016/j.cviu.2003.07.001

A.W. Van Der Vaart. 1998. *Asymptotic Statistics*. Cambridge University Press, One Liberty Plaza, New York, NY 10006. https://books.google.com/books?id=udhfQgAACAAJ

Grace Wahba. 1983. Bayesian C̈onfidence IntervalsF̈or the Cross-Validated Smoothing Spline. *Journal of the Royal Statistical Society. Series B* 45 (1983), 133–150. Issue 1.

Lanhui Wang and Amit Singer. 2013. Exact and Stable Recovery of Rotations for Robust Synchronization. *Information and Inference: A Journal of the IMA* 2 (December 2013), 145âĂŞ193. Issue 2.

Thomas Whelan, Michael Kaess, Hordur Johannsson, Maurice F. Fallon, John J. Leonard, and John McDonald. 2015. Real-time large-scale dense RGB-D SLAM with volumetric fusion. *I. J. Robotics Res.* 34, 4-5 (2015), 598–626. https://doi.org/10.1177/0278364914551008

Thomas Whelan, Renato F Salas-Moreno, Ben Glocker, Andrew J Davison, and Stefan Leutenegger. 2016. ElasticFusion. *Int. J. Rob. Res.* 35, 14 (Dec. 2016), 1697âĂŞ1716. https://doi.org/10.1177/0278364916669237

Hong-Tzong Yau, Tsan-Jui Yang, and Yi-Chen Chen. 2014. Tooth model reconstruction based upon data fusion for orthodontic treatment simulation. *Comp. in Bio. and Med.* 48 (2014), 8–16. http://dblp.uni-trier.de/db/journals/cbm/cbm48.html#YauYC14

Fuzhen Zhang. 2005. *The Schur Complement and its Applications*. Numerical Methods and Algorithms, Vol. 4. Springer, New York. https://doi.org/10.1007/b105056

Zhengyou Zhang. 1998. Determining the Epipolar Geometry and its Uncertainty: A Review. *Int. J. Comput. Vis.* 27, 2 (1998), 161–195. https://doi.org/10.1023/A:1007941100561

Zaiwei Zhang, Zhenxiao Liang, Lemeng Wu, Xiaowei Zhou, and Qixing Huang. 2019. Path-Invariant Map Networks. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019*. IEEE, New York City, NY, USA, 11084–11094. https://doi.org/10.1109/CVPR.2019.01134

Michael Zollhöfer, Patrick Stotko, Andreas Görlitz, Christian Theobalt, Matthias Nießner, Reinhard Klein, and Andreas Kolb. 2018. State of the Art on 3D Reconstruction with RGB-D Cameras. *Comput. Graph. Forum* 37, 2 (2018), 625–652. https://doi.org/10.1111/cgf.13386

A PROOFS OF PROPOSITIONS AND THEOREMS IN SECTION 4

In this section, we present a formal statement of the main analysis result of the paper. We begin with introducing the notations in Section A.1. We then discuss the main result, key lemmas, and a proof of the main result in Section A.2. In Section A.3, we show a bound for the error Taylor's approximation of the implicit function induced from some energy function that fits some regularity conditions. Finally in Section A.4, we discuss the error due to the reconstruction. Except in Section A.4 we always use the ground truth reconstruction parameter, i.e. $\boldsymbol{p}^{gt}$, as in (5).

A.1 Notations

Definition 2. *Given a vector $\boldsymbol{y} \in \mathbb{R}^n$, we define the k-order product tensor induced from $\boldsymbol{y}$ as $\otimes^k \boldsymbol{y} \in \mathbb{R}^{n\times\cdots\times n}$.*

Definition 3. *Given two k-order tensors $A, B \in \mathbb{R}^{n_1\times\cdots\times n_k}$, we define their inner products as*

$$\langle A, B\rangle = \sum_{i_1=1}^{n_1}\cdots\sum_{i_k=1}^{n_k} a_{i_1\cdots i_k}\cdot b_{i_1\cdots i_k}, \tag{29}$$

where $a_{i_1\cdots i_k}$ and $b_{b_{i_1\cdots i_k}}$ denote the corresponding elements of A and B.

Definition 4. *Suppose the implicit function $\boldsymbol{x}(\boldsymbol{y})$ is given by binary function f such that $f(\boldsymbol{x}(\boldsymbol{y}), \boldsymbol{y}) = 0$. For simplicity we assume $f(0,0) = 0$ and now we are interested about the local behavior near $(0,0)$. The k-th order approximation for $\boldsymbol{x}(\boldsymbol{y})$ is referred to as*

$$S_n(\boldsymbol{y}) := \sum_{k=0}^{n}\frac{1}{k!}\langle \boldsymbol{x}^{(k)}(0), \otimes^k \boldsymbol{y}\rangle$$

where $\boldsymbol{x}^{(k)}$ is the k-th order derivative of $\boldsymbol{x}$ with respect to $\boldsymbol{y}$. The k-th order remainder is naturally defined as

$$R_n(\boldsymbol{y}) := \boldsymbol{x}(\boldsymbol{y}) - S_n(\boldsymbol{y}) = \sum_{k=n+1}^{\infty}\frac{1}{k!}\langle \boldsymbol{x}^{(k)}(0), \otimes^k \boldsymbol{y}\rangle.$$

For convenience we use abbreviation $f_{x^b y^s}$ to denote the partial derivative $\frac{\partial^{b+s}}{\partial x^b \partial y^s} f(0,0)$ which is a $(b+s)$-order tensor. Sometimes we treat $f_{x^b y^s}$ as a matrix of shape

$$\text{size}(\boldsymbol{x})^b \times \text{size}(\boldsymbol{y})^s$$

where size$(\cdot)$ *takes the number of components of operand.*

Lemma 2.

$$\sum_{k=0}^{n}\frac{1}{k!}\big(\otimes^k R_n(\boldsymbol{y})\big)\cdot f_{x^k}(S_n(\boldsymbol{y}), \boldsymbol{y}) = 0$$

$$\sum_{p\geq 0}\frac{1}{p!}\big(\otimes^p R_n(\boldsymbol{y})\big)\cdot\sum_{q,s\geq 0}\frac{1}{q!s!}\big(\otimes^q S_n(\boldsymbol{y})\big)\cdot f_{x^{p+q}y^s}\cdot\big(\otimes^s \boldsymbol{y}\big) = 0$$

Proof. By the Taylor's theorem for binary function $f(x,y)$ we have that

$$\sum_{b,s\geq 0}\frac{1}{b!s!} f_{x^b y^s}\big(\otimes^b \boldsymbol{x}(\boldsymbol{y})\big)\otimes\big(\otimes^s \boldsymbol{y}\big) = 0 \tag{30}$$

Substituting $\boldsymbol{x}(\boldsymbol{y})$ by $R_n(\boldsymbol{y}) + S_n(\boldsymbol{y})$, (30) turns into

$$\begin{aligned}
0 &= \sum_{b,s\geq 0}\frac{1}{b!s!} f_{x^b y^s}\big(\otimes^b \boldsymbol{x}(\boldsymbol{y})\big)\otimes\big(\otimes^s \boldsymbol{y}\big)\\
&= \sum_{p,q,s\geq 0}\frac{1}{p!q!s!}\big(\otimes^p R_n(\boldsymbol{y})\big)\otimes\big(\otimes^q S_n(\boldsymbol{y})\big)\cdot f_{x^{p+q}y^s}\cdot\big(\otimes^s \boldsymbol{y}\big) \qquad (31)\\
&= \sum_{p\geq 0}\frac{1}{p!}\big(\otimes^p R_n(\boldsymbol{y})\big)\cdot\sum_{q,s\geq 0}\frac{1}{q!s!}\big(\otimes^q S_n(\boldsymbol{y})\big)\cdot f_{x^{p+q}y^s}\cdot\big(\otimes^s \boldsymbol{y}\big)\\
&= \sum_{p\geq 0}\frac{1}{p!}\big(\otimes^p R_n(\boldsymbol{y})\big)\cdot f_{x^p}(S_n(\boldsymbol{y}), \boldsymbol{y}) \qquad (32)
\end{aligned}$$

The equation (32) was obtained by shrinking the Taylor expansion of $f_{x^p}(\boldsymbol{x}, \boldsymbol{y})$ at point $(S_n(\boldsymbol{y}), \boldsymbol{y})$. In general the binomial theorem cannot apply to tensor product because of the lack of commutativity. As a matter of fact in (31) we utilized the symmetric property of the tensor $f_{x^{p+q}y^s}$.

□

Lemma 2 allows us to estimate $|R_n(\boldsymbol{y})|$ only depending on partial derivatives of binary function f and low-order Taylor's expansion of $\boldsymbol{x}(\boldsymbol{y})$.

Corollary 2 (Scalar Version). *Suppose*

$$\left|\sum_{k\geq 2}\frac{1}{k!}\big(\otimes^k R_n(\boldsymbol{y})\big)\cdot f_{x^k}(S_n(\boldsymbol{y}), \boldsymbol{y})\right| \leq c R_n(\boldsymbol{y})\cdot f_x(S_n(\boldsymbol{y}), \boldsymbol{y}) \tag{33}$$

holds for some constant $c < 1$, Then

$$|R_n(\boldsymbol{y})| \leq \frac{1}{1-c}|f_x(S_n(\boldsymbol{y}), \boldsymbol{y})|^{-1}|f(S_n(\boldsymbol{y}), \boldsymbol{y})|.$$

remark. It is worth noting that the condition (33) is not that strong as it seems at first look. In the neighborhood of $\boldsymbol{y}_0 = 0$, $|R_n(\boldsymbol{y})|$ has the order of magnitude $O(|\boldsymbol{y}|^{n+1})$, which means $|f_{x^k}|$ just needs to grow in the rate slower than $O\left(\frac{k!}{|\boldsymbol{y}|^{(n+1)k}}\right)$. It can be shown that

$$f_x(S_n(\boldsymbol{y}), \boldsymbol{y}) = -\frac{1}{(n+1)!}\langle \boldsymbol{x}^{(n+1)}, \otimes^{n+1}\boldsymbol{y}\rangle + o(|\boldsymbol{y}|^{n+1}),$$

hence the dominating term in $|R_n(y)|$ would be

$$\frac{1}{(n+1)!}|\boldsymbol{x}^{(n+1)}|_F\cdot|\boldsymbol{y}|^{n+1},$$

compliant with the Taylor's expansion of $\boldsymbol{x}(\boldsymbol{y})$.

A.2 Main Result and Key Lemmas

In this subsection we consider implicit function $\boldsymbol{x}(\boldsymbol{y})$ induced by energy function $T(\boldsymbol{x}, \boldsymbol{y})$:

$$\boldsymbol{x}(\boldsymbol{y}) = \arg\min_{\boldsymbol{x}} T(\boldsymbol{x}, \boldsymbol{y}).$$

Assume $\boldsymbol{x}(0) = 0$ as before. As we focus on the local behavior of $\boldsymbol{x}$ near $(0,0)$ with some appropriate smooth conditions, this definition is equivalent to

$$T_x(\boldsymbol{x}, \boldsymbol{y}) = 0$$

where $T_x = \frac{\partial T}{\partial x}$.

A subtle difference between T_x and f introduced in A.1 is that T_x is a vector instead of scalar. But Lemma 2 still holds true by adding one more dimension from vector-valued T_x to tensors f_{x^k} and $f_{x^p+q_{y^s}}$ through proof. Similar to Corollary 2, we have the following lemma when $T(\boldsymbol{x}, \boldsymbol{y})$ is known:

Lemma 3 (Vector Version). *Denote by $\lambda_{\max}$ and $\lambda_{\min}$ the largest and smallest eigenvalues of matrix $T_{xx}(S_n(\boldsymbol{y}), \boldsymbol{y})$ respectively, and suppose*

$$\left|\sum_{k\geq 2}\frac{1}{k!}\left(\otimes^k R_n(\boldsymbol{y})\right)\cdot T_{x^{k+1}}(S_n(\boldsymbol{y}), \boldsymbol{y})\right| \leq c|R_n(\boldsymbol{y})|^2\cdot\lambda_{\max} \tag{34}$$

holds for some constant $c < 1$. Then

$$\begin{aligned}|R_n(\boldsymbol{y})| &\leq \frac{2|T_x(S_n, \boldsymbol{y})|}{\lambda_{\min}+\sqrt{\lambda_{\min}^2 - 4c\lambda_{\max}|T_x(S_n, \boldsymbol{y})|}}\\ &\leq \frac{2|T_x(S_n, \boldsymbol{y})|}{\lambda_{\min}}\end{aligned}$$

whenever $|T_x(S_n, \boldsymbol{y})| \leq \lambda_{\min}^2/(4c\lambda_{\max})$. Here $\|\cdot\|$ represents the spectral norm of matrices.

Proof. By Lemma 2 and the triangle inequality we have

$$\begin{aligned}&|T_{xx}(S_n(\boldsymbol{y}), \boldsymbol{y})\cdot R_n(\boldsymbol{y})|\\ \leq&|T_x(S_n(\boldsymbol{y}), \boldsymbol{y})| + \left|\sum_{k=2}^{n}\frac{1}{k!}\left(\otimes^k R_n(\boldsymbol{y})\right)\cdot T_{x^{k+1}}(S_n(\boldsymbol{y}), \boldsymbol{y})\right|\\ \leq&|T_x(S_n(\boldsymbol{y}), \boldsymbol{y})| + c\lambda_{\max}|R_n(\boldsymbol{y})|^2\end{aligned}$$

By definition of eigenvalues,

$$|T_{xx}(S_n(\boldsymbol{y}), \boldsymbol{y})\cdot R_n(\boldsymbol{y})| \geq \lambda_{\min}|R_n(\boldsymbol{y})|.$$

Thus

$$\lambda_{\min}|R_n(\boldsymbol{y})| \leq |T_x(S_n(\boldsymbol{y}), \boldsymbol{y})| + c\cdot\lambda_{\max}|R_n(\boldsymbol{y})|^2$$

which is a quadratic inequality about $|R_n(\boldsymbol{y})|$ thus has two branches. Since $R_n(y) = O(y^{n+1})$ in the neighborhood of $y_0 = 0$, by continuity we could assume $|R_n(\boldsymbol{y})|$ falls into the smaller branch. Solving this inequality reveals that

$$|R(\boldsymbol{y})| \leq \frac{2|T_x(S_n, \boldsymbol{y})|}{\lambda_{\min}+\sqrt{\lambda_{\min}^2 - 4c\lambda_{\max}|T_x(S_n, \boldsymbol{y})|}},$$

which completes our proof. □

We next bound the difference between $\overline{C}$ and $\mathrm{Var}[\boldsymbol{x}(\boldsymbol{y})]$ using tools introduced above. Recall that

$$\overline{C} := \mathrm{Var}(\boldsymbol{x}'\cdot\boldsymbol{y}) = \mathrm{Var}[S_1(\boldsymbol{y})]$$

where S_n was introduced in A.1. In the following we would discuss the more general situation where

$$\overline{C}_n := \mathrm{Var}[S_n(\boldsymbol{y})]$$

in which we use $S_n(\boldsymbol{y})$ to approximate $\boldsymbol{x}(\boldsymbol{y})$ rather than the linear approximation $S_1(\boldsymbol{y})$. In this way

$$\begin{aligned}&\mathrm{Var}[\boldsymbol{x}(\boldsymbol{y})] - \overline{C}_n\\ =&\mathrm{Var}[\boldsymbol{x}(\boldsymbol{y})] - \mathrm{Var}[S_n(\boldsymbol{y})]\\ =&\mathrm{Var}[R_n(\boldsymbol{y})] + \mathrm{Cov}[R_n(\boldsymbol{y}), S_n(\boldsymbol{y})] + \mathrm{Cov}[S_n(\boldsymbol{y}), R_n(\boldsymbol{y})]\end{aligned}$$

where $\mathrm{Cov}(X, Y) := \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])^T]$ is the covariance matrix. By Cauchy's inequality we have

$$\begin{aligned}\|\mathrm{Var}[\boldsymbol{x}(\boldsymbol{y})] - \overline{C}_n\| &\leq \mathbb{E}\left[|R_n(\boldsymbol{y})|^2\right] + 2\sqrt{\mathbb{E}\left[|R_n(\boldsymbol{y})|^2\right]\mathbb{E}\left[|S_n(\boldsymbol{y})|^2\right]}\\ &= \sqrt{\mathbb{E}\left[|R_n(\boldsymbol{y})|^2\right]}\left(\sqrt{\mathbb{E}\left[|R_n(\boldsymbol{y})|^2\right]} + 2\sqrt{\mathbb{E}\left[|S_n(\boldsymbol{y})|^2\right]}\right).\end{aligned} \tag{35}$$

Lemma 3 together with inequality (35) provides a general framework to bound the error of our approximated variance. In general $S_n(\boldsymbol{y})$ can be calculated explicitly and $T_x(S_n(\boldsymbol{y}), \boldsymbol{y})$ has the order of magnitude $O(|\boldsymbol{y}|^{n+1})$ in the neighborhood of $\boldsymbol{y}_0 = 0$. Although $\lambda_{\min}$ and $\lambda_{\max}$ are functions with respect to $\boldsymbol{y}$, we could expect they are stable within the area of our interest since the implicit function $\boldsymbol{x}(\boldsymbol{y})$ would be ill-conditioned if $\lambda_{\min}(T_{xx})$ tends to zero.

Lemma 4.

$$T_x(S_n(\boldsymbol{y}), \boldsymbol{y}) = -\frac{1}{(n+1)!}\langle\otimes^{n+1}\boldsymbol{y}, \boldsymbol{x}^{(n+1)}(0)\rangle + o\left(|\boldsymbol{y}|^{n+2}\right)$$

Proof. By definition of $S_n(\boldsymbol{y})$ we have

$$T_x(S_n(\boldsymbol{y}) + R_n(\boldsymbol{y}), \boldsymbol{y}) = 0 \tag{36}$$

The Taylor's expansion of T_x has the form like Lemma 2.

It is worth noting that if $R_n(\boldsymbol{y})$ appeared in some product then this product would have no terms of degree less or equal to n. Hence the all terms of degree at most n on left side of Lemma 2 is contributed by $T_x(S_n(\boldsymbol{y}), \boldsymbol{y})$, which is of course zero from the right side, i.e., constant zero.

As for the coefficient of $\otimes^{n+1}\boldsymbol{y}$, it suffices to compare the coefficient corresponding to degree $n+1$ on both side of Lemma 2. □

As far as the linear approximation $S_1(\boldsymbol{y})$ is concerned, (35) can be simplified further.

Theorem 5. *Suppose the random variable $\boldsymbol{y}$ satisfies $\mathbb{E}[\boldsymbol{y}] = 0$. The function $T(\boldsymbol{x}, \boldsymbol{y})$ satisfies the condition in (34) and*

$$|T_x(S_1(\boldsymbol{y}), \boldsymbol{y})| \leq \|T_{xx}\|/\left(4c\kappa(\boldsymbol{y})\right)$$

where

$$\kappa(\boldsymbol{y}) := \lambda_{\max}(T_{xx}(S_1, \boldsymbol{y}))/\lambda_{\min}(T_{xx}(S_1, \boldsymbol{y})).$$

Then we have

$$\begin{aligned}&|\mathrm{Var}[\boldsymbol{x}(\boldsymbol{y})] - \overline{C}_1|\\ \leq&4\mathbb{E}\left[\frac{|T_x(S_1, \boldsymbol{y})|^2\kappa(\boldsymbol{y})^2}{\|T_{xx}(S_1, \boldsymbol{y})\|^2}\right] + 2|\boldsymbol{x}'(0)|\mathbb{E}\left[\frac{2|T_x(S_1, \boldsymbol{y})|\cdot\kappa(\boldsymbol{y})\cdot|\boldsymbol{y}|}{\|T_{xx}(S_1, \boldsymbol{y})\|}\right]\end{aligned}$$

Proof. Just note that

$$\mathrm{Cov}[R_1(\boldsymbol{y}), S_1(\boldsymbol{y})] = \mathbb{E}[R_1(\boldsymbol{y})S_1(\boldsymbol{y})^T]$$

and

$$\|\mathrm{Var}(R_1(\boldsymbol{y}))\| \leq \mathbb{E}[|R_1(\boldsymbol{y})|^2].$$

Then directly using Lemma 3 completes the proof. □

A.3 An Explicit Bound for Functions with Regularity

Finally we show that a family of functions with regularity conditions satisfy the condition (34).

Theorem 6. *Suppose there are positive constants $M, \alpha, \beta, \zeta, \kappa$ such that:*

- $$\|T_{\boldsymbol{x}^{b+1}\boldsymbol{y}^s}(0,0)\| \le M\alpha^b\beta^s$$
  Here $T_{\boldsymbol{x}^{b+1}\boldsymbol{y}^s}$ is regarded as a matrix whose column size is the dimension of variable $\boldsymbol{x}$ and $\|\cdot\|$ takes the spectral norm of matrices.
- $$\lambda_{\min}(T_{xx}(0,0)) > \frac{M\alpha}{\kappa},$$
  which implies that
  $$\frac{\lambda_{\max}(T_{xx}(0,0))}{\lambda_{\min}(T_{xx}(0,0))} \le \kappa.$$
- $$|R_n(\boldsymbol{y})| \le \frac{\zeta}{\alpha} \tag{37}$$
  and
  $$|T_x(S_n(\boldsymbol{y}),\boldsymbol{y})| \le \frac{(1/\kappa - \tau e^\tau)^2}{2M\alpha^2 e^{\zeta+\tau}}, \tag{38}$$
  where $\tau(\boldsymbol{y}) = \alpha|S_n(\boldsymbol{y})| + \beta|\boldsymbol{y}|$.

Then we have

$$|R_n(\boldsymbol{y})| \le \frac{2|T_x(S_n(\boldsymbol{y}),\boldsymbol{y})|}{\|T_{xx}(0,0)\|\big(1/\kappa - \tau(\boldsymbol{y})e^{\tau(\boldsymbol{y})}\big)} \tag{39}$$

remark. We make some explanation for the last two conditions above. (37) means $|R_n(\boldsymbol{y})|$ is constrained inside the area what we consider into the neighborhood of $(\boldsymbol{x}_0, \boldsymbol{y}_0) = (0,0)$. This condition is necessary because the higher-order terms would play a critical role and the polynomial approximation fails when $\boldsymbol{y}$ go beyond the area of constant radius.

The (38) is in fact rather weak. We have shown that

$$|T_x(S_n(\boldsymbol{y}),\boldsymbol{y})| = -\frac{1}{(n+1)!}|\boldsymbol{x}^{(n+1)}(0)|_F \cdot |\boldsymbol{y}|^{n+1} + o\big(|\boldsymbol{y}|^{n+1}\big)$$

and the right side of (38) has a lower bound when constrained in the neighborhood of $(0,0)$ where the growth of τ would be nearly linear.

Proof. By Taylor's expansion, for $b \ge 1$ we have

$$\begin{aligned}
\|T_{\boldsymbol{x}^{b+1}}(S_n(\boldsymbol{y}),\boldsymbol{y})\| &\le \sum_{p,q\ge 0}\frac{1}{p!q!}\|T_{\boldsymbol{x}^{b+p+1}\boldsymbol{y}^q}(0,0)\| \cdot |R_n(\boldsymbol{y})|^p \cdot |\boldsymbol{y}|^q \\
&\le M\alpha^b \sum_{p,q\ge 0}\frac{1}{p!q!}\alpha^p\beta^q|R_n(\boldsymbol{y})|^p \cdot |\boldsymbol{y}|^q \\
&= M\alpha^b \sum_{k\ge 0}\frac{1}{k!}\big(\alpha|R_n(\boldsymbol{y})| + \beta|\boldsymbol{y}|\big)^k \\
&= M\alpha^b e^{\tau(\boldsymbol{y})}.
\end{aligned} \tag{40}$$

For T_{xx} we have

$$\begin{aligned}
&\|T_{xx}(S_n(\boldsymbol{y},\boldsymbol{y}) - T_{xx}(0,0)\| \\
\le &\sum_{p+q\ge 1}\frac{1}{p!q!}\|T_{x^{p+2}y^q}(0,0)\| \cdot |R_n(\boldsymbol{y})|^p \cdot |\boldsymbol{y}|^q \\
\le &M\alpha\sum_{k\ge 1}\sum_{p+q=k}\frac{1}{p!q!}\alpha^p\beta^q|R_n(\boldsymbol{y})|^p \cdot |\boldsymbol{y}|^q \\
= &M\alpha\sum_{k\ge 1}\frac{1}{k!}\tau(\boldsymbol{y})^k \\
\le &M\alpha\tau(\boldsymbol{y})e^{\tau(\boldsymbol{y})} < \kappa\tau(\boldsymbol{y})e^{\tau(\boldsymbol{y})} \cdot \lambda_{\min}(T_x x(0,0)).
\end{aligned}$$

Thus

$$\begin{aligned}
\lambda_{\min}(T_{xx}(S_n(\boldsymbol{y},\boldsymbol{y})) &\ge \lambda_{\min}(T_{xx}(0,0))(1 - \kappa\tau(\boldsymbol{y})e^{\tau(\boldsymbol{y})}) \\
&\ge \|T_{xx}(0,0)\|(1/\kappa - \tau(\boldsymbol{y})e^{\tau(\boldsymbol{y})})
\end{aligned} \tag{41}$$

Dividing left side of the equation

$$\sum_{b\ge 0}\frac{1}{b!}T_{\boldsymbol{x}^{b+1}}(S_n(\boldsymbol{y}),\boldsymbol{y}) \cdot \otimes^b R_n(\boldsymbol{y}) = 0$$

into three parts we obtain that

$$\begin{aligned}
&T_{xx}(S_n(\boldsymbol{y}),\boldsymbol{y}) \cdot R_n(\boldsymbol{y}) \\
= &-T_x(S_n(\boldsymbol{y}),\boldsymbol{y}) - \sum_{b\ge 2}\frac{1}{b!}T_{\boldsymbol{x}^{b+1}}(S_n(\boldsymbol{y}),\boldsymbol{y}) \cdot \otimes^b R_n(\boldsymbol{y}).
\end{aligned} \tag{42}$$

Inequality (41) implies that

$$\|T_{xx}(S_n(\boldsymbol{y}),\boldsymbol{y}) \cdot R_n(\boldsymbol{y})\| \ge \|T_{xx}(0,0)\|(1/\kappa - \tau(\boldsymbol{y})e^{\tau(\boldsymbol{y})})|R_n(\boldsymbol{y})|. \tag{43}$$

Inequality (40) implies that

$$\begin{aligned}
&\left|\sum_{b\ge 2}\frac{1}{b!}T_{\boldsymbol{x}^{b+1}}(S_n(\boldsymbol{y}),\boldsymbol{y}) \cdot \otimes^b R_n(\boldsymbol{y})\right| \\
\le &M\sum_{b\ge 2}\frac{1}{b!}\alpha^b e^{\tau(\boldsymbol{y})}|R_n(\boldsymbol{y})|^b \\
\le &\frac{1}{2}M\alpha^2\exp\big(\tau(\boldsymbol{y}) + \alpha|R_n(\boldsymbol{y})|\big).
\end{aligned} \tag{44}$$

Combining (42), (43) and (44) we obtain the inequality

$$\begin{aligned}
&\|T_{xx}(0,0)\|(1/\kappa - \tau(\boldsymbol{y})e^{\tau(\boldsymbol{y})})|R_n(\boldsymbol{y})| \\
\le &|T_x(S_n(\boldsymbol{y},\boldsymbol{y}))| + \frac{1}{2}M\alpha^2\exp\big(\tau(\boldsymbol{y}) + \alpha|R_n(\boldsymbol{y})|\big)|R_n(\boldsymbol{y})|^2 \\
\le &|T_x(S_n(\boldsymbol{y},\boldsymbol{y}))| + \frac{1}{2}M\alpha^2\exp\big(\tau(\boldsymbol{y}) + \zeta\big)|R_n(\boldsymbol{y})|^2
\end{aligned} \tag{45}$$

The inequality (45) is a quadratic inequality about variable $|R_n(\boldsymbol{y})|$ and the solution has two branches. Since now we are interested in the neighborhood of $\boldsymbol{y}_0 = 0$ and $R_n(\boldsymbol{y}) = O(|\boldsymbol{y}|^{n+1})$, we assume the solution falls into the branch corresponding to the smaller solutions. Solving this quadratic inequality and utilizing (38), the conclusion (39) immediately follows. □

A.4 The Error from Reconstruction Bias

In the last section we mainly consider the difference between $\mathrm{Var}[\boldsymbol{x}(\boldsymbol{y})]$ and $\overline{C}$ where the reconstruction parameter $\boldsymbol{p}$ in (4) is assumed to be the ground truth $\boldsymbol{p}^{gt}$, but such an assumption is unpractical in general. Instead we use $\boldsymbol{p}^\star$ that is derived from one set of scans. In other words, $\boldsymbol{p}^\star$ is a function of $\boldsymbol{y}$. In particular $\boldsymbol{p}^\star(\boldsymbol{0}) = \boldsymbol{p}^{gt}$. In this section we are about to estimate the error of (6) compared with (5) where the ground truth reconstruction $\boldsymbol{p}^{gt}$ is replaced by $\boldsymbol{p}^\star$.

The first order apprximation of $\boldsymbol{p}^\star$ is

$$\boldsymbol{p}^\star(\boldsymbol{y}) \approx \boldsymbol{p}^{gt} + \boldsymbol{p}'(0)\cdot\boldsymbol{y}. \tag{46}$$

In our noise model we assumed $\boldsymbol{y}$ has zero mean and covariance matrix $\sigma^2 I$, which means the components of random variable $\boldsymbol{y}$ are pariwisely independent. In the following argument we would use the first order approximation to analyze the error caused by the reconstruction bias. This reduction is reasonable since the reconstruction function $\boldsymbol{p}$ can be regarded locally linear about the input noise $\boldsymbol{y}$.

Thus by Markov's inequality,

$$\begin{aligned}\Pr[|\boldsymbol{p}^\star(\boldsymbol{y}) - \boldsymbol{p}^{gt}| \geq \gamma\sigma] &\leq \frac{\mathrm{diag}\big(\boldsymbol{p}'(0)^T\boldsymbol{p}'(0)\big)}{\gamma^2}\\ &= \frac{|\boldsymbol{p}'(0)|_F^2}{\gamma^2}\end{aligned}$$

where diag means the sum of diagonal entries.

For some fixed $\boldsymbol{y}$, we consider solution $\boldsymbol{x}$ and $\boldsymbol{x}'$ for $\boldsymbol{p}^{gt}$ and $\boldsymbol{p}^\star$ respectively:

$$f(\boldsymbol{x},\boldsymbol{y},\boldsymbol{p}^{gt}) = 0 \tag{47}$$

$$f(\boldsymbol{x}',\boldsymbol{y},\boldsymbol{p}^\star) = 0 \tag{48}$$

The Tyalor's expansion for f with respect to variables $\boldsymbol{x}$ and $\boldsymbol{p}$ would be

$$\sum_{b,s\geq 0}\frac{1}{b!s!}\big(\otimes^b(\boldsymbol{x}'-\boldsymbol{x})\big)\cdot f_{x^b p^s}(\boldsymbol{x},\boldsymbol{y},\boldsymbol{p}^{gt})\cdot\big(\otimes^s(\boldsymbol{p}^\star-\boldsymbol{p}^{gt})\big) = 0. \tag{49}$$

The first order approximation of $\boldsymbol{x}'$ would be

$$\boldsymbol{x}' \approx \boldsymbol{x} - T_{xx}^{-1}T_{xp}(\boldsymbol{p}^\star - \boldsymbol{p}^{gt}).$$

Thus

$$\mathrm{Var}[\boldsymbol{x}'] \approx \mathrm{Var}[\boldsymbol{x}] + T_{xx}^{-1}T_{xp}(\boldsymbol{p}^\star-\boldsymbol{p}^{gt})(\boldsymbol{p}^\star-\boldsymbol{p}^{gt})^T T_{px}T_{xx}^{-1}. \tag{50}$$

Formula (50) is rather crude since we have not taken y into consideration. The following theorem gives a more accurate and formal estimation for the discussion above,

Theorem 7. *Function $\boldsymbol{x}(\boldsymbol{y},\boldsymbol{p}^\star)$ is determined by equation*

$$\boldsymbol{x}(\boldsymbol{y},\boldsymbol{p}^\star) = \arg\min_{\boldsymbol{x}} T(\boldsymbol{x},\boldsymbol{y},\boldsymbol{p}^\star) \tag{51}$$

where $\boldsymbol{p}^\star$ is the predetermined parameter. Suppose $\boldsymbol{x}(\boldsymbol{y}^{gt},\boldsymbol{p}^{gt}) = \boldsymbol{x}^{gt}$ and $\boldsymbol{y}$ follows the noise model that $\mathrm{Var}[\boldsymbol{y}] = \sigma I^2$ as the main text.

For brevity we assume that

$$\boldsymbol{x}^{gt} = \boldsymbol{y}^{gt} = \boldsymbol{p}^{gt} = \boldsymbol{0}.$$

Let $S_1(\boldsymbol{y};\boldsymbol{p}^\star)$ be the first order approximation of function $\boldsymbol{x}(\boldsymbol{y},\boldsymbol{p}^\star)$ with respect to variable $\boldsymbol{y}$, i.e.,

$$S_1(\boldsymbol{y};\boldsymbol{p}^\star) = \frac{\partial\boldsymbol{x}}{\partial y}(\boldsymbol{0},\boldsymbol{p}^\star)\cdot\boldsymbol{y}.$$

Here we ignored the constant term $\boldsymbol{x}(\boldsymbol{0},\boldsymbol{p}^\star)$ because $\boldsymbol{x}(\boldsymbol{0},\boldsymbol{p}^{gt}) = \boldsymbol{0}$.

The following inequality provides a bound for bias between variances of $S_1(\boldsymbol{y},\boldsymbol{p}^\star)$ and $S_1(\boldsymbol{y},\boldsymbol{p}^{gt})$.

$$-Z(\boldsymbol{p}^\star) \preceq \mathrm{Var}[S_1(\boldsymbol{y};\boldsymbol{p}^\star)] - \mathrm{Var}[S_1(\boldsymbol{y};\boldsymbol{p}^{gt})] \preceq Z(\boldsymbol{p}^\star) \tag{52}$$

in which $Z(\boldsymbol{p})$ is a matrix such that

$$\begin{aligned}&\|Z(\boldsymbol{p}^\star)\|\\ \leq&2\sigma^2\max\Big\{\|T_{xx}^{-1}\|^2\cdot\|T_{xy}\|\big(\|T_{xx}^{-1}\|\cdot\|T_{xy}\|\cdot\|T_{xxp}\| + \|T_{xyp}\|\big)\Big\}\\ &\cdot|\boldsymbol{p}^\star-\boldsymbol{p}^{gt}|\cdot\dim(\boldsymbol{y})\end{aligned} \tag{53}$$

where max *takes arguments along the segment joining $(0,0,\boldsymbol{p}^\star)$ and $(0,0,\boldsymbol{p}^{gt})$ for partial derivatives T_{xx}, T_{xy}, T_{xxp} and T_{xyp}.*

remark. *If the partial derivatives of T is stable in the neighborhood of $\boldsymbol{p}^{gt}$ and $|\boldsymbol{p}^\star-\boldsymbol{p}^{gt}| < \delta$, we can simplify (53) into*

$$\|Z(\boldsymbol{p}^\star)\| \leq \sigma^2\cdot|\boldsymbol{p}^\star-\boldsymbol{p}^{gt}|\cdot C(\boldsymbol{p}^{gt},\delta) \tag{54}$$

where $C(\boldsymbol{p}^{gt},\delta)$ depends on the ground truth reconstruction $\boldsymbol{p}^{gt}$ and the reconstruction error $\boldsymbol{\delta}$. When $\boldsymbol{y}$ is normally distributed, $\boldsymbol{p}^\star$ can be also regarded as approximately normally distributed conentrated in $\boldsymbol{p}^{gt}$ in local sense. In this way the error matrix $\boldsymbol{Z}$ is typically smaller than $\mathrm{Var}[S_1(\boldsymbol{y};\boldsymbol{p}^{gt})]$ by several orders of magnitude.

Proof. Using the mean value theorem of vector-valued functions we have

$$\begin{aligned}&|S_1(\boldsymbol{y};\boldsymbol{p}^\star) - S_1(\boldsymbol{y};\boldsymbol{p}^{gt})|\\ \leq&\left\|\frac{\partial S_1}{\partial\boldsymbol{p}}(\boldsymbol{y},\boldsymbol{p}_1)\right\|\cdot|\boldsymbol{p}^\star-\boldsymbol{p}^{gt}|\\ \leq&\Big(\|T_{xx}^{-1}\|^2\cdot\|T_{xy}\|\cdot\|T_{xxp}\| + \|T_{xx}^{-1}\|\cdot\|T_{xyp}\|\Big)\cdot|\boldsymbol{y}|\cdot|\boldsymbol{p}^\star-\boldsymbol{p}^{gt}|.\end{aligned} \tag{55}$$

where $\boldsymbol{p}_1$ is some point on segment joining $\boldsymbol{p}^\star$ and $\boldsymbol{p}^{gt}$, and T_{xx}, T_{xy}, T_{xxp}, T_{xyp} are all taken at point $(\boldsymbol{x}^{gt},\boldsymbol{y}^{gt},\boldsymbol{p}_1)$.

Next we consider the variance. First we have ineuqality

$$\mathrm{Var}[S_1(\boldsymbol{y},\boldsymbol{p}^\star)] \preceq \mathbb{E}[S_1(\boldsymbol{y},\boldsymbol{p}^\star)S_1^T(\boldsymbol{y},\boldsymbol{p}^\star)].$$

Utilizing the face that $\mathbb{E}[S_1(\boldsymbol{y},\boldsymbol{p}^{gt})] = 0$, we have

$$\begin{aligned}&\mathrm{Var}[S_1(\boldsymbol{y};\boldsymbol{p}^\star)] - \mathrm{Var}[S_1(\boldsymbol{y};\boldsymbol{p}^{gt})]\\ \preceq&\mathbb{E}[S_1(\boldsymbol{y};\boldsymbol{p}^\star)S_1^T(\boldsymbol{y};\boldsymbol{p}^\star) - S_1(\boldsymbol{y};\boldsymbol{p}^{gt})S_1^T(\boldsymbol{y};\boldsymbol{p}^{gt})].\end{aligned}$$

Similarly, again applying mean value theorem we have

$$\begin{aligned}&|S_1(\boldsymbol{y};\boldsymbol{p}^\star)S_1^T(\boldsymbol{y};\boldsymbol{p}^\star) - S_1(\boldsymbol{y};\boldsymbol{p}^{gt})S_1^T(\boldsymbol{y};\boldsymbol{p}^{gt})|\\ \leq&2\left\|\frac{\partial S_1}{\partial p}(\boldsymbol{y};\boldsymbol{p}_1)S_1^T(\boldsymbol{y};\boldsymbol{p}_1)\right\|\cdot|\boldsymbol{p}^\star-\boldsymbol{p}^{gt}|\\ \leq&2\|T_{xx}^{-1}\|^2\cdot\|T_{xy}\|\Big(\|T_{xx}^{-1}\|\cdot\|T_{xy}\|\cdot\|T_{xxp}\| + \|T_{xyp}\|\Big)\\ &\cdot|\boldsymbol{y}|^2\cdot|\boldsymbol{p}^\star-\boldsymbol{p}^{gt}|\end{aligned}$$

where $\boldsymbol{p}_1$ is some point on segment joining $\boldsymbol{p}^\star$ and $\boldsymbol{p}^{gt}$, and T_{xx}, T_{xy}, T_{xxp}, T_{xyp} are all taken at point $(\boldsymbol{x}^{gt},\boldsymbol{y}^{gt},\boldsymbol{p}_1)$.

Since $\mathbb{E}[|\boldsymbol{y}|^2] = \sigma^2\dim(\boldsymbol{y})$ we completes our proof. □

B PROOFS OF PROPOSITIONS AND THEOREMS IN SECTION 5

B.1 Proof of Prop. 2

Since we only care about the local behavior near the ground truth, we use the $\boldsymbol{p}_{ij}^{gt}, \boldsymbol{n}_{ij}^{gt}$ to replace the optimized poses $\boldsymbol{p}_{ij}^{\star}, \boldsymbol{n}_{ij}^{\star}$. Also for convenience we assume the output $(\boldsymbol{c}_i, \boldsymbol{t}_i)$ is $(I, 0)$ corresponding to $\boldsymbol{p}_{ij}^{gt}, \boldsymbol{n}_{ij}^{gt}$ since otherwise we can just consider the shift from the ground truth.

$\boldsymbol{x}^{\star}(\boldsymbol{y})$ is determined by

$$\begin{aligned}\boldsymbol{x}^{\star}(\boldsymbol{y}) &:= \arg\min_{\{R_i, \boldsymbol{t}_i\}} \sum_{(i,i')\in\mathcal{E}} d^2(S_i, T_i, S_{i'}, T_{i'}) \\ &= \arg\min_{\boldsymbol{x}} f(\boldsymbol{x}, \boldsymbol{y}) \\ &= \arg\min_{\boldsymbol{x}} \sum_{(i,i')\in\mathcal{E}} \sum_{(\boldsymbol{p}_{ij}, \boldsymbol{p}_{i'j'})\in C_{ii'}} \alpha_{iji'j'}^2(\boldsymbol{x}, \boldsymbol{y})\end{aligned}$$

in which

$$d^2(S_i, T_i, S_{i'}, T_{i'}) := \sum_{(\boldsymbol{p}_{ij}, \boldsymbol{p}_{i'j'})\in C_{ii'}} \left((R_i\boldsymbol{p}_{ij} + \boldsymbol{t}_i - R_{i'}\boldsymbol{p}_{i'j'} - \boldsymbol{t}_{i'})^T (R_{i'}\boldsymbol{n}_{i'j'})\right)^2,$$

$$\alpha_{iji'j'}(\boldsymbol{x}, \boldsymbol{y}) = (R_i\boldsymbol{p}_{ij} + \boldsymbol{t}_i - R_{i'}\boldsymbol{p}_{i'j'} - \boldsymbol{t}_{i'})^T (R_{i'}\boldsymbol{n}_{i'j'}),$$

and

$$f(\boldsymbol{x}, \boldsymbol{y}) = \sum_{(\boldsymbol{p}_{ij}, \boldsymbol{p}_{i'j'})\in C} \alpha_{iji'j'}^2(\boldsymbol{x}, \boldsymbol{y}).$$

Introducing notation $\alpha_{iji'j'}$ is to emphasize the dependence upon variables $\boldsymbol{x}$ and $\boldsymbol{y}$. It is worth noting that $\alpha_{iji'j'}(\boldsymbol{x}^{gt}, \boldsymbol{y}^{gt}) = 0$ for all $(i, i,') \in \mathcal{E}$.

By definition

$$\begin{aligned}C &:= \mathrm{Var}[\boldsymbol{x}'(\boldsymbol{y}^{gt}) \cdot \boldsymbol{y}] \\ &= \mathbb{E}\left[\boldsymbol{x}'(\boldsymbol{y}^{gt}) \cdot \boldsymbol{y}\boldsymbol{y}^T \cdot \boldsymbol{x}'(\boldsymbol{y}^{gt})^T\right] \\ &= \sigma^2 \boldsymbol{x}'(\boldsymbol{y}^{gt}) \boldsymbol{x}'(\boldsymbol{y}^{gt})^T,\end{aligned}$$

where $\boldsymbol{x}'(\boldsymbol{y}^{gt})$ is the first order derivative of $\boldsymbol{x}$ at the point of ground truth. It is known that

$$\boldsymbol{x}'(\boldsymbol{y}^{gt}) = \left(\frac{\partial^2}{\partial x^2} f(\boldsymbol{x}^{gt}, \boldsymbol{y}^{gt})\right)^{-1} \frac{\partial^2}{\partial x \partial y} f(\boldsymbol{x}^{gt}, \boldsymbol{y}^{gt}).$$

Here we clarify the format of the variable $\boldsymbol{x}$. Recall that $\boldsymbol{y}$ collects input $\boldsymbol{p}_{ij}$ and $\boldsymbol{n}_{ij}$, and $\boldsymbol{x}$ consists of pose estimation $R_i = \exp(\boldsymbol{c}_i\times)$ and $\boldsymbol{t}_i$. Note $\boldsymbol{x}$ is a $6(N-1)$-dimensional vector since R_1 is always I and $\boldsymbol{t}_1$ is 0. We put $\boldsymbol{t}_i$ in $(6i-11)$th to $(6i-9)$th components of $\boldsymbol{x}$, $\boldsymbol{c}_i$ in $(6i-8)$th to $(6i-6)$th components of $\boldsymbol{x}$.

Define $\boldsymbol{h}_{ij} \in \mathbb{R}^{6(N-1)}$ whose non-zero elements are

$$\begin{aligned}V_{6i-11:6i-9}(\boldsymbol{h}_{ij}) &:= \boldsymbol{n}_{ij}^{gt} \\ V_{6i-8:6i-6}(\boldsymbol{h}_{ij}) &:= \boldsymbol{p}_{ij}^{gt} \times \boldsymbol{n}_{ij}^{gt}\end{aligned}$$

Using such a definition we have

$$\begin{aligned}\frac{\partial^2}{\partial x^2} f(\boldsymbol{x}^{gt}, \boldsymbol{y}^{gt}) &= 2 \sum_{(\boldsymbol{p}_{ij}, \boldsymbol{p}_{i'j'})\in C} \frac{\partial \alpha_{ii'}}{\partial x}(\boldsymbol{x}^{gt}, \boldsymbol{y}^{gt}) \cdot \left(\frac{\partial \alpha_{ii'}}{\partial x}(\boldsymbol{x}^{gt}, \boldsymbol{y}^{gt})\right)^T \\ &= 2 \sum_{(\boldsymbol{p}_{ij}, \boldsymbol{p}_{i'j'})\in C} (\boldsymbol{h}_{ij} - \boldsymbol{h}_{i'j'})(\boldsymbol{h}_{ij} - \boldsymbol{h}_{i'j'})^T,\end{aligned}$$

$$\begin{aligned}\frac{\partial^2}{\partial x \partial y} f(\boldsymbol{x}^{gt}, \boldsymbol{y}^{gt}) &= 2 \sum_{(\boldsymbol{p}_{ij}, \boldsymbol{p}_{i'j'})\in C} \frac{\partial \alpha_{ii'}}{\partial x}(\boldsymbol{x}^{gt}, \boldsymbol{y}^{gt}) \cdot \left(\frac{\partial \alpha_{ii'}}{\partial y}(\boldsymbol{x}^{gt}, \boldsymbol{y}^{gt})\right)^T \\ &= 2 \sum_{(\boldsymbol{p}_{ij}, \boldsymbol{p}_{i'j'})\in C} (\boldsymbol{h}_{ij} - \boldsymbol{h}_{i'j'})\left(\boldsymbol{n}_{i'j'}^T (P_{ij} - P_{i'j'})\right)\end{aligned}$$

where $P_{ij} \in \mathcal{P}_{3,|\boldsymbol{y}|}$ are column permutations so that all P_{ij} have disjoint non-zero columns. Let $\boldsymbol{e}_{ij} \in \mathbb{R}^{\sum N_i}$ be the canonanical basis with the i, j-th element being one. Since

$$\begin{aligned}&(P_{i_1 j_1} - P_{i'_1 j'_1})(P_{i_2 j_2} - P_{i'_2 j'_2})^T \\ =&\big(\mathbb{I}[(i_1, j_1) = (i_2, j_2)] + \mathbb{I}[(i'_1, j'_1) = (i'_2, j'_2)] \\ &- \mathbb{I}[(i'_1, j'_1) = (i_2, j_2)] - \mathbb{I}[(i_1, j_1) = (i'_2, j'_2)]\big) I \\ =&\left(\boldsymbol{e}_{i_1 j_1} - \boldsymbol{e}_{i'_1 j'_1}\right)^T \left(\boldsymbol{e}_{i_2 j_2} - \boldsymbol{e}_{i'_2 j'_2}\right) I.\end{aligned}$$

where $\mathbb{I}$ is the indicator function. Thus

$$\begin{aligned}&\frac{\partial^2 f}{\partial \boldsymbol{x} \partial \boldsymbol{y}} \cdot \mathrm{Var}[\boldsymbol{y}] \cdot \frac{\partial^2 f}{\partial \boldsymbol{y} \partial \boldsymbol{x}} \\ =&4 A_{dis} A_{dis}^T\end{aligned}$$

in which

$$A_{dis} = \sigma \sum_{(\boldsymbol{p}_{ij}, \boldsymbol{p}_{i'j'})\in C} (\boldsymbol{h}_{ij} - \boldsymbol{h}_{i'j'}) \boldsymbol{n}_{i'j'}^T \left((\boldsymbol{e}_{ij} - \boldsymbol{e}_{i'j'})^T \otimes I\right)$$

Finally we obtain that

$$C := \mathrm{Var}[\boldsymbol{x}'(\boldsymbol{y}^{gt}) \cdot \boldsymbol{y}] = C_m^{-1} A_{dis} A_{dis}^T C_m^{-1}.$$

□

C PROOFS OF PROPOSITIONS AND THEOREMS IN SECTION 6

In this section, we provide proofs of the propositions and theorems in Section 6.

C.1 Proof of Prop. 3

The energy function is

$$f = \sum_{i=1}^{N} \sum_{j=1}^{N_i} h_{ij}^2$$

where

$$h_{ij} = (R_i\boldsymbol{p}_{ij} + \boldsymbol{t}_i)^T \boldsymbol{n}_{k_{ij}} - d_{k_{ij}}.$$

Similar to Prop. 2, we have

$$\begin{aligned}\mathrm{Var}[\boldsymbol{x}^{\star}(\boldsymbol{y})] &= \left(\frac{\partial^2 f}{\partial \boldsymbol{x}^2}\right)^{-1} \cdot \frac{\partial^2 f}{\partial \boldsymbol{x} \partial \boldsymbol{y}} \cdot \mathrm{Var}[\boldsymbol{y}] \cdot \frac{\partial^2 f}{\partial \boldsymbol{y} \partial \boldsymbol{x}} \cdot \left(\frac{\partial^2 f}{\partial \boldsymbol{x}^2}\right)^{-1} \\ &= \sigma^2 \left(\frac{\partial^2 f}{\partial \boldsymbol{x}^2}\right)^{-1} \cdot \frac{\partial^2 f}{\partial \boldsymbol{x} \partial \boldsymbol{y}} \cdot \frac{\partial^2 f}{\partial \boldsymbol{y} \partial \boldsymbol{x}} \cdot \left(\frac{\partial^2 f}{\partial \boldsymbol{x}^2}\right)^{-1}, \qquad (56)\end{aligned}$$

in which all partial derivatives were taken at the point of ground truth $(\boldsymbol{x}^{gt}, \boldsymbol{y}^{gt})$.

Note that $h_{ij}(\boldsymbol{x}^{gt}, \boldsymbol{y}^{gt}) = 0$ for ground truth $(\boldsymbol{x}^{gt}, \boldsymbol{y}^{gt})$. Taking advantage of this property we have

$$\frac{\partial^2 h_{ij}^2}{\partial \boldsymbol{x}^2} = 2\left(h_{ij} \frac{\partial^2 h_{ij}}{\partial \boldsymbol{x}^2} + \frac{\partial h_{ij}}{\partial \boldsymbol{x}} \left(\frac{\partial h_{ij}}{\partial \boldsymbol{x}}\right)^T\right) = 2 \frac{\partial h_{ij}}{\partial \boldsymbol{x}} \left(\frac{\partial h_{ij}}{\partial \boldsymbol{x}}\right)^T$$

$$\frac{\partial^2 h_{ij}^2}{\partial \boldsymbol{x}\partial \boldsymbol{y}} = 2\left(h_{ij}\frac{\partial^2 h_{ij}}{\partial \boldsymbol{x}\partial \boldsymbol{y}} + \frac{\partial h_{ij}}{\partial \boldsymbol{x}}\left(\frac{\partial h_{ij}}{\partial \boldsymbol{y}}\right)^T\right) = 2\frac{\partial h_{ij}}{\partial \boldsymbol{x}}\left(\frac{\partial h_{ij}}{\partial \boldsymbol{y}}\right)^T$$

Observe that the only non-zero components in $\frac{\partial h_{ij}}{\partial y}$ is corresponding to parameter $\boldsymbol{p}_{ij}$, which leads to

$$\left(\frac{\partial h_{ij}}{\partial \boldsymbol{y}}\right)^T \frac{\partial h_{i'j'}}{\partial \boldsymbol{y}} = \begin{cases} \boldsymbol{p}_{ij}^T \boldsymbol{p}_{ij} = 1, & \text{if } (i,j) = (i',j') \\ 0, & \text{otherwise} \end{cases}$$

Thus

$$\frac{\partial^2 f}{\partial \boldsymbol{x}\partial \boldsymbol{y}} = 4\sum_{i=1}^{N}\sum_{j=1}^{N_i}\frac{h_{ij}}{\partial \boldsymbol{x}}\left(\frac{h_{ij}}{\partial \boldsymbol{x}}\right)^T = 2\frac{\partial^2 f}{\partial \boldsymbol{x}^2},$$

and furthermore,

$$\mathrm{Var}[\boldsymbol{x}^\star(\boldsymbol{y})] = \sigma^2\left(\sum_{i=1}^{N}\sum_{j=1}^{N_i}\frac{\partial h_{ij}}{\partial \boldsymbol{x}}\left(\frac{\partial h_{ij}}{\partial \boldsymbol{x}}\right)^T\right)^{-1}.$$

Simple calculus figures out that

$$\frac{\partial h_{ij}}{\partial \boldsymbol{x}} = \boldsymbol{g}_{ij}$$

where $\boldsymbol{g}_{ij}$ is just as defined in Prop. 3. Hence the proof completes. □

C.2 Efficient Inverse

One challenge of (22) in the main paper is that for large-scale problems, e.g., hundreds of scans with hundreds of thousands of latent planes, it is infeasible to compute the inverse of C_{XX} directly. In the following, we describe how to leverage the structure in C_{XX} to efficiently compute the pose component $C_{\mathrm{SRAR,Pose}}$ of C_{SRAR}.

The following proposition characterizes the structure of C_{XX}.

Proposition 4. *C_{XX} can be expressed as*

$$C_{XX} = \begin{pmatrix} A & B \\ B^T & \Sigma \end{pmatrix} \tag{57}$$

where $A \in \mathbb{R}^{6(N-1)\times 6(N-1)}$ is a $(N-1)\times(N-1)$ block matrix, where the (i,i')-th block is non-zero if and only if sets of the corresponding planes of the data points of S_{i+1} and $S_{i'+1}$ overlap. $B \in \mathbb{R}^{6(N-1)\times 3K}$ is a $(N-1)\times K$ block matrix, where the (i,k)-th block is non-zero if and only if the k-th plane is the corresponding plane of a data point in S_{i+1}. $\Sigma \in \mathbb{R}^{3K\times 3K}$ is a $K\times K$ block diagonal matrix.

The proof is straight-forward and we omit it for brevity.

The 2×2 block structure of C_{XX} offers a simple way to express the its inverse:

Proposition 5. *The inverse of C_{XX} is given by*

$$\begin{pmatrix} 0 & 0 \\ 0 & \Sigma^{-1} \end{pmatrix} + \begin{pmatrix} I \\ \Sigma^{-1}B^T \end{pmatrix}(A - B\Sigma^{-1}B^T)^{-1}\begin{pmatrix} I \\ \Sigma^{-1}B^T \end{pmatrix}^T \tag{58}$$

Proof. See Appendix C.3. □

Since Σ is a block diagonal matrix, we can use (58) to efficiently encode C_{XX}^{-1}.

C.3 Proof of Proposition 5

Denote

$$\begin{pmatrix} A & B \\ B^T & \Sigma \end{pmatrix}^{-1} = \begin{pmatrix} X & Y \\ Y^T & Z \end{pmatrix}. \tag{59}$$

It follows that

$$AX + BY^T = I \tag{60}$$

$$AY + BZ = 0 \tag{61}$$

$$B^T Y + \Sigma Z = I \tag{62}$$

(62) leads to

$$Z = \Sigma^{-1}(I - B^T Y) \tag{63}$$

Substitute (63) into (61), we obtain

$$(A - B\Sigma^{-1}B^T)Y = -B\Sigma^{-1} \Leftrightarrow Y = -(A - B\Sigma^{-1}B^T)^{-1}B\Sigma^{-1}. \tag{64}$$

Substitute (64) into (62), we obtain

$$\begin{aligned} Z &= \Sigma^{-1}(I + B^T(A - B\Sigma^{-1}B^T)^{-1}B\Sigma^{-1}) \\ &= \Sigma^{-1} + (B\Sigma^{-1})^T(A - B\Sigma^{-1}B^T)^{-1}(B\Sigma^{-1}). \end{aligned} \tag{65}$$

Substitute (64) into (60), we obtain

$$\begin{aligned} X &= A^{-1}(I + B\Sigma^{-1}B^T(A - B\Sigma^{-1}B^T)^{-1}) \\ &= A^{-1} + A^{-1}B\Sigma^{-1}B^T(I - A^{-1}B\Sigma^{-1}B^T)^{-1}A^{-1} \\ &= (I + A^{-1}B\Sigma^{-1}B^T(I - A^{-1}B\Sigma^{-1}B^T)^{-1})A^{-1} \\ &= (I - A^{-1}B\Sigma^{-1}B^T)^{-1}A^{-1} \\ &= (A - B\Sigma^{-1}B^T)^{-1}. \end{aligned} \tag{66}$$

Combing (64),(65), and (66), we end the proof.

□

D PROOFS OF PROPOSITIONS AND THEOREMS IN SECTION 7

D.1 Proof of Theorem 1

First of all, we have

$$\sum_{i=1}^{n}\begin{pmatrix} w_i G_i \\ G_i \end{pmatrix}\cdot\begin{pmatrix} w_i G_i^T & G_i \end{pmatrix} \succeq 0. \tag{67}$$

Denote the left side of (67) as $\begin{pmatrix} A & B \\ B^T & C \end{pmatrix}$. Then we have

$$\begin{pmatrix} A & B \\ B^T & C \end{pmatrix} \succeq 0, \quad A > 0. \tag{68}$$

It follows that

$$\begin{aligned} 0 &\preceq \begin{pmatrix} I & 0 \\ -B^T A^{-1} & I \end{pmatrix}\cdot\begin{pmatrix} A & B \\ B^T & C \end{pmatrix}\cdot\begin{pmatrix} I & -A^{-1}B \\ 0 & I \end{pmatrix} \\ &= \begin{pmatrix} A & 0 \\ 0 & C - B^T A^{-1}B \end{pmatrix} \end{aligned} \tag{69}$$

This means

$$C - B^T A^{-1}B \succeq 0,$$

which ends the proof.