

SPARE: Symmetrized Point-to-Plane Distance for Robust Non-Rigid 3D Registration

Yuxin Yao, Bailin Deng, *Member, IEEE*, Junhui Hou, *Senior Member, IEEE*, Juyong Zhang, *Member, IEEE*

Abstract—Existing optimization-based methods for non-rigid registration typically minimize an alignment error metric based on the point-to-point or point-to-plane distance between corresponding point pairs on the source surface and target surface. However, these metrics can result in slow convergence or a loss of detail. In this paper, we propose *SPARE*, a novel formulation that utilizes a symmetrized point-to-plane distance for robust non-rigid registration. The symmetrized point-to-plane distance relies on both the positions and normals of the corresponding points, resulting in a more accurate approximation of the underlying geometry and can achieve higher accuracy than existing methods. To solve this optimization problem efficiently, we introduce an as-rigid-as-possible regulation term to estimate the deformed normals and propose an alternating minimization solver using a majorization-minimization strategy. Moreover, for effective initialization of the solver, we incorporate a deformation graph-based coarse alignment that improves registration quality and efficiency. Extensive experiments show that the proposed method greatly improves the accuracy of non-rigid registration problems and maintains relatively high solution efficiency. The code is publicly available at <https://github.com/yaoyx689/spare>.

Index Terms—non-rigid registration, symmetrized point-to-plane distance, numerical optimization, surrogate function.

I. INTRODUCTION

GIVEN a source surface and a target surface, non-rigid registration aims to compute a deformation field that aligns the source surface with the target surface. This problem is fundamental in computer vision, with various applications such as 3D shape acquisition and tracking.

In non-rigid registration, the challenge lies in aligning two surfaces without knowing their correspondence in advance. Motivated by the ICP algorithm [1] for rigid registration, many non-rigid registration methods iteratively update the correspondence using closest point queries from the source points to the target surface, and then update the deformation by minimizing an alignment error metric between the corresponding points along with some regularization terms for the deformation field [2]–[4]. The quality and speed of the registration are heavily influenced by the alignment error metric. A commonly

used metric is the “point-to-point” distance, measured between points on the source surface and their corresponding closest points on the target surface [2]–[8]. However, due to the unreliability of such correspondence, the registration may converge to a local minimum and produce unsatisfactory results. Another frequently used metric is the “point-to-plane” distance, which measures the distance between source surface points and the tangent planes at their closest points on the target surface [9]–[13]. As the point-to-plane distance is a first-order approximation of the target surface shape around the closest point, it provides a more accurate proxy of the target shape than the point-to-point distance and enables faster alignment of the two surfaces. Despite these efforts, non-rigid registration can still be challenging and time-consuming. In the rigid registration literature, alignment error metrics that incorporate higher-order geometric properties have been utilized to achieve faster convergence than the point-to-plane distance [14], [15], but such higher-order properties are expensive to compute. Recently, a symmetrized point-to-plane metric was proposed in [16] for rigid registration. It measures the consistency between the positions and normals of each source point and its corresponding closest point, and achieves the minimal value when the pair of points lies on a second-order patch of surface [16]. When used for rigid registration, this gains similar benefits of fast convergence as registration methods that exploit second-order surface properties, but without the need for explicit computation of such properties.

In this paper, we propose *SPARE*, a novel method that utilizes the Symmetrized Point-to-plane distance for Robust non-rigid 3D rEgistration. Although the symmetrized point-to-plane distance works well on rigid registration, deploying it for non-rigid registration is a non-trivial task. First, compared to rigid transformations, accurately modeling the point normals on surfaces undergoing complex non-rigid deformations presents significant computational challenges due to their nonlinear geometric dependencies. Second, there is a large space of point pair positions and normals that can minimize the symmetrized point-to-plane distance. While this is beneficial for rigid registration problems with a limited degree of freedom, it can be problematic for non-rigid registration: due to the excessive degrees of freedom, the symmetrized point-to-plane distance may be minimized without actually aligning the two surfaces. Finally, in real-world applications, non-rigid registration is often carried out between surface data with noise, outliers or partial overlaps, which may lead to erroneous correspondence for closest-point queries and impact the effectiveness of the symmetrized point-to-plane distance. To estimate point normals after deformation, we assume local surface deformation is

This research was in part supported by the National Natural Science Foundation of China (No.62441224, No.62272433, No.62422118), the Hong Kong Research Grants Council under Grants 11219324 and 11219422, and the USTC Fellowship (No.S19582024). (*Corresponding author: Juyong Zhang.*)

Y. Yao is with the School of Mathematical Sciences, University of Science and Technology of China, and also with the Department of Computer Science, City University of Hong Kong. Email: yaoyuxin@mail.ustc.edu.cn.

B. Deng is with the School of Computer Science and Informatics, Cardiff University. Email: DengB3@cardiff.ac.uk.

J. Hou is with the Department of Computer Science, City University of Hong Kong. Email: jh.hou@cityu.edu.hk.

J. Zhang is with the School of Mathematical Sciences, University of Science and Technology of China. Email: juyong@ustc.edu.cn.

nearly rigid. We therefore introduce an as-rigid-as-possible (ARAP) regularization term. This ARAP term, in conjunction with the symmetrized point-to-plane distance, contributes to a more accurate approximation of the distance between surfaces. Furthermore, to address the problem of excessive degrees of freedom, we incorporate a deformation graph-based coarse alignment. This not only helps the deformation field to maintain local structures of the source shape, but also reduces the degrees of freedom and restores the effectiveness of the symmetrized point-to-plane distance for non-rigid registration. In addition, to address the issue of erroneous correspondence due to noise, outliers and partial overlaps, we incorporate an adaptive weight into the symmetrized point-to-plane metric, which is computed according to the positions and normals of each point pair and automatically reduces the influence of unreliable correspondence.

As the resulting optimization problem is non-convex and highly non-linear, we devise an iterative solver that alternately updates the subsets of the variables, decomposing the optimization into simple sub-problems that are easy to solve. In particular, in each iteration we adopt a majorization-minimization strategy [17] and construct a proxy problem that minimizes a simple surrogate function that bounds the target function from above, enabling us to derive closed-form update steps that can effectively reduce the original target function. Moreover, to effectively initialize the solver and avoid undesirable local minima, we propose a strategy that uses a deformation graph [18] to roughly align the two surfaces while maintaining the structure of the source shape.

We perform extensive experiments to evaluate our method on both synthetic and real datasets. Our method outperforms state-of-the-art optimization-based and learning-based methods in terms of accuracy and efficiency. To summarize, our contributions include:

- We formulate a novel optimization-based approach for non-rigid registration problem. Our formulation adopts a robustified symmetrized point-to-plane distance in conjunction with an as-rigid-as-possible regularization term, significantly enhancing the robustness and accuracy of the registration solution.
- We propose an efficient alternating minimization solver for the resulting non-convex optimization problem, using a majorization-minimization strategy to derive simple sub-problems that effectively reduce the target function.
- We devise an effective initialization strategy for the solver, by using a deformation graph-based coarse alignment that maintains the overall structure of the source shape.

II. RELATED WORK

In the following, we review existing works closely related to our approach. More complete overviews of non-rigid registration can be found in recent surveys such as [19], [20].

A. Alignment Error Metric

For optimization-based non-rigid registration, the alignment error metric is an important component in the formulation. Many existing methods adopt the simple point-to-point distance

from the classical ICP algorithm [2], [6], [8], [21], [22]. Others use the point-to-plane distance instead [11]–[13], [23]–[26], which benefits from faster convergence than the point-to-point distance albeit being more complicated to solve. Some methods utilize both metrics to complement each other [7], [9], [27]–[31]. Another class of methods, such as the coherent point drift (CPD) [32], models the sample points using Gaussian mixtures and formulates the alignment error from a probabilistic perspective [32]–[35]. Recently, a Bayesian formulation of CPD has been proposed in [36], with further work in [37], [38] to improve its performance and accuracy. In [39], CPD is also generalized to non-Euclidean domains to improve its robustness on data with large deformations. However, when the data quality is poor, e.g., with the presence of noise or partial overlap, the registration quality will degrade rapidly.

To address this problem, some methods assign individual weights to each point pair in the alignment metric to reduce the impact of erroneous correspondence [2], [6], [7]. Others apply a robust norm (such as the ℓ_0 -norm or the Welsch's function) to the alignment metric, which automatically reduces the contribution of point pairs that are less reliable [8], [12], [30], [40], [41]. Recently, DDM [42] proposes a distance metric based on directional distance fields (DDF) featuring a confidence score for enhanced alignment of 3D geometric surfaces. Separately, [16] introduces a symmetrized point-to-plane metric that considers higher-order properties of surfaces. Our method incorporates this latter metric from [16] with an adaptive weighting scheme for non-rigid registration; this approach yields a robust alignment error metric and ultimately leads to improved alignment performance.

B. Deformation Field

The performance of non-rigid registration also depends on the representation of the deformation field. A simple approach is to introduce a variable for the new position or transformation of each source point [3], [21], [43]. Such a representation provides abundant degrees of freedom to represent complex deformations, but the resulting problem is often expensive to solve due to the large number of variables. Other methods address this problem using an embedded deformation graph [18], where each source point is deformed according to the transformations associated with the nearby nodes of a small graph on the surface [6], [8], [9], [44]. This reduces the number of variables and improves computational efficiency, at the cost of less expressiveness due to fewer degrees of freedom. [45], [46] leverage model priors to parameterize deformations into a set of pose and shape parameters. Recent advancements in neural networks and deep learning have introduced new strategies for representing deformation fields. Some methods directly leverage neural networks for this task [47], [48]. Others focus on reduced-dimensional representations, employing deformation nodes, control points, or neural bones to represent the reduced-dimensional deformation field [49]–[51]. Various specific techniques have also emerged: for instance, [52], [53] iteratively predict the offset of each point based on a per-vertex neural network; [54] introduces a basis point set representation to predict the positions of deformed mesh

points; [55] utilizes a combination of rigid transformations to predict non-rigid deformations. [56], [57] adopt a bijective map to model the reversible deformation field between two shapes; and [58] employs the structure of an encoder and decoder to predict deformation. However, these learning-based approaches can exhibit data dependencies. In this paper, we propose a combination of coarse alignment using a deformation graph and fine registration based on pointwise variables, designed to achieve high solution accuracy while maintaining fast solution speed.

C. Correspondences

Due to initially unknown correspondences, optimization-based methods generally follow the ICP algorithm's paradigm [1], which involves iteratively regarding the closest points as correspondences and applying optimization steps to gradually converge to an optimal solution. In recent years, deep learning techniques have also been adopted to deal with challenging non-rigid registration problems. Many methods rely on supervised learning to improve the quality of the alignment. Some methods learn the correspondences between source and target surfaces from ground-truth correspondences, which helps to handle data with large deformations [59]–[65]. Others use the ground-truth positions of the deformed model as the supervision to train the network to learn the correspondences or the deformation fields [60], [66], [67]. Despite their strong capabilities, these methods rely on training data with ground-truth correspondence or deformation, which is not always easy to obtain. Some unsupervised methods that do not require ground-truth correspondence have also been proposed to address this problem [53], [55], [68], [69]. Such methods still rely on a properly designed loss function that is minimized during training to help the network learn the registration. Some methods [60], [67] incorporate additional information, such as texture or colors, to assist in the establishment of corresponding points. Despite the rapid progress in learning-based methods, research into optimization approaches still advances the field in a complementary way: such approaches do not rely on data and have stronger generalization capability. On the data with relatively small deformation, they can usually achieve higher accuracy. They may be used to compute the required ground-truth information for supervised learning, and they can also contribute to the knowledge of loss function design for unsupervised methods. Furthermore, optimization-based methods can be integrated with correspondences predicted by learning-based methods to reduce reliance on the spatial positions of the source and target, thereby achieving better performance.

III. PROPOSED METHOD

Consider a source surface represented by sample points $\mathcal{V} = \{\mathbf{v}_1, \dots, \mathbf{v}_{|\mathcal{V}|} \in \mathbb{R}^3\}$ equipped with normals $\mathcal{N}_s = \{\mathbf{n}_1, \dots, \mathbf{n}_{|\mathcal{V}|} \in \mathbb{R}^3\}$, and a target surface represented by sample points $\mathcal{U} = \{\mathbf{u}_1, \dots, \mathbf{u}_{|\mathcal{U}|} \in \mathbb{R}^3\}$ with normals $\mathcal{N}_t = \{\mathbf{n}_1^t, \dots, \mathbf{n}_{|\mathcal{U}|}^t \in \mathbb{R}^3\}$. We assume that the neighboring relation between the source sample points in \mathcal{V} is represented by a set of edges $\mathcal{E} = \{(\mathbf{v}_i, \mathbf{v}_j)\}$. This representation is applicable

to both meshes and point clouds: the sample points are either mesh vertices or points within a point cloud, while the edges are either mesh edges or derived from the k -nearest neighbors in the point cloud. The normals can come directly from the input data, be estimated from nearby points using PCA, or calculated by averaging the normals of adjacent faces in a mesh.

We aim to compute a deformation field for the source surface to align it with the target surface. In practical applications, the non-rigid registration problem is inherently challenging due to partial overlap, the lack of known corresponding point pairs, and the presence of noise and outliers in both the source and target surfaces. To this end, we propose a novel optimization formulation for non-rigid registration, utilizing a robust symmetrized point-to-plane (SP2P) distance metric and an as-rigid-as-possible (ARAP) regularization term in the target function (Sections III-A & III-B). To efficiently solve the resulting non-convex optimization, we derive a majorization-minimization (MM) solver that decomposes it into simple sub-problems with closed-form solutions (Section III-C). Furthermore, to avoid local minima and improve the solution accuracy and speed, we introduce a deformation graph-based coarse alignment to initialize our MM solver (Section III-D).

A. Preliminary: Symmetrized Point-To-Plane Distance

A central component for optimization-based registration is the alignment error metric that penalizes the deviation between the source surface and the target surface. Many existing non-rigid registration methods adopt a simple point-to-point metric that originates from rigid registration [1]:

$$E_{PP}^i = \|\hat{\mathbf{v}}_i - \mathbf{u}_{\rho_i}\|^2, \quad (1)$$

where $\hat{\mathbf{v}}_i$ is the new position of \mathbf{v}_i after the deformation, and \mathbf{u}_{ρ_i} is the closest target point to $\hat{\mathbf{v}}_i$. This allows for a simple solver that iteratively updates the closest point correspondence and re-computes the deformation according to the correspondence [2], similar to the classical ICP algorithm for rigid registration [1]. However, this metric may not accurately measure the alignment error: as the source and target point clouds may be sampled from different locations of the underlying surface, they may not fully align under the ground-truth deformation, and the point-to-point metric may not reach its minimum in this case. Moreover, the fixed closest points in each iteration fail to account for the change of correspondence for the moving source points, which can lead to slow convergence, especially when the alignment requires tangential movement along the target surface [15]. To address such issues, other methods adopt a point-to-plane metric from the rigid registration literature [70]:

$$E_{PPL}^i = [\mathbf{n}_{\rho_i}^t \cdot (\hat{\mathbf{v}}_i - \mathbf{u}_{\rho_i})]^2, \quad (2)$$

where $\mathbf{n}_{\rho_i}^t$ is the normal vector at \mathbf{u}_{ρ_i} . The metric penalizes the deviation from the tangent planes at the closest target points. Since the tangent plane is a first-order approximation of the target surface around the closest point, this metric helps to account for the target surface shape away from the sample points and leads to faster convergence than the point-to-point

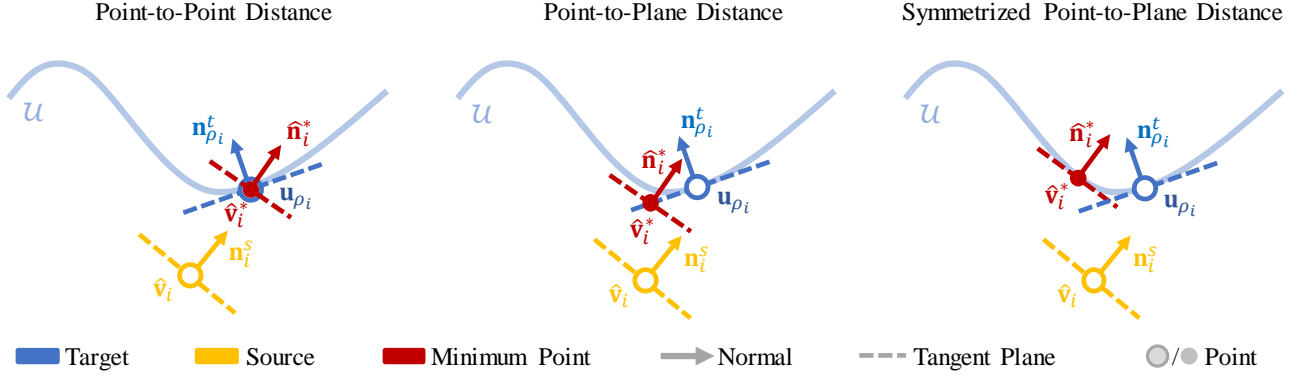


Fig. 1. The target positions of the points on the source surface when minimizing different distance metrics. Here, we only show the position and the normal vector of a potential minimum point. When considering the point-to-plane distance, the position of the minimum point may slide along the tangent plane of the target point. For the symmetrized point-to-plane distance, the minimum point may slide along a symmetric surface that passes through the source point. Additionally, for the point-to-point distance and the point-to-plane distance, the normal vector of the minimum point may rotate around the minimum point.

metric [15]. In the rigid registration literature, alignment error metrics that encode higher-order geometric properties (such as local curvatures) have also been adopted for even faster convergence than the point-to-plane metric [14], [15]. However, such higher-order properties can be expensive to compute.

To achieve faster convergence than the point-to-plane distance without evaluating higher-order properties, a symmetrized point-to-plane (SP2P) metric (see Fig. 1) has been proposed recently in [16] for rigid registration:

$$E_{\text{SPPL}}^i = [(\hat{\mathbf{n}}_i + \mathbf{n}_{\rho_i}^t) \cdot (\hat{\mathbf{v}}_i - \mathbf{u}_{\rho_i})]^2, \quad (3)$$

where $\hat{\mathbf{n}}_i$ is the normal at the point $\hat{\mathbf{v}}_i$ on the deformed surface. Here, the term $(\hat{\mathbf{n}}_i + \mathbf{n}_{\rho_i}^t) \cdot (\hat{\mathbf{v}}_i - \mathbf{u}_{\rho_i})$ vanishes when the corresponding point pair $(\hat{\mathbf{v}}_i, \mathbf{u}_{\rho_i})$ and their normals $(\hat{\mathbf{n}}_i, \mathbf{n}_{\rho_i}^t)$ are consistent with a locally quadratic surface centered between them [16]. In other words, the metric can be considered as penalizing the deviation between the deformed source point $\hat{\mathbf{v}}_i$ and a family of quadratic approximations of the target surface that are consistent with the corresponding target point \mathbf{u}_{ρ_i} and its normal $\mathbf{n}_{\rho_i}^t$. It has been observed in [16] that the symmetrized point-to-plane metric (3) leads to faster convergence of ICP than the point-to-plane metric (2).

B. Non-rigid Registration with Robust SP2P Distance

In this paper, we adopt the SP2P distance in Eq. (3) to propose a new optimization formulation for non-rigid registration, to benefit from its fast convergence. However, in real-world applications, the closest-point correspondence can become unreliable due to partial overlap, noise, outliers, and large initial position differences in the input surfaces. In such cases, the error metric on unreliable point pairs may steer the optimization toward an erroneous alignment. One common solution is to use dynamically adjusted weights for individual source points to control their contribution to the alignment energy based on the reliability of their correspondence [71]. Following this idea, we introduce a weighted alignment term for the source point \mathbf{v}_i as

$$E_{\text{align}}^i = \alpha_i [(\hat{\mathbf{n}}_i + \mathbf{n}_{\rho_i}^t) \cdot (\hat{\mathbf{v}}_i - \mathbf{u}_{\rho_i})]^2, \quad (4)$$

where the weight α_i is computed based on the deformed position $\hat{\mathbf{v}}_i$, the corresponding point \mathbf{u}_{ρ_i} , and their normals $\hat{\mathbf{n}}_i, \mathbf{n}_{\rho_i}^t$ in the previous iteration:

$$\alpha_i = \begin{cases} 0, & \hat{\mathbf{n}}_i' \cdot \mathbf{n}_{\rho_i}^t < 0, \\ \exp(-\frac{\|\hat{\mathbf{v}}_i' - \mathbf{u}_{\rho_i}\|^2}{2\sigma^2}), & \text{otherwise,} \end{cases} \quad (5)$$

where σ is a user-specified parameter. We set σ to the median Euclidean distance from the initial source points to their closest target points. The weight α_i has a large value only if both the positions and normals of the point pair are close to each other, thus reducing the influence of unreliable correspondence.

Another challenge in using the SP2P metric for non-rigid registration is that although it works well for rigid registration, using it alone for non-rigid registration is usually insufficient to achieve good results. This is because compared to the point-to-point and point-to-plane metrics, the SP2P metric has a much larger space of position and normal values with a zero metric value: on the zero level set, the positions and normals of the point pair are only required to be consistent with a certain quadratic surface patch, but the quadratic patch is not necessarily consistent with the target surface. This is not an issue for rigid registration: since all the source points must undergo the same rigid transformation, this induces an implicit regulation for the deformed source points and their normals, forcing them to align with the target surface eventually. Furthermore, in non-rigid registration, the source surface may undergo different deformations, making it challenging to accurately model the normal direction after deformation. We notice that in many non-rigid registration problems such as human body tracking, the surface exhibits local rigidity during deformation. Therefore, we introduce for each source point an as-rigid-as-possible (ARAP) term [72] to help estimate the normal direction after deformation. The ARAP term constrains the deformation in its local neighborhood to be close to a rigid transformation:

$$E_{\text{ARAP}}^i = \frac{1}{|\mathcal{N}(\mathbf{v}_i)|} \sum_{j \in \mathcal{N}(\mathbf{v}_i)} \|(\hat{\mathbf{v}}_i - \hat{\mathbf{v}}_j) - \mathbf{R}_i(\mathbf{v}_i - \mathbf{v}_j)\|^2. \quad (6)$$

Here $\mathcal{N}(\mathbf{v}_i)$ is the index set of neighboring points for \mathbf{v}_i , and $\mathbf{R}_i \in \mathbb{R}^{3 \times 3}$ is an auxiliary rotation matrix variable for a rigid

transformation associated with the source point \mathbf{v}_i (i.e., it needs to satisfy $\mathbf{R}_i^T \mathbf{R}_i = \mathbf{I}$ and $\det(\mathbf{R}_i) = 1$). Therefore, we can use the auxiliary rotation matrix \mathbf{R}_i to approximate the normal of $\hat{\mathbf{v}}_i$ after the deformation as

$$\hat{\mathbf{n}}_i = \mathbf{R}_i \mathbf{n}_i, \quad (7)$$

so that the alignment term in Eq. (4) becomes

$$E_{\text{align}}^i = \alpha_i [(\mathbf{R}_i \mathbf{n}_i + \mathbf{n}_{\rho_i}^t) \cdot (\hat{\mathbf{v}}_i - \mathbf{u}_{\rho_i})]^2. \quad (8)$$

Combining the alignment term (8) and the ARAP term (6) for all source points, we obtain the overall alignment term and ARAP term as

$$E_{\text{align}} = \frac{1}{|\mathcal{V}|} \sum_{\mathbf{v}_i \in \mathcal{V}} E_{\text{align}}^i, \quad E_{\text{ARAP}} = \frac{1}{2|\mathcal{E}|} \sum_{\mathbf{v}_i \in \mathcal{V}} E_{\text{ARAP}}^i. \quad (9)$$

Our non-rigid registration is formulated as an optimization problem

$$\begin{aligned} \min_{\{\hat{\mathbf{v}}_i\}, \{\mathbf{R}_i\}} \quad & E_{\text{align}} + w_{\text{ARAP}} E_{\text{ARAP}} \\ \text{s.t.} \quad & \mathbf{R}_i^T \mathbf{R}_i = \mathbf{I}, \det(\mathbf{R}_i) = 1, \quad \forall i, \end{aligned} \quad (10)$$

where w_{ARAP} is a user-specified parameter.

C. Numerical Solver

The optimization problem (10) is challenging due to the non-linearity, non-convexity, and constraints of rotation matrices. To solve it efficiently and effectively, we devise an iterative solver that alternately updates the closest points $\{\mathbf{u}_{\rho_i}\}$, the position variables $\{\hat{\mathbf{v}}_i\}$, and the rotation matrix variables $\{\mathbf{R}_i\}$ while fixing the remaining variables. We denote their values after the k -th iteration as $\{\mathbf{u}_{\rho_i}^{(k)}\}$, $\{\hat{\mathbf{v}}_i^{(k)}\}$, and $\{\mathbf{R}_i^{(k)}\}$, respectively. In the following, we explain how to update their values in the $(k+1)$ -th iteration.

Update of $\{\mathbf{u}_{\rho_i}\}$. Following the solver in [16], we fix the source point positions $\hat{\mathbf{v}}_i^{(k)}$ and update closest points via

$$\rho_i^{(k+1)} = \arg \min_{\rho_i \in \{1, \dots, |\mathcal{U}|\}} \|\mathbf{u}_{\rho_i} - \hat{\mathbf{v}}_i^{(k)}\|. \quad (11)$$

Afterward, we also compute the updated robust weights $\{\alpha_i^{(k+1)}\}$ according to Eq. (5).

Update of $\{\hat{\mathbf{v}}_i\}$. After updating $\{\mathbf{u}_{\rho_i}\}$, we fix the closest points $\{\mathbf{u}_{\rho_i^{(k+1)}}\}$, rotations $\{\mathbf{R}_i^{(k)}\}$, and robust weights $\{\alpha_i^{(k+1)}\}$, and minimize the target function in Eq. (10) with respect to $\{\hat{\mathbf{v}}_i\}$. It is a linear least-squares problem and can be solved via a linear system. Details can be found in Appx. A-A.

Update of $\{\mathbf{R}_i\}$. Finally, we update the rotation matrices $\{\mathbf{R}_i\}$ by fixing $\{\hat{\mathbf{v}}_i^{(k+1)}\}$ and $\{\mathbf{u}_{\rho_i^{(k+1)}}\}$ and minimizing the target function with respect to $\{\mathbf{R}_i\}$. This is reduced to an independent sub-problem for each \mathbf{R}_i :

$$\begin{aligned} \min_{\mathbf{R}_i} \quad & \alpha_i^{(k+1)} [(\mathbf{R}_i \mathbf{n}_i + \mathbf{n}_{\rho_i}^t) \cdot (\hat{\mathbf{v}}_i^{(k+1)} - \mathbf{u}_{\rho_i^{(k+1)}})]^2 \\ & + \omega \sum_{\mathbf{v}_j \in \mathcal{N}(\mathbf{v}_i)} \|(\hat{\mathbf{v}}_i^{(k+1)} - \hat{\mathbf{v}}_j^{(k+1)}) - \mathbf{R}_i(\mathbf{v}_i - \mathbf{v}_j)\|^2, \\ \text{s.t.} \quad & \mathbf{R}_i^T \mathbf{R}_i = \mathbf{I}, \det(\mathbf{R}_i) = 1, \end{aligned} \quad (12)$$

where $\omega = \frac{w_{\text{ARAP}} \cdot |\mathcal{V}|}{|\mathcal{N}(\mathbf{v}_i)| \cdot 2|\mathcal{E}|}$. As far as we are aware, there is no closed-form solution to this problem. For an efficient update, we adopt the idea of majorization-minimization (MM) [17] and minimize a convex surrogate function $\bar{F}(\mathbf{R}_i | \mathbf{R}_i^{(k)})$ that is constructed based on the current variable value $\mathbf{R}_i^{(k)}$. The surrogate function should bound the original target function F from above and have the same value as F at $\mathbf{R}_i^{(k)}$, i.e.,

$$\begin{aligned} \bar{F}(\mathbf{R}_i | \mathbf{R}_i^{(k)}) &\geq F(\mathbf{R}_i) \quad \forall \mathbf{R}_i, \\ \bar{F}(\mathbf{R}_i^{(k)} | \mathbf{R}_i^{(k)}) &= F(\mathbf{R}_i^{(k)}). \end{aligned} \quad (13)$$

As a result, the minimizer of $\bar{F}(\mathbf{R}_i | \mathbf{R}_i^{(k)})$ is guaranteed to decrease the original target function F compared to $\mathbf{R}_i^{(k)}$, unless $\mathbf{R}_i^{(k)}$ is already a local minimum of F (in which case the minimizer of $\bar{F}(\mathbf{R}_i | \mathbf{R}_i^{(k)})$ will be exactly $\mathbf{R}_i^{(k)}$) [17]. In the following, we will derive a simple surrogate function that allows for a closed-form solution, enabling fast and effective update of $\{\mathbf{R}_i\}$.

First, we note that the squared term $[(\mathbf{R}_i \mathbf{n}_i + \mathbf{n}_{\rho_i}^t) \cdot (\hat{\mathbf{v}}_i^{(k+1)} - \mathbf{u}_{\rho_i^{(k+1)}})]^2$ in Eq. (12) can be written as

$$f(\mathbf{R}_i) = [(\mathbf{R}_i \mathbf{n}_i + \mathbf{n}_i^t) \cdot \mathbf{d}]^2, \quad (14)$$

where $\mathbf{d} = \hat{\mathbf{v}}_i^{(k+1)} - \mathbf{u}_{\rho_i^{(k+1)}}$ and $\mathbf{n}_i^t = \mathbf{n}_{\rho_i^{(k+1)}}^t$. Then, we have

Proposition 1. If $\mathbf{d} \neq \mathbf{0}$, then $f(\mathbf{R}_i)$ in Eq. (14) satisfies

$$f(\mathbf{R}_i) = \|\mathbf{d}\|^2 \cdot \min_{\mathbf{h} \in \mathcal{P}} \|\mathbf{R}_i \mathbf{n}_i - \mathbf{h}\|^2, \quad (15)$$

where \mathcal{P} is a plane containing all vectors $\mathbf{h} \in \mathbb{R}^3$ that satisfy

$$(\mathbf{h} + \mathbf{n}_i^t) \cdot \mathbf{d} = 0, \quad (16)$$

i.e., $f(\mathbf{R}_i)$ is the squared distance from $\mathbf{R}_i \mathbf{n}_i$ to the plane \mathcal{P} , scaled by a factor $\|\mathbf{d}\|^2$.

Proof. If $\mathbf{d} \neq \mathbf{0}$, then for any vector \mathbf{h} satisfying (16) we have

$$\begin{aligned} f(\mathbf{R}_i) &= [(\mathbf{R}_i \mathbf{n}_i - \mathbf{h} + \mathbf{h} + \mathbf{n}_i^t) \cdot \mathbf{d}]^2 \\ &= [(\mathbf{R}_i \mathbf{n}_i - \mathbf{h}) \cdot \mathbf{d} + (\mathbf{h} + \mathbf{n}_i^t) \cdot \mathbf{d}]^2 \\ &= [(\mathbf{R}_i \mathbf{n}_i - \mathbf{h}) \cdot \mathbf{d}]^2. \end{aligned} \quad (17)$$

Moreover, the plane \mathcal{P} defined by (16) has a unit normal vector $\frac{\mathbf{d}}{\|\mathbf{d}\|}$. Since \mathbf{h} belongs to the plane \mathcal{P} , we have

$$[(\mathbf{R}_i \mathbf{n}_i - \mathbf{h}) \cdot \mathbf{d}]^2 = \|\mathbf{d}\|^2 \cdot \left[(\mathbf{R}_i \mathbf{n}_i - \mathbf{h}) \cdot \frac{\mathbf{d}}{\|\mathbf{d}\|} \right]^2. \quad (18)$$

Note that $(\mathbf{R}_i \mathbf{n}_i - \mathbf{h}) \cdot \frac{\mathbf{d}}{\|\mathbf{d}\|}$ is the signed distance from $\mathbf{R}_i \mathbf{n}_i$ to the plane \mathcal{P} . Therefore,

$$\left[(\mathbf{R}_i \mathbf{n}_i - \mathbf{h}) \cdot \frac{\mathbf{d}}{\|\mathbf{d}\|} \right]^2 = \min_{\mathbf{h} \in \mathcal{P}} \|\mathbf{R}_i \mathbf{n}_i - \mathbf{h}\|^2. \quad (19)$$

Using Eqs. (17), (18) and (19), we obtain Eq. (15). \square

Based on Eq. (15), we can construct a surrogate function for f at the variable value $\mathbf{R}_i^{(k)}$ as:

$$\bar{f}(\mathbf{R}_i | \mathbf{R}_i^{(k)}) = \|\mathbf{d}\|^2 \cdot \|\mathbf{R}_i \mathbf{n}_i - \mathbf{h}_*^{(k)}\|^2, \quad (20)$$

where $\mathbf{h}_*^{(k)}$ is the closest projection of $\mathbf{R}_i^{(k)} \mathbf{n}_i$ onto the plane \mathcal{P} , i.e., $\mathbf{h}_*^{(k)} = \arg \min_{\mathbf{h} \in \mathcal{P}} \|\mathbf{R}_i^{(k)} \mathbf{n}_i - \mathbf{h}\|$ (see Appx. A-B for

a proof that $\bar{f}(\mathbf{R}_i | \mathbf{R}_i^{(k)})$ satisfies the conditions of a surrogate function).

It is easy to show that $\mathbf{h}_*^{(k)}$ can be computed as

$$\mathbf{h}_*^{(k)} = \mathbf{R}_i^{(k)} \mathbf{n}_i - \mathbf{d} \frac{(\mathbf{n}_i^t + \mathbf{R}_i^{(k)} \mathbf{n}_i) \cdot \mathbf{d}}{\|\mathbf{d}\|^2}. \quad (21)$$

By replacing $f(\mathbf{R}_i)$ with its surrogate function in Eq. (20), we replace the optimization problem (12) with the following proxy problem:

$$\begin{aligned} \min_{\mathbf{R}_i} & \alpha_i^{(k+1)} \|\mathbf{d}\|^2 \cdot \|\mathbf{R}_i \mathbf{n}_i - \mathbf{h}_*^{(k)}\|^2 \\ & + \omega \sum_{\mathbf{v}_j \in \mathcal{N}(\mathbf{v}_i)} \|(\hat{\mathbf{v}}_i^{(k+1)} - \hat{\mathbf{v}}_j^{(k+1)}) - \mathbf{R}_i(\mathbf{v}_i - \mathbf{v}_j)\|^2, \\ \text{s.t. } & \mathbf{R}_i^T \mathbf{R}_i = \mathbf{I}, \quad \det(\mathbf{R}_i) = 1. \end{aligned}$$

This problem has a closed-form solution [73]:

$$\mathbf{R}_i^{(k+1)} = \mathbf{V}_S \begin{bmatrix} 1 & & \\ & 1 & \\ & & \det(\mathbf{V}_S \mathbf{U}_S^T) \end{bmatrix} \mathbf{U}_S^T, \quad (22)$$

where the matrices \mathbf{U}_S , \mathbf{V}_S are from the SVD

$$\mathbf{S} = \mathbf{U}_S \mathbf{\Sigma}_S \mathbf{V}_S^T$$

for the matrix

$$\begin{aligned} \mathbf{S} &= \alpha_i^{(k+1)} \|\mathbf{d}\|^2 \mathbf{n}_i (\mathbf{h}_*^{(k)})^T \\ &+ \omega \sum_{\mathbf{v}_j \in \mathcal{N}(\mathbf{v}_i)} (\mathbf{v}_i - \mathbf{v}_j) (\hat{\mathbf{v}}_i^{(k+1)} - \hat{\mathbf{v}}_j^{(k+1)})^T \\ &= \alpha_i^{(k+1)} \left(\|\mathbf{d}\|^2 \mathbf{n}_i (\hat{\mathbf{n}}_i^{(k)})^T - \mathbf{n}_i \mathbf{d}^T ((\mathbf{n}_{\rho_i^{(k+1)}}^t + \hat{\mathbf{n}}_i^{(k)}) \cdot \mathbf{d}) \right) \\ &+ \omega \sum_{\mathbf{v}_j \in \mathcal{N}(\mathbf{v}_i)} (\mathbf{v}_i - \mathbf{v}_j) (\hat{\mathbf{v}}_i^{(k+1)} - \hat{\mathbf{v}}_j^{(k+1)})^T \end{aligned} \quad (23)$$

where $\hat{\mathbf{n}}_i^{(k)} = \mathbf{R}_i^{(k)} \mathbf{n}_i$. Note that although the above derivation is based on Proposition 1 which requires $\mathbf{d} \neq \mathbf{0}$, the solution in Eq. (22) remains effective when $\mathbf{d} = \mathbf{0}$: in this case, the target function in (12) reduces to

$$\omega \sum_{\mathbf{v}_j \in \mathcal{N}(\mathbf{v}_i)} \|(\hat{\mathbf{v}}_i^{(k+1)} - \hat{\mathbf{v}}_j^{(k+1)}) - \mathbf{R}_i(\mathbf{v}_i - \mathbf{v}_j)\|^2,$$

and the matrix \mathbf{S} in Eq. (23) becomes

$$\mathbf{S} = \omega \sum_{\mathbf{v}_j \in \mathcal{N}(\mathbf{v}_i)} (\mathbf{v}_i - \mathbf{v}_j) (\hat{\mathbf{v}}_i^{(k+1)} - \hat{\mathbf{v}}_j^{(k+1)})^T.$$

Then the matrix $\mathbf{R}_i^{(k+1)}$ in Eq. (22) is exactly the solution to the reduced optimization problem [73]. Later in Sec. IV-C, we will showcase the benefits of this solution for updating $\{\mathbf{R}_i\}$.

Termination criteria. We stop the iteration if at least one of the following conditions is satisfied: (1) the ℓ_2 norm of the point position changes in an iteration is less than a threshold, i.e., $\|\hat{\mathbf{V}}^{(k+1)} - \hat{\mathbf{V}}^{(k)}\|/\sqrt{|\mathcal{V}|} < \epsilon$, where ϵ is a user-specified parameter (we set $\epsilon = 10^{-4}$ in our experiments); (2) the number of iterations reaches an upper bound K (we set $K = 30$). Algorithm 1 summarizes our solver for the optimization problem in Eq. (10).

Algorithm 1: Non-rigid registration with robust SP2P distance.

Input: $\{\mathbf{v}_i, \mathbf{n}_i\}_{i=1}^{|\mathcal{V}|}$: the source points and normals;
 $\{\mathbf{u}_i, \mathbf{n}_i^t\}_{i=1}^{|\mathcal{U}|}$: the target points and normals;
 K : maximum number of iterations;
 ϵ : convergence threshold.

Result: The deformed point positions $\{\hat{\mathbf{v}}_i\}_{i=1}^{|\mathcal{V}|}$.

```

1 Set  $\mathbf{R}_i^{(0)} = \mathbf{I}$  and  $\hat{\mathbf{v}}_i^{(0)} = \mathbf{v}_i$  for all  $i$ ;
2  $k = 0$ ;
3 while  $k < K$  and  $\|\hat{\mathbf{V}}^{(k+1)} - \hat{\mathbf{V}}^{(k)}\|/\sqrt{|\mathcal{V}|} < \epsilon$  do
4   For each  $i \in \mathcal{V}$ , find the closest point  $\mathbf{u}_{\rho_i^{(k+1)}}$  for
      $\hat{\mathbf{v}}_i^{(k)}$ ;
5   Compute weight  $\alpha_i^{(k+1)}$  with Eq. (5);
6   Compute  $\{\hat{\mathbf{v}}_i^{(k+1)}\}$  via linear system (32);
7   Compute  $\{\mathbf{R}_i^{(k+1)}\}$  with Eq. (22);
8    $\hat{\mathbf{n}}_i^{(k+1)} = \mathbf{R}_i^{(k+1)} \mathbf{n}_i$ ;
9    $k = k + 1$ ;
10 end while

```

D. Coarse Alignment Using a Deformation Graph

Our numerical solver presented in Sec. III-C is a local solver that searches for a stationary point near the initial solution. Therefore, proper initialization is crucial for achieving desirable results. To this end, we initialize the solver with a coarse alignment computed using a deformation graph [18]. The deformation graph controls the source surface shape with a reduced number of variables, allowing us to efficiently determine a deformation that roughly aligns the two surfaces while preserving the structure of the source shape.

Specifically, to construct a deformation graph, we follow [4] and first uniformly sample a subset $\mathcal{V}_G = \{\mathbf{p}_1, \dots, \mathbf{p}_{|\mathcal{V}_G|}\}$ from \mathcal{V} as the deformation graph nodes, so that the number of nodes is much smaller than the number of source points. Next, we establish an edge set \mathcal{E}_G by connecting the neighboring nodes, thereby deriving a deformation graph $\mathcal{G} = \{\mathcal{V}_G, \mathcal{E}_G\}$. We then assign to each node an affine transformation, represented by a matrix $\mathbf{A}_j \in \mathbb{R}^{3 \times 3}$ and a vector $\mathbf{t}_j \in \mathbb{R}^3$. The transformations at the nodes determine the deformation of each source point \mathbf{v}_i as

$$\hat{\mathbf{v}}_i = \sum_{\mathbf{p}_j \in \mathcal{I}(\mathbf{v}_i)} w_{ij} \cdot (\mathbf{A}_j(\mathbf{v}_i - \mathbf{p}_j) + \mathbf{p}_j + \mathbf{t}_j), \quad (24)$$

where $\mathcal{I}(\mathbf{v}_i) = \{\mathbf{p}_j \in \mathcal{V}_G \mid D(\mathbf{v}_i, \mathbf{p}_j) < R\}$ is the node set that affects \mathbf{v}_i , with $D(\cdot, \cdot)$ denoting the geodesic distance and R being the sampling radius. We set $R = 10 \cdot \bar{l}_s$ by default, where \bar{l}_s is the average edge length of the source surface. The weight w_{ij} for the influence of \mathbf{p}_j on \mathbf{v}_i is defined as [4]:

$$w_{ij} = \frac{(1 - D^2(\mathbf{v}_i, \mathbf{p}_j)/R^2)^3}{\sum_{\mathbf{p}_k \in \mathcal{I}(\mathbf{v}_i)} (1 - D^2(\mathbf{v}_i, \mathbf{p}_k)/R^2)^3}.$$

Using the deformation graph, we perform a coarse alignment by optimizing the transformation variables $\{(\mathbf{A}_j, \mathbf{t}_j)\}$ associated with the nodes.

The target function is a combination of the following terms:

- **Alignment and ARAP terms.** We use an alignment term E_{align}^C and an ARAP term E_{ARAP}^C similar to the ones presented in Sec. III-B but with two main differences: (1) the deformed source point position $\hat{\mathbf{v}}_i$ is computed from the deformation graph according to Eq. (24); (2) as only a rough alignment is needed, we apply the alignment term only to a sampled subset \mathcal{S} of the source points \mathcal{V} to reduce computation, i.e.,

$$E_{\text{align}}^C = \frac{1}{|\mathcal{S}|} \sum_{\mathbf{v}_s \in \mathcal{S}} E_{\text{align}}^s, \quad (25)$$

where E_{align}^s is defined in the same way as Eq. (4). By default, we set the number of sampling points to 3000.

- **Other regularization terms.** To ensure the deformation graph induces a smooth deformation that preserves the structure of the source shape, we introduce two additional regularization terms E_{smo} and E_{rot} from [4] to enforce the smoothness and rigidity of the transformations, respectively.

Using these terms, our optimization problem for coarse alignment can be written as

$$\min_{\{(\mathbf{A}_j, \mathbf{t}_j)\}, \{\mathbf{R}_i\}} E_{\text{align}}^C + w_{\text{ARAP}}^C E_{\text{ARAP}}^C + w_{\text{smo}} E_{\text{smo}} + w_{\text{rot}} E_{\text{rot}}, \quad (26)$$

where $\{\mathbf{R}_i\}$ are auxiliary rotation matrix variables for the alignment and ARAP terms. Similar to Sec. III-C, we solve this problem by alternating updates of the variables using an MM strategy. Details of the solver can be found in Appx. A-C.

IV. RESULTS

We conducted comprehensive performance comparisons of the proposed method with the state-of-the-art non-rigid registration methods. In addition, we evaluated the effectiveness of the components in our formulation. This section provides details of our experiment settings and results.

A. Experiment Settings

To assess the effectiveness and accuracy of our method, we conducted comparisons with several existing methods: the non-rigid ICP method (N-ICP) from [2]; the Welsch function-based formulation from [4] (AMM); the Bayesian Coherent Point Drift method (BCPD) [36] and its variants BCPD++ [37] and GBCPD/GBCPD++ [38]. Additionally, we compared our method with state-of-the-art learning-based methods, including LNDP [74], SyNoRiM [75], and GraphSCNet [63]. The comparisons were performed using the open-source implementations of these methods^{1,2,3,4,5,6}. All methods were evaluated using a synthetic dataset, the DeformingThings4D (DT4D) dataset [76], and real datasets including the articulated mesh animation (AMA) dataset [77], the SHREC'20 Track dataset [78] the DFAUST dataset [46] and the BEHAVE dataset [79]. Additionally, we assessed all optimization-based methods using real datasets including DFAUST [46], DeepDeform [60] and the

face sequence from [27], to test their effectiveness for non-rigidly tracking. For each dataset, we tuned the parameters of each optimization-based method to achieve the best overall performance. For learning-based methods, we used the pre-trained models provided by the authors for testing. Detailed parameter settings can be found in Appx. C.

The optimization-based methods were run on a PC with 32GB of RAM and a 6-core Intel Core i7-8700K CPU at 3.70GHz. They were all running on Ubuntu 20.04 LTS system built with Docker and all leveraged multi-thread acceleration on the CPU. All learning-based methods were running on a server equipped with an NVIDIA RTX A6000 GPU and utilized CUDA acceleration. For all problem instances, we scaled the source surface and target surface with the same scaling factor, such that the two surfaces are contained in a bounding box with a unit diagonal length to test all comparison methods. To have a clear error scale, we rescale the model to its original size and calculate the error in meters. In the subsequent presentation of numerical results in tables, we highlight the best results using bold fonts, while underlining the second-best results for clarity and emphasis. In all result figures, we render the target surface in yellow, while the source surface and the deformed surface are rendered in blue.

For the evaluation metric, we calculated the root-mean-square error between the deformed positions and the ground-truth positions, referring to [4], [8], to measure the registration error:

$$\text{RMSE} = \sqrt{\frac{1}{|\mathcal{V}|} \sum_{\mathbf{v}_i \in \mathcal{V}} \|\hat{\mathbf{v}}_i - \hat{\mathbf{v}}_i^*\|^2}, \quad (27)$$

where $\hat{\mathbf{v}}_i^*$ is the ground-truth positions for $\hat{\mathbf{v}}_i$. Since some data do not have ground-truth deformed positions but provide ground-truth correspondences, we also calculate the distance error between the correspondences, similar to [80], [81]:

$$\text{Corr}_{\text{err}} = \frac{1}{|\mathcal{C}|} \sum_{(i,j) \in \mathcal{C}} \mathcal{D}_{\mathcal{V}}(\mathbf{v}_{\tau_j}, \mathbf{v}_i), \quad (28)$$

where \mathcal{C} is the set of ground-truth correspondence pairs. For a given target point \mathbf{u}_j , \mathbf{v}_i is the ground-truth corresponding point on the source surface, while \mathbf{v}_{τ_j} is the corresponding point obtained from the registration result where its deformed position $\hat{\mathbf{v}}_{\tau_j}$ is the closest to \mathbf{u}_j on the deformed source surface. When the geodesic distance can be calculated for surface \mathcal{V} , $\mathcal{D}_{\mathcal{V}}(\cdot, \cdot)$ represents the geodesic distance between two points on surface \mathcal{V} . Otherwise, for example, if the source surface is discretized as a mesh with multiple connected parts, $\mathcal{D}_{\mathcal{V}}(\cdot, \cdot)$ represents the Euclidean distance between two points. In the following experiments, only the ‘‘crane’’ sequences from the AMA dataset [77] with partial overlaps use the Euclidean distance for error calculation. In addition, following established practices [80], [81], we plotted cumulative curves (representing the percentage of correspondences with errors below a varying threshold) and computed the area under these curves (AUC).

B. Comparison with State-of-the-Art Methods

Clean data. We evaluated different registration methods using the AMA dataset [77] to assess their performance in continuous

¹https://github.com/Juyong/Fast_RNRR

²https://github.com/yaoyx689/AMM_NRR

³<https://github.com/ohirose/bcpd>

⁴<https://github.com/rabbityl/DeformationPyramid>

⁵<https://github.com/huangjh-pub/synorim>

⁶<https://github.com/qinzheng93/GraphSCNet>

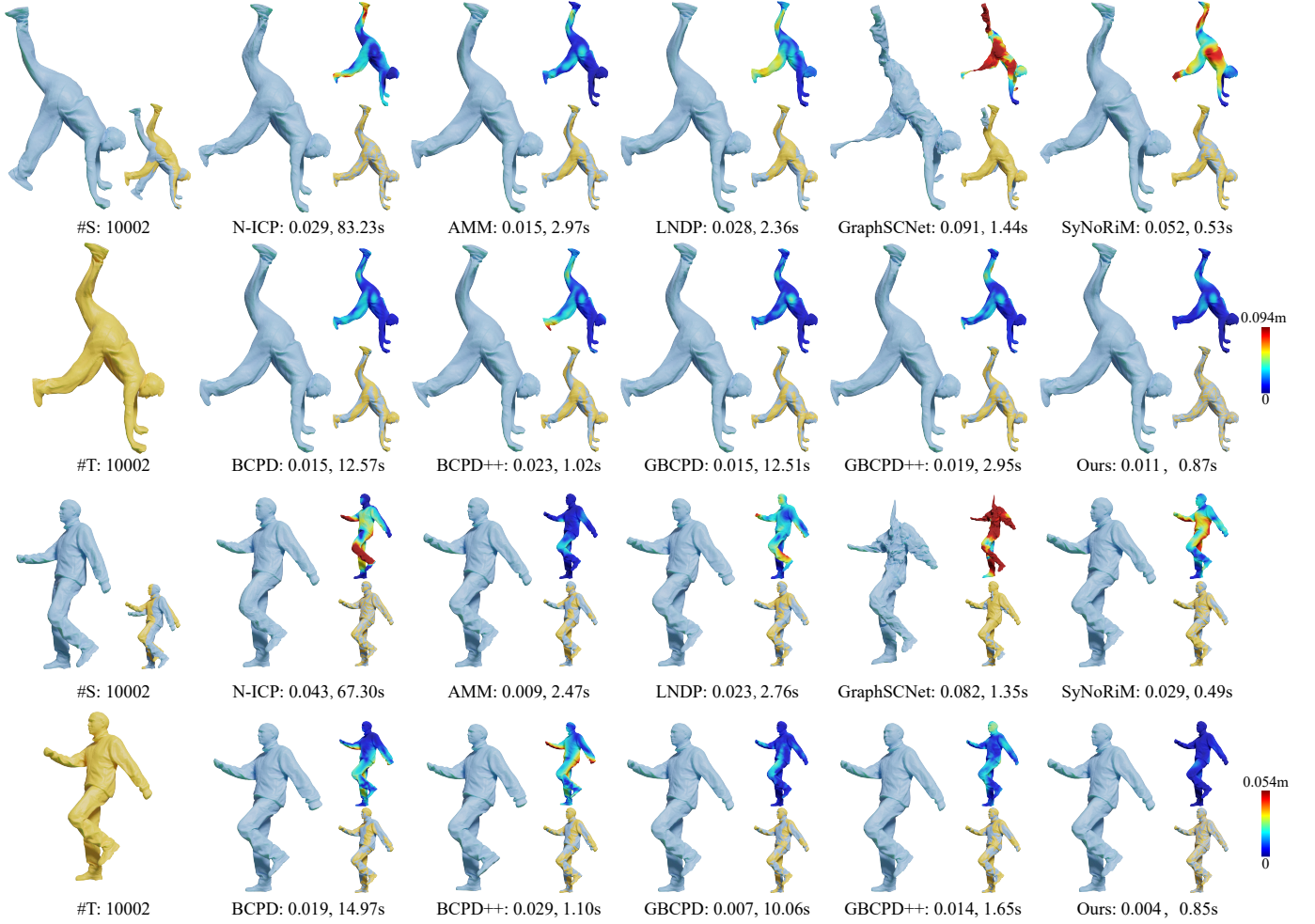


Fig. 2. Results from different methods on two problem instances from the “handstand”(top) and “march1”(bottom) sequences from [77]. For each method, we show the deformed mesh (left), alignment result (right-bottom), and an error map (right-top) that visualizes the distance between each point and its ground-truth corresponding points, as well as the RMSE and the computational time.

TABLE I
AVERAGE VALUES OF RMSE ($\times 0.01$) ↓ / CORR_{err} ($\times 0.01$) ↓ / AUC ↑ AND COMPUTATIONAL TIME (S) ON “HANDSTAND” AND “MARCH1” SEQUENCES FROM [77].

Method	handstand		march1	
	Accuracy	Time	Accuracy	Time
N-ICP [2]	3.03 / 2.17 / 0.81	66.58	1.67 / 1.18 / 0.89	39.84
AMM [4]	1.19 / 0.62 / 0.94	2.35	0.70 / 0.21 / 0.98	1.51
BCPD [36]	1.29 / 0.79 / 0.92	4.97	1.18 / 0.76 / 0.93	5.40
BCPD++ [37]	1.71 / 1.31 / 0.87	3.21	1.74 / 1.29 / 0.88	3.33
GBCPD [38]	1.14 / 0.62 / 0.94	10.07	0.55 / 0.15 / 0.99	9.53
GBCPD++ [38]	1.46 / 1.09 / 0.89	2.68	1.08 / 0.73 / 0.93	2.38
GraphSCNet [63]	9.44 / 7.61 / 0.40	1.95	7.23 / 6.96 / 0.46	1.85
LNDDP [74]	2.23 / 1.93 / 0.81	1.90	1.85 / 1.39 / 0.87	1.43
SyNoRiM [75]	3.20 / 2.83 / 0.73	0.54	2.20 / 1.93 / 0.83	0.48
Ours	0.86 / 0.38 / 0.96	<u>0.96</u>	0.26 / 0.03 / 1.00	<u>0.80</u>

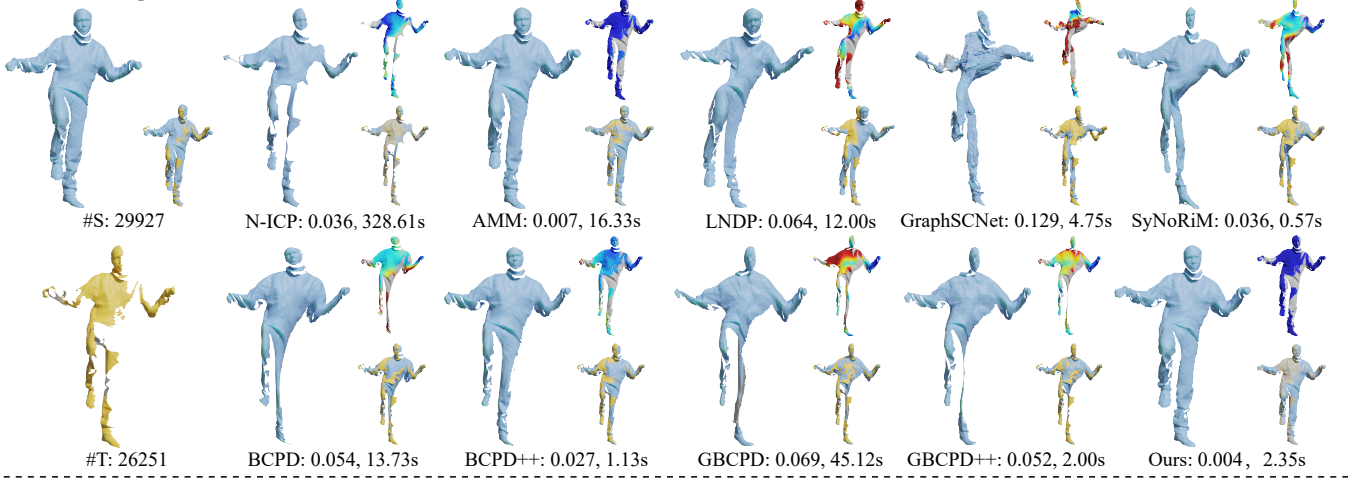
sequence scenarios. The dataset consists of 10 sequences of human continuous motion captured from real-world scenarios, and data have been processed into triangular meshes with the same connectivity structure in each sequence. Following [4], we focused on the “handstand” and “march1” sequences from [77]. For each sequence, we selected 50 pairs of models to test,

considering the i -th mesh as the source model and the $(i+2)$ -th mesh as the target model, where $8 \leq i < 48$.

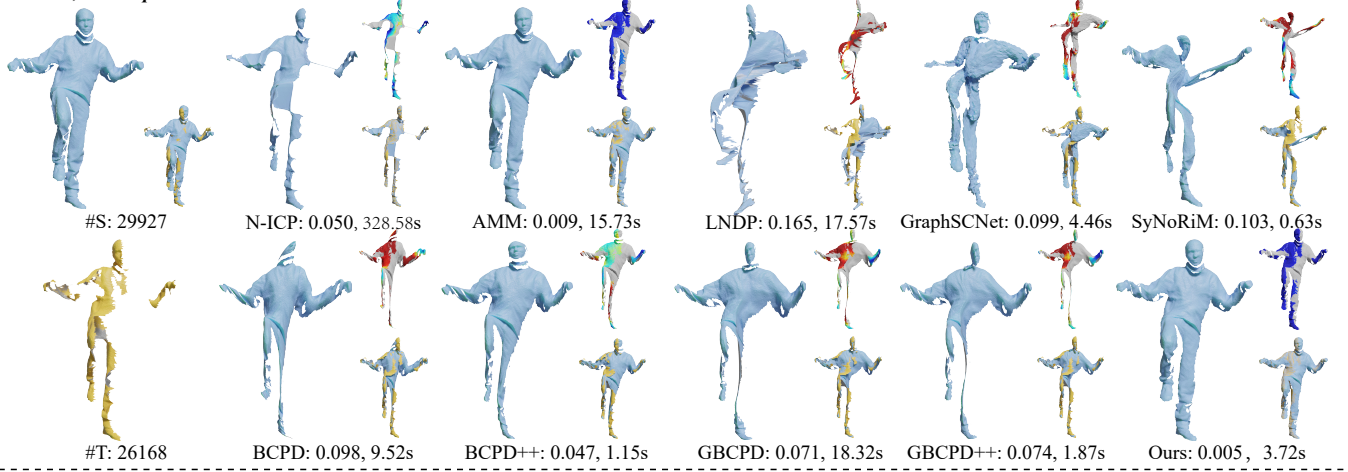
In Tab. I, we present the average values of RMSE, Corr_{err}, AUC and average computation time for each sequence. Additionally, we provide visualizations of two specific cases in Fig. 2 and the cumulative curves in Fig. 6. From the results, it can be observed that our method achieves the highest accuracy in terms of RMSE, Corr_{err}, and AUC among all compared methods. Furthermore, the average computation time of our method is the second shortest and only longer than SyNoRiM, which is a learning-based method with GPU acceleration.

Partial overlaps. In practice, many non-rigid registration problems involve surface pairs with partial overlaps that increase the difficulty of registration. To evaluate the methods on such data, we used the “crane” sequence from the AMA dataset [77] as test cases. We first selected ten pairs of meshes $\{(m_j^s, m_j^t) \mid j = 1, \dots, 10\}$ from the crane sequence, each consisting of two adjacent frames in the sequence. For each mesh pair, we derived nine pairs of point clouds by simulating depth cameras from a fixed view angle v^s for the source mesh m_j^s , and from nine different view angles for angle $\{v_1^t, \dots, v_9^t\}$ for the target mesh m_j^t with increasing deviation from the

1st view, Overlap rate: 71%



4th view, Overlap rate: 60%



7th view, Overlap rate: 42%

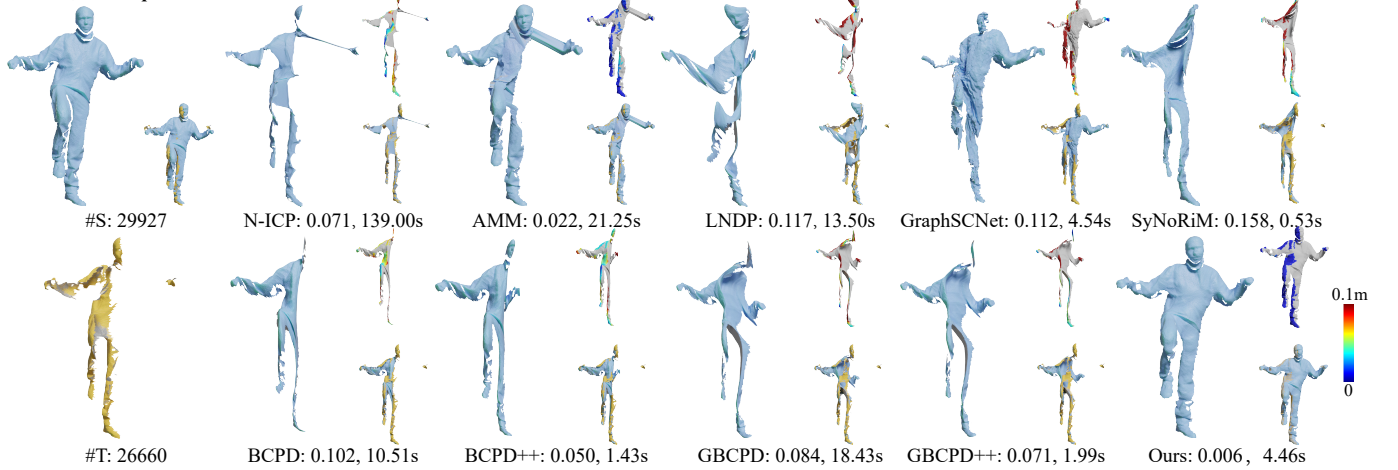


Fig. 3. The results obtained from different methods on problem instances with partial overlaps from [77]. For each method, we show the deformed mesh (left), the alignment result (right-bottom), and an error map (right-top) that visualizes the distance between the ground-truth correspondences in the overlapping area defined by Eq. (29) (Points with no correspondence are marked in gray), and label the value of Corr_{err} and the computational time.

source view angle. This results in nine pairs of point clouds for each mesh pair, with decreasing overlap ratios. we report the average values of the following overlap ratio o for all pairs

using the same view angles:

$$o = \frac{|\{\mathbf{v}_i \mid \|\hat{\mathbf{v}}_i^* - \mathbf{u}_{\rho_i^{(*)}}\| < \bar{l}_t/\sqrt{3}\}|}{|\mathcal{V}|}, \quad (29)$$

where $\hat{\mathbf{v}}_i^*$ is \mathbf{v}_i 's position under the ground-truth deformation, $\mathbf{u}_{\rho_i^{(*)}}$ is the closest target point to $\hat{\mathbf{v}}_i^*$, and \bar{l}_t is the average

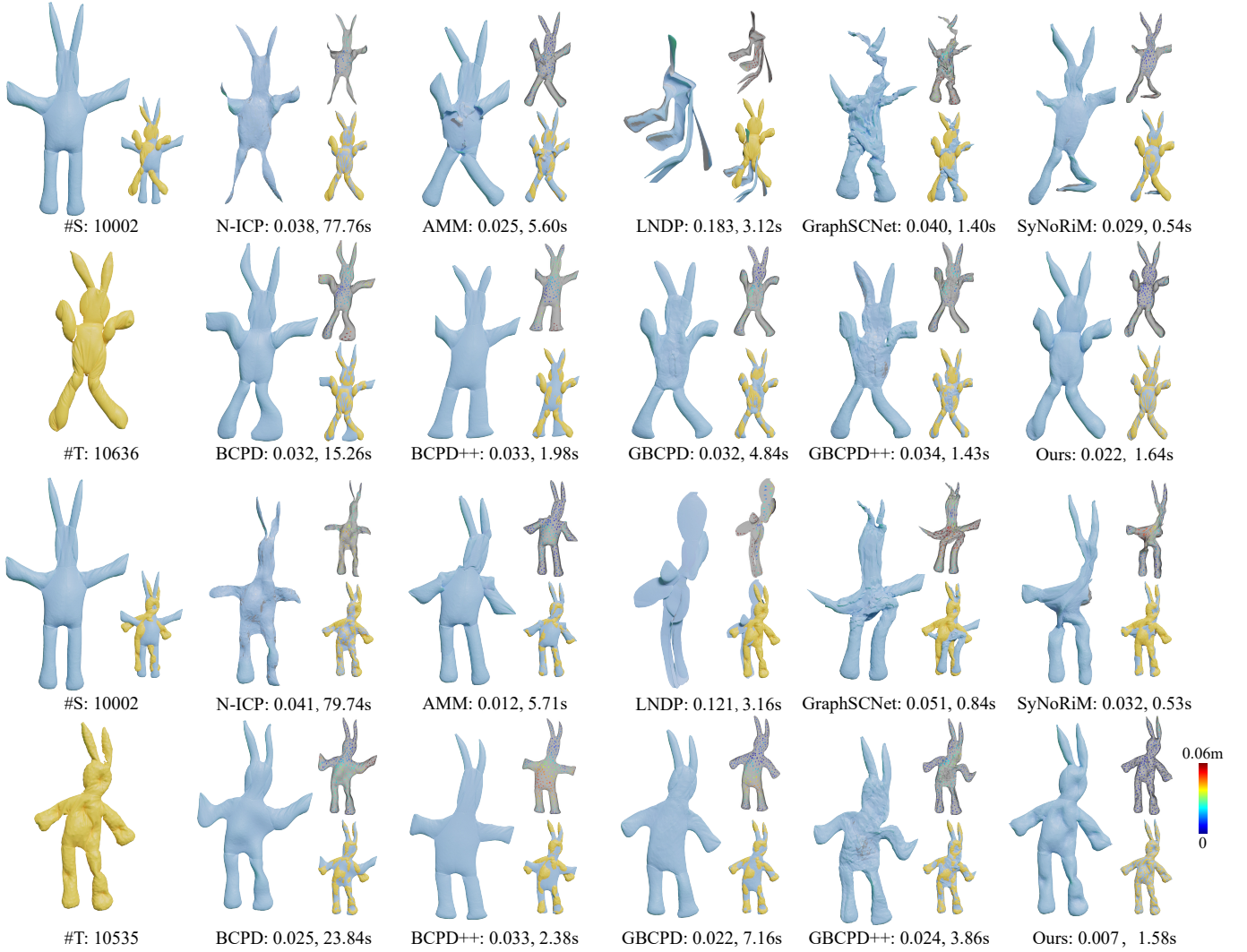


Fig. 4. The results obtained from different methods on two problem instances from the SHREC'20 non-rigid correspondence dataset [78]. For each method, we show the deformed mesh (left), the alignment result (right-bottom), and an error map (right-top) that visualizes the distance between the ground-truth correspondences (Points with no correspondences are marked in gray). We also label the value of Corr_{err} and the computational time for each method.

TABLE II

MEAN VALUES OF $\text{CORR}_{\text{ERR}} \downarrow (\times 0.001)$ AND $\text{AUC} \uparrow$ ON “CRANE” SEQUENCE FROM AMA DATASET [77] WITH PARTIAL OVERLAPS. EACH COLUMN (VIEW ANGLE PAIR (THE AVERAGE OVERLAP RADIO ϕ (%))) SHOWS THE STATISTICS FOR A PARTICULAR PAIR OF VIEW ANGLES.

Method	1 (76.30)	2 (73.55)	3 (70.94)	4 (67.98)	5 (64.43)	6 (60.71)	7 (56.60)	8 (52.50)	9 (48.61)
N-ICP [2]	3.04 / 0.71	3.23 / 0.69	3.28 / 0.68	3.42 / 0.67	3.64 / 0.65	3.98 / 0.62	4.26 / 0.60	4.62 / 0.57	4.79 / 0.56
AMM [4]	0.79 / 0.92	0.79 / 0.92	0.83 / 0.92	0.85 / 0.92	1.05 / 0.90	0.99 / 0.90	1.17 / 0.89	1.43 / 0.87	2.00 / 0.83
BCPD [36]	6.49 / 0.51	7.12 / 0.45	7.80 / 0.41	9.78 / 0.29	10.68 / 0.24	11.37 / 0.22	12.34 / 0.20	12.41 / 0.18	13.36 / 0.17
BCPD++ [37]	3.30 / 0.70	3.54 / 0.68	3.70 / 0.66	4.29 / 0.61	4.50 / 0.59	4.85 / 0.58	6.02 / 0.47	6.48 / 0.43	7.81 / 0.38
GBCPD [38]	7.59 / 0.43	9.02 / 0.34	9.68 / 0.32	9.67 / 0.30	10.33 / 0.27	10.57 / 0.25	10.88 / 0.24	11.39 / 0.23	11.96 / 0.22
GBCPD++ [38]	7.30 / 0.41	8.09 / 0.37	8.34 / 0.34	8.52 / 0.34	9.61 / 0.29	9.86 / 0.27	10.10 / 0.26	10.60 / 0.25	11.21 / 0.23
GraphSCNet [63]	11.55 / 0.21	11.78 / 0.19	12.34 / 0.17	12.91 / 0.17	13.23 / 0.18	12.78 / 0.17	13.60 / 0.16	13.21 / 0.15	12.88 / 0.15
LNDP [74]	9.51 / 0.44	11.67 / 0.39	10.85 / 0.38	14.28 / 0.32	14.33 / 0.35	13.76 / 0.34	17.33 / 0.28	17.62 / 0.30	13.73 / 0.35
SyNoRiM [75]	3.89 / 0.65	4.42 / 0.60	5.01 / 0.57	4.79 / 0.56	6.41 / 0.49	7.35 / 0.45	6.85 / 0.46	8.93 / 0.40	11.17 / 0.35
Ours	0.41 / 0.96	0.42 / 0.96	0.42 / 0.96	0.44 / 0.96	0.46 / 0.96	0.48 / 0.95	0.49 / 0.95	0.50 / 0.95	0.57 / 0.95

distance between neighboring points on the target shape. That is, ϕ represents the proportion of source points whose distance to the target shape, under the ground-truth deformation, is smaller than a threshold related to the sampling density. Based on the results shown in Tab. II, Fig. 3 and Fig. 6, it is evident that the proposed method achieved the highest matching accuracy

among all the compared methods.

In addition, we utilized the models from the SHREC'20 dataset [78] to evaluate the performance of the registration methods. The dataset consists of a complete mesh and 11 partial meshes of a real-world object, where the partial meshes exhibit diverse shapes with fine details, resulting from different

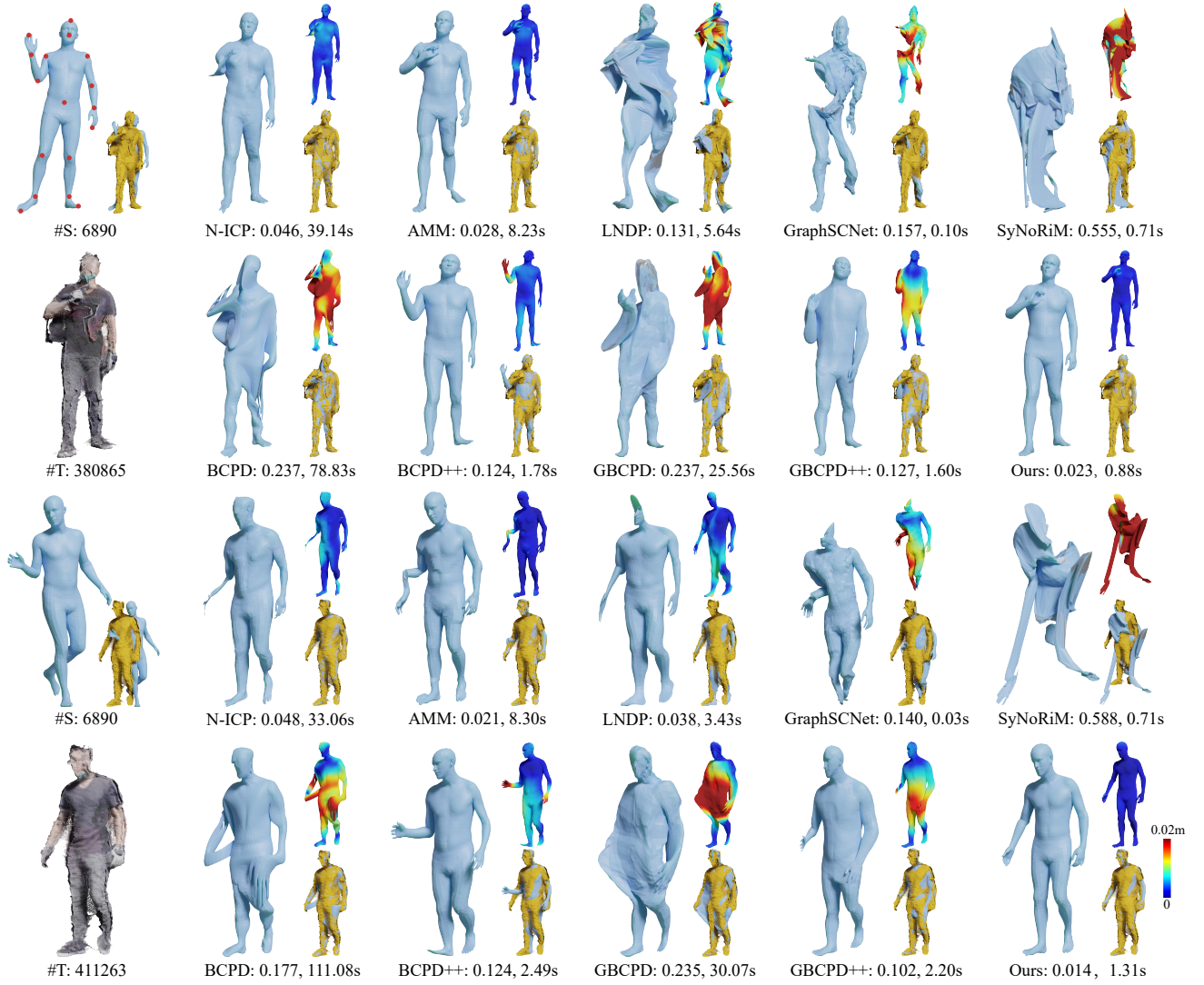


Fig. 5. The results obtained from different methods on two problem instances from the BEHAVE dataset [79]. For each method, we show the deformed mesh (left), the alignment results (right-bottom), and an error map (right-top) that visualizes the distance between each point and the ground-truth positions (right-top). We also label the value of RMSE and the computational time. The red dots on the source surface in the top-left corner mark the locations of the landmark points.

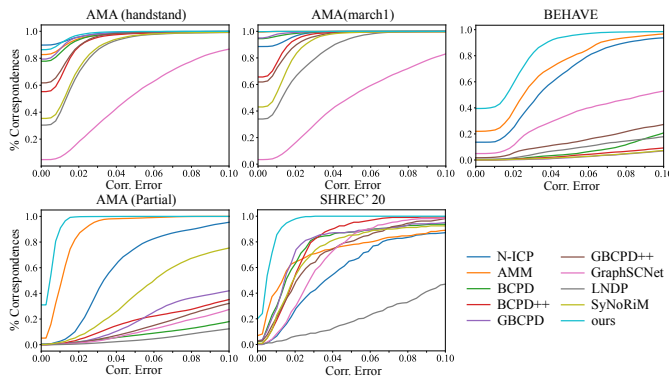


Fig. 6. The cumulative curves for the correspondence error Corr_{err} of different methods on various datasets.

TABLE III
MEAN VALUES OF $D_{\text{PP}}^{\text{T-S}} (\times 10^{-5})$, $D_{\text{PPL}}^{\text{T-S}} (\times 10^{-5})$, $\text{Corr}_{\text{err}} (\times 0.01)$, AUC AND AVERAGE COMPUTATIONAL TIME (S) USING DIFFERENT METHODS ON THE SHREC'20 NON-RIGID CORRESPONDENCE DATASET [78].

Method	$D_{\text{PP}}^{\text{T-S}} \downarrow$	$D_{\text{PPL}}^{\text{T-S}} \downarrow$	$\text{Corr}_{\text{err}} \downarrow$	AUC \uparrow	Time
N-ICP [2]	66.43	37.12	4.59	0.58	79.59
AMM [4]	16.81	13.05	2.70	0.76	5.95
BCPD [36]	6.81	4.56	3.23	0.69	16.80
BCPD++ [37]	10.11	6.70	3.26	0.68	2.05
GBCPD [38]	1.07	0.58	2.79	0.73	7.17
GBCPD++ [38]	1.12	0.56	2.98	0.72	1.84
GraphSCNet [63]	37.10	23.99	5.33	0.52	1.59
LNDP [74]	203.71	114.46	14.11	0.23	3.06
SyNoRiM [75]	47.03	28.27	8.58	0.50	0.54
Ours	0.41	0.10	1.47	0.86	1.69

deformations (stretch, indent, twist and inflate) and different filling materials. We considered the full mesh as the source

surface, and the partial meshes as the target surfaces. We compute the Corr_{err} with the provided sparse ground-truth correspondences. To further evaluate the errors of dense points,

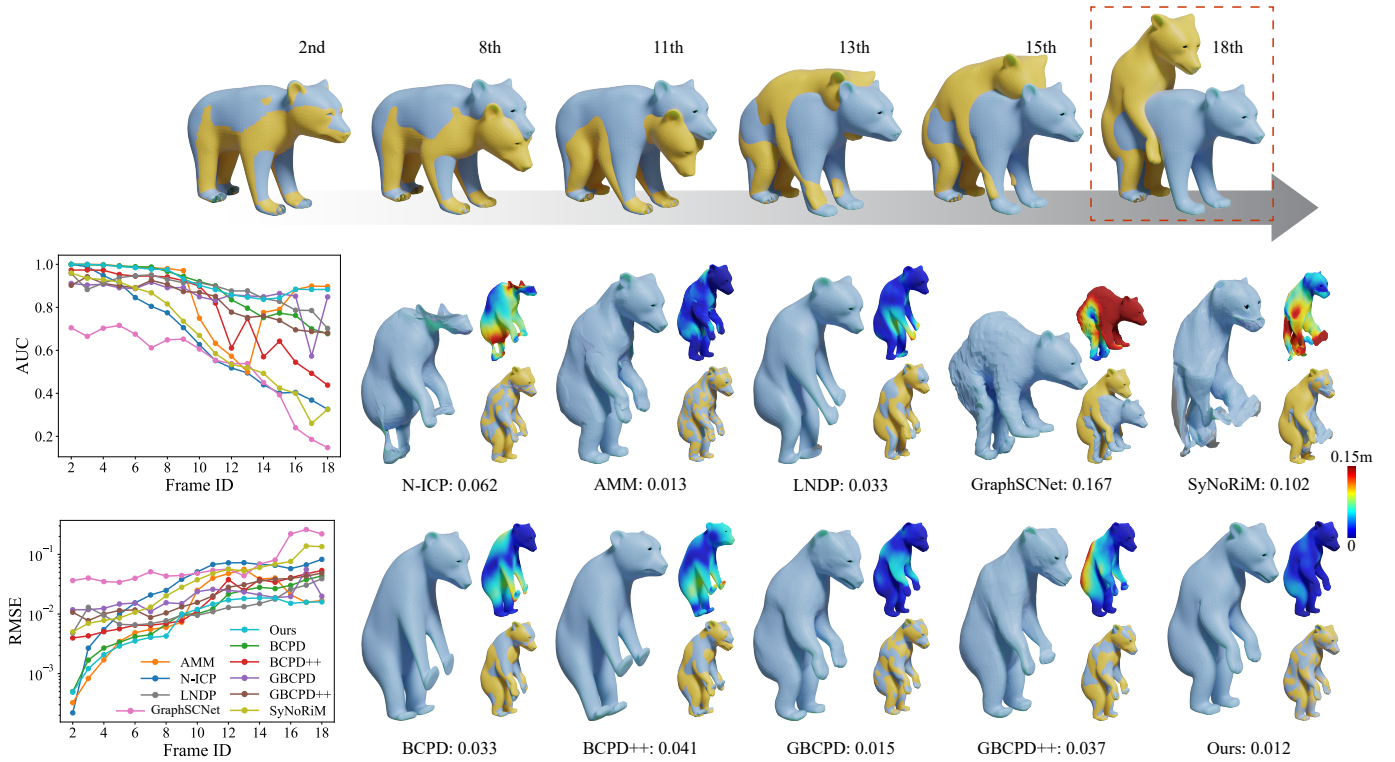


Fig. 7. The results obtained from different methods on DT4D data [76] with gradually increasing deformation differences. The top shows the source and target models for several frames. The line graphs in the lower left corner show the results of AUC \uparrow and RMSE \downarrow changing with the indices of the frame respectively. The result of the 18th frame is visualized in the lower right corner. For each method, we show the deformed mesh (left), alignment result (right-bottom), and an error map (right-top) that visualizes the distance between the ground-truth correspondences, as well as label the RMSE.

we calculate the point-to-point distance D_{pp}^{T-S} and the point-to-plane distance D_{ppl}^{T-S} from the partial meshes to the deformed source surfaces. Tab. III shows the numerical results, and Fig. 4 visualizes two specific cases. We can see that our proposed method achieves notably higher accuracy compared to other methods in the evaluation. This superior performance can be attributed to the high degrees of freedom provided by our dense deformation field, as well as the reweighting scheme that can effectively handle partial overlaps. Moreover, our method achieves the best speed among optimization-based methods, and only falls behind the GPU-accelerated learning-based methods SyNoRiM and GraphSCNet.

Noisy data. We assessed our method’s performance on noisy data using the BEHAVE dataset [79], which depicts human-object interactions and includes typical outliers produced by depth sensors. From this dataset, we selected the “Data01” sequence, containing 35 distinct motion sequences of an individual. Each sequence in BEHAVE provides frame-by-frame human body point clouds and their corresponding fitted SMPL [82] mesh. In total, we constructed 1586 test cases, using the SMPL mesh of the i -th frame as the source model, the point cloud of the $(i+1)$ -th frame as the target model, and the SMPL mesh of the $(i+1)$ -th frame as the ground-truth mesh for error calculation. A significant challenge arose from the large inter-frame differences within each sequence (as processed by [79], see Fig. 5), which impacted the performance of all evaluated methods. To mitigate this, we manually annotated 17 landmark corresponding points (marked on the SMPL mesh

TABLE IV
MEAN VALUES OF RMSE($\times 0.01$), CORR_{err}($\times 0.01$), AUC AND AVERAGE COMPUTATIONAL TIME (S) USING DIFFERENT METHODS ON THE BEHAVE DATASET [79].

Method	RMSE \downarrow	CORR _{err} \downarrow	AUC \uparrow	Time
N-ICP [2]	4.04	4.51	0.66	37.15
AMM [4]	4.11	4.73	0.68	8.32
BCPD [36]	23.53	33.89	0.17	82.34
BCPD++ [37]	15.94	21.22	0.40	2.52
GBCPD [38]	26.99	36.93	0.16	47.76
GBCPD++ [38]	18.12	25.12	0.31	2.14
GraphSCNet [63]	16.19	20.76	0.22	0.04
LNBP [74]	10.81	15.69	0.42	2.17
SyNoRiM [75]	74.32	56.86	0.03	<u>0.71</u>
Ours	2.37	2.62	0.81	1.13

and transferred to raw point clouds.) and integrated a landmark term similar to that proposed in [2] into the registration process. These landmarks were utilized by our method as well as all other compared approaches, with the exception of the BCPD-family of methods, which do not accommodate landmark inputs. As shown in Table IV, Fig. 5 and Fig. 6, the comparison results demonstrate that our method outperforms existing techniques.

Large deformation. To verify the robustness of our method to the magnitude of deformation, we conducted evaluations using data with gradually increasing deformation differences. Specifically, we selected a sequence (“bear3EP_Aggression”) from the DT4D dataset [76] and two sequences (“hips” and “jumping-jacks”) from the DFAUST dataset [46] to test our

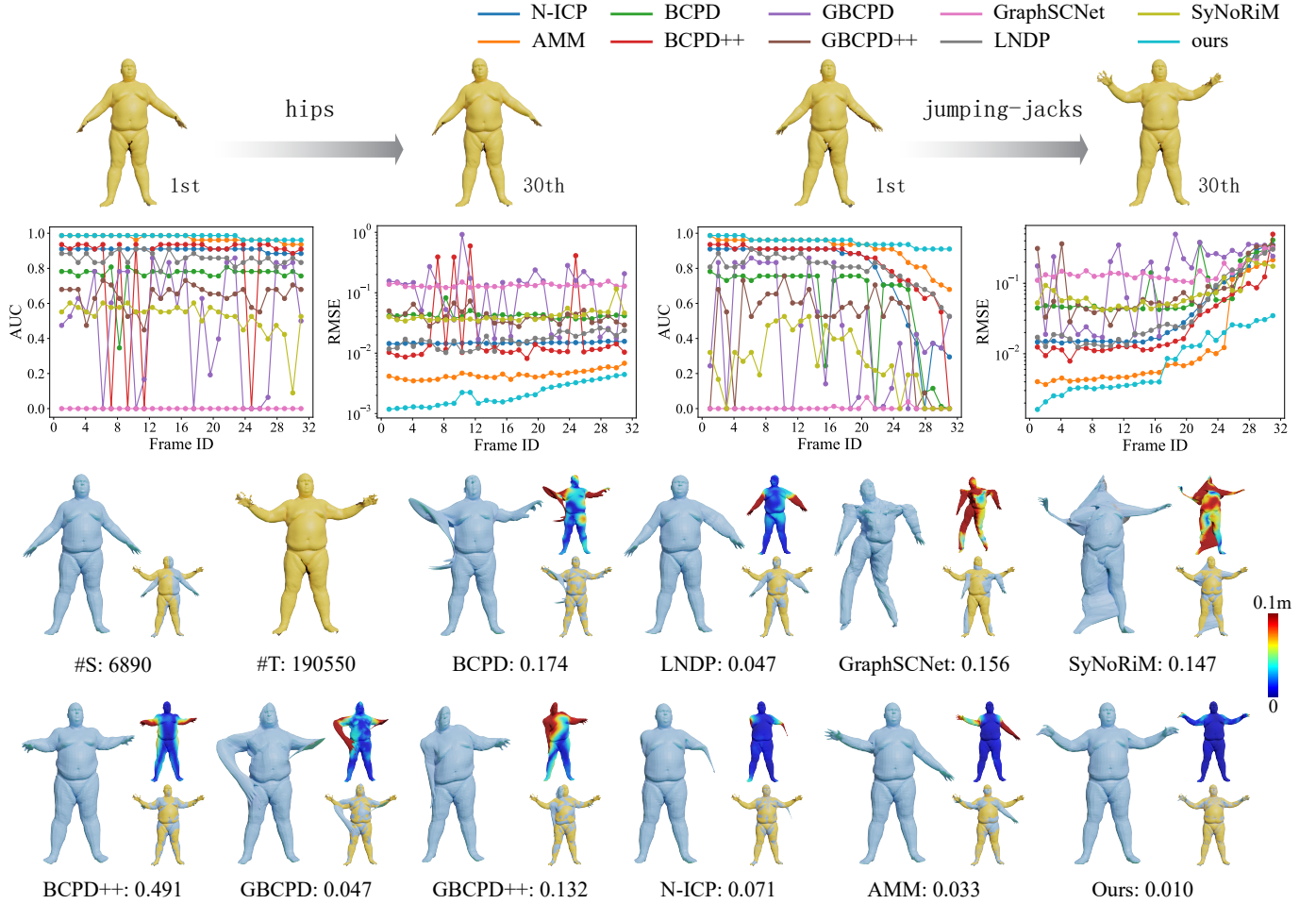


Fig. 8. The results obtained from different methods on DFAUST dataset [46] with gradually increasing deformation difference. The line graphs in the top show the results of AUC \uparrow and RMSE \downarrow changing with the indices of the frame respectively in “hips” and “jumping-jacks” sequences. The result of the 29th frame in “jumping-jacks” is visualized in the bottom. For each method, we show the deformed mesh (left), alignment result (right-bottom), and an error map (right-top) that visualizes the distance between the ground-truth correspondences, as well as label the RMSE.

method. DT4D is a synthetic dataset that includes continuous motion sequences of various animals and humanoids. The shapes in each sequence have ground-truth correspondences. We set the first mesh as the source model and the i -th mesh as the target model ($i=2, \dots, 18$) for this sequence. DFAUST contains multiple continuously moving human scans and the corresponding SMPL mesh [82]. We set the SMPL mesh corresponding to the first frame as the source model and the i -th scan as the target model ($i=2, \dots, 30$) to test the performance. The corresponding i -th SMPL mesh is used to obtain ground-truth correspondences for calculating the correspondence error Corr_{err} . From Fig. 7 and Fig. 8, we can observe that as the deformation increases, our method achieves higher or comparable accuracy compared to existing methods.

Non-rigid tracking. To evaluate the practicality of the proposed method, we conducted non-rigid tracking experiments on real-world data, including the DFAUST dataset [46], the DeepDeform dataset [60] and the face sequence from [27]. For a real scan sequence $\{\mathcal{T}_1, \dots, \mathcal{T}_N\}$, we choose a template model \mathcal{S}_0 , and register it to \mathcal{T}_1 and obtain the deformed model $\hat{\mathcal{S}}_1$. Then we deform $\hat{\mathcal{S}}_1$ to align with \mathcal{T}_{i+1} ($i = 1, \dots, N - 1$) and

obtain a mesh sequence $\{\hat{\mathcal{S}}_i\}_{i=1}^N$ aligned with $\{\mathcal{T}_i\}_{i=1}^N$. Due to the accumulation of errors in the registration process, this setting increases the difficulty of the problem and makes it easier to demonstrate the performance of different methods. Since there is usually a small difference between two adjacent frames, it is generally solved using optimization-based methods. Since learning-based methods excel in acquiring global correspondences but often struggle with capturing intricate details and exhibit limited stability in continuous registration, we ignore the comparisons with learning-based methods in these examples.

We utilized the template mesh from the SMPL model matching the first frame as \mathcal{S}_0 , and deformed it to align the following i -th human scan, where $1 \leq i < 200$. We also used the template meshes of the SMPL model matched to these scans as the ground-truth meshes. We chose four sequences: “hips”, “jiggle-on-toes”, “jumping-jacks” and “chicken-wings” of the identity labeled as 50002 to test, where the deformations of the latter two sequences are much larger than those of the first two sequences. We compute RMSE between the registered results with the ground-truth meshes, and show their average values and the average running time in Tab. V. We also provide

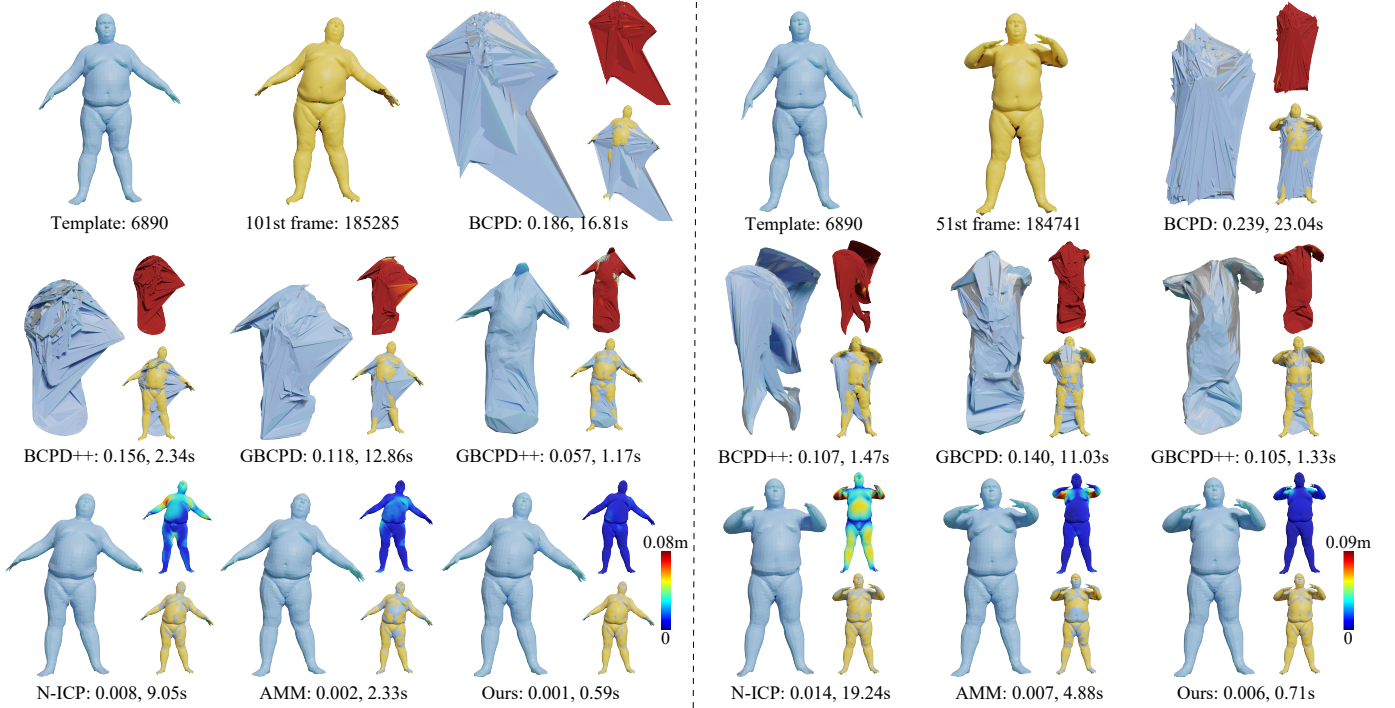


Fig. 9. Results from different methods on two problem instances from “hips”(left) and “chicken-wings” (right) of the DFAUST dataset [46]. For each method, we show the deformed mesh (left), the alignment result (right-bottom), and an error map (right-top) that visualizes the distance between each point and its ground-truth positions, and label the value of RMSE and the computational time.

TABLE V

MEAN VALUES OF RMSE($\times 0.01$) AND AVERAGE COMPUTATIONAL TIME (s) USING DIFFERENT METHODS ON THE DFAUST DATASET [46]. DUE TO ERROR ACCUMULATION, N-ICP ONLY OBTAINED VALID RESULTS ON SOME SEQUENCES (195 FOR “JIGGLE-ON-TOES”, 74 FOR “JUMPING-JACKS”, AND 97 FOR “CHICKEN-WINGS”). WE ONLY MAKE STATISTICS ON THESE VALID VALUES.

Method	hips	jiggle-on-toes	jumping-jacks	chicken-wings
N-ICP	3.70 / 8.54	1.73 / 12.21	10.16 / 14.18	4.74 / 14.29
AMM	0.97 / 2.18	0.93 / 2.60	4.50 / 4.51	2.49 / 4.32
BCPD	73.11 / 34.31	77.24 / 22.11	84.56 / 34.27	74.53 / 46.46
BCPD++	57.98 / 3.63	77.23 / 2.86	80.28 / 3.77	51.77 / 5.84
GBCPD	59.75 / 13.71	79.39 / 12.32	83.90 / 13.66	55.22 / 14.10
GBCPD++	53.08 / <u>1.45</u>	44.95 / <u>1.45</u>	35.56 / <u>1.38</u>	36.85 / <u>1.62</u>
Ours	0.55 / 0.53	0.92 / 0.58	3.51 / 0.64	2.11 / 0.63

visualizations of two cases in Fig. 9.

The face sequence from [27] is represented as depth maps and captures facial expressions and muscle movements, which were acquired using a single depth sensor. We utilized the provided template model from [27] as the template model S_0 , and deformed it to align with the continuous 299 point clouds, which were obtained by converting depth maps.

The DeepDeform dataset [60] is a real RGB-D video dataset, including various scenes such as humans, clothes, animals, etc. Since it does not provide well-defined template meshes, we adopted the reconstructed mesh from [57] as our template, first adjusting it to match the initial frame via a rigid transformation, and then deforming it to align with the subsequent 29 point clouds converted from the depth maps. We notice that the BCPD-class methods are too ineffective, so we omit their

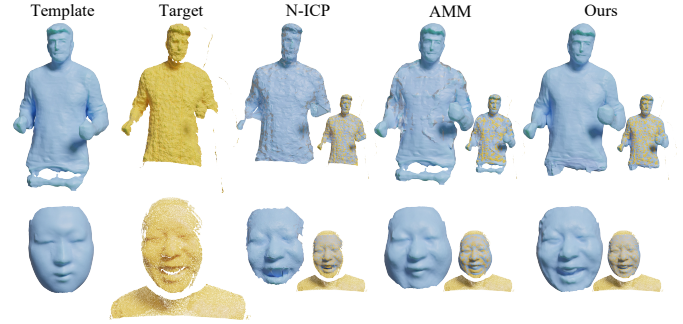


Fig. 10. The results obtained from different methods on two problem instances from the DeepDeform dataset [60](top: 14th frame) and the face sequence from [27](bottom: 100th frame). For each method, we show the deformed mesh (left) and alignment result (right).

results. We show the 2 cases of registration results in Fig. 10 and we can observe that the proposed method is significantly better than other methods. It shows that our method can perform non-rigid registration more stably and reliably. To more intuitively show the performance, we also render the frame-by-frame rendering results of different methods for these three datasets in the *Supplementary Video*.

C. Effectiveness of Components

To measure the effectiveness of different components in our method, we conducted experiments by either removing a specific term or replacing it with an existing method. We compared these variants on the AMA dataset [77] with small differences (“handstand” and “march1” sequences in Tab. I)

TABLE VI
MEAN VALUE OF $\text{CORR}_{\text{ERR}} (\times 0.001) \downarrow$ AND $\text{AUC} \uparrow$ USING DIFFERENT VARIANTS OF OUR METHODS ON THE AMA DATASET [77] AND THE SHREC'20 TRACK NON-RIGID CORRESPONDENCE DATASET [78].

Variants	$E_{\text{align}}^C / E_{\text{align}}$	Robust Weights	handstand	march1	SHREC'20 [78]	Partial data from [77]
With E_{PP}	E_{PP}	✓	2.91 / 0.77	1.28 / 0.89	2.81 / 0.74	2.42 / 0.77
With E_{PPL}	E_{PPL}	✓	0.56 / 0.95	0.09 / 0.99	1.76 / 0.83	0.84 / 0.93
Welsch	✓	Welsch	0.44 / 0.96	0.10 / 0.99	1.53 / 0.85	0.58 / 0.95
Huber	✓	Huber	0.41 / 0.96	0.19 / 0.99	1.88 / 0.84	0.78 / 0.93
Geman McClure	✓	Geman McClure	0.43 / 0.96	0.21 / 0.99	1.90 / 0.84	1.09 / 0.91
Without Weight	✓		0.75 / 0.95	0.39 / 0.99	1.89 / 0.84	1.35 / 0.89
Hard Thres.	✓	Eq. (30)	0.63 / 0.95	0.25 / 0.99	1.93 / 0.83	1.34 / 0.89
Ours	✓	✓	0.38 / 0.96	0.03 / 1.00	1.47 / 0.86	0.46 / 0.96

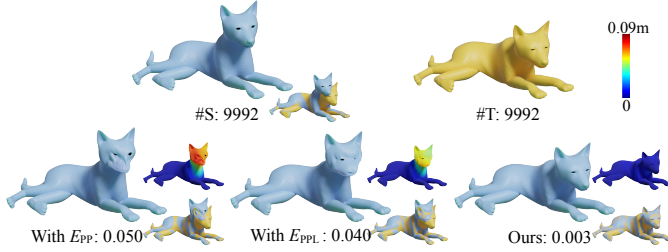


Fig. 11. Comparisons of our method and the variants with point-to-point distance and point-to-plane distance on a problem instance of “doggieMN5_Sleep” from [76]. For each variant, we show the deformed mesh (left), alignment result (right-bottom), and an error map (right-top) that visualizes the distance between the ground-truth correspondences, as well as label the RMSE.

and partial overlaps (“crane” sequence in Tab. II). For the “crane” sequence, we present the average values of the accuracy on all examples. We also performed comparisons on the SHREC'20 non-rigid correspondence dataset [78] and the DT4D dataset [76]. We show all numerical comparisons in Tab. VI and visualized results in Figs. 11, 12 and 13.

Effectiveness of symmetrized point-to-plane distance. We replaced the symmetrized point-to-plane distance metric E_{SPPL} in our target function by the point-to-point distance metric E_{PP} (With E_{PP}) and the point-to-plane distance metric E_{PPL} (With E_{PPL}) respectively while keeping other components and the reliable weights $\alpha_i^{(k)}$ the same. From Tab. VI and Fig. 11, we observe that the symmetrized point-to-plane distance achieves better results than the variants based on E_{PP} and E_{PPL} . We also show a test case on the DT4D dataset [76], which has more obvious visual differences due to its relatively large deformation.

Effectiveness of the robust weights. Moreover, to test the effectiveness of the adaptive robust weights, we compared with the following strategies:

- *Without Weight*: setting $\alpha_i^{(k)} = 1$ for all k and i (i.e., no weight);
- *Hard Thres.*: setting

$$\alpha_i^{(k+1)} = \begin{cases} 0, & \text{if } \|\hat{\mathbf{v}}_i^{(k)} - \mathbf{u}_{\rho_i^{(k+1)}}\| > 3\sigma \\ & \text{or } \hat{\mathbf{n}}_i^{(k)} \cdot \mathbf{n}_{\rho_i^{(k+1)}}^t < 0, \\ 1, & \text{Otherwise.} \end{cases} \quad (30)$$

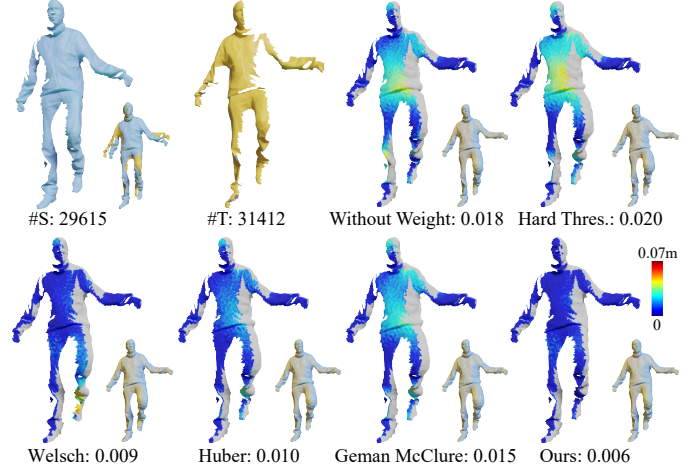


Fig. 12. Comparisons of our method and the variants with different weights for the alignment on a problem instance of partial overlapping data from “crane” sequence [77]. For each variant, we show the deformed mesh with an error map that visualizes the distance between the ground-truth correspondences (left) and alignment result (right-bottom), as well as label the RMSE.

i.e., a hard thresholding of the weight based on the distance between the corresponding points.

We also evaluated a common strategy for enhancing robustness, which involves using established robust functions to measure the alignment error. Specifically, we replace E_{align}^i with

$$\tilde{E}_{\text{align}} = \beta_i \phi([\hat{\mathbf{n}}_i + \mathbf{n}_{\rho_i}^t] \cdot (\hat{\mathbf{v}}_i - \mathbf{u}_{\rho_i})),$$

where β_i is a binary weight designed to filter out point pairs based on normal information; β_i is set to 0 if $\hat{\mathbf{n}}_i \cdot \mathbf{n}_{\rho_i}^t < 0$ and 1 otherwise. $\phi(\cdot)$ is a robust function, and we set it as Welsch loss, Huber loss and Geman-McClure loss respectively to conduct the experiments. These new optimization problems were solved using iteratively reweighted least squares (IRLS) algorithms, drawing on methodologies from [4] [83] [84]. From Tab. VI and Fig. 12, we can observe that our method with robust weights achieves the highest accuracy.

Efficiency of solving algorithms. To validate the effectiveness of our proposed solution for updating $\{\mathbf{R}_i\}$ in the optimization problem (12), we compared it with gradient-based solvers that directly optimize the problem in (12) by the iterative gradient-based optimization method: the steepest descent method and the L-BFGS algorithm. Since $\{\mathbf{R}_i\}$ needs to be in the rotation

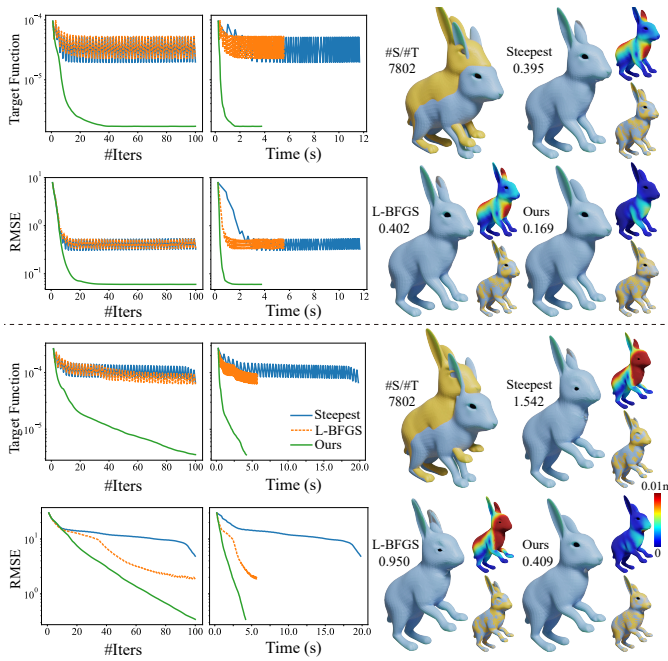


Fig. 13. Comparisons of our method and the variants of updating rotations with the steepest descent method or L-BFGS solver on two problem instances from “bunnyQ_Attack1” sequence of [76]. For each instance, we show the curves of the objective function E and RMSE changing with the number of iterations or time on the left. On the right, for each variant, we show the deformed mesh (left), alignment result (right-bottom), and an error map (right-top) that visualizes the distance between the ground-truth correspondences, as well as label the RMSE ($\times 0.01$).

space, we transformed them in Lie algebra space $so(3)$ and solved them accordingly. Specifically, we replaced $\mathbf{R}_i \in \mathcal{R}$ with $\mathbf{r}_i = [r_1, r_2, r_3]^T \in so(3)$ via Rodrigues’ rotation formula. By utilizing the transformed variables \mathbf{r}_i as independent variables and removing the equality constraint in problem (12), we can convert it into an unconstrained optimization problem. This allows us to employ optimization algorithms such as the steepest descent method or the L-BFGS algorithm to solve it. To ensure optimal convergence of the algorithm, we set a maximum limit of 10^3 iterations for solving each sub-problem of \mathbf{r}_i . Additionally, the algorithm is terminated prematurely if the difference in objective function values between two consecutive iterations falls below a threshold of 10^{-10} . To eliminate potential discrepancies caused by adaptive weights, we set $\alpha_i^{(k)} = 1$ for any k and i in this part. Fig. 13 displays the curves depicting the changes in the objective function values $E = E_{\text{align}} + w_{\text{ARAP}}E_{\text{ARAP}}$ in Eq. (10) and RMSE over the iterations or time. Due to the high degree of nonlinearity of the problem, gradient-based optimization methods are susceptible to getting trapped in local optima. It is important to note that the curve may exhibit fluctuations since using the closest point as the corresponding point does not necessarily result in a reduction of the symmetrized point-to-plane distance. From Fig. 13, it is evident that our proposed method achieves stable and rapid convergence toward the optimal solution.

V. LIMITATIONS AND FUTURE WORK

Although our method has achieved good performance in many cases, it is important to note that our method primarily

focuses on local non-rigid registration. By relying on the nearest point in Euclidean space for establishing correspondences, our method is limited by the initial spatial position. As a result, when significant differences exist between the source and target shapes, the incorporation of global information becomes crucial. This can be achieved through the integration of semantic clues, overall shape analysis, other global features, and initial correspondences predicted by learning-based methods. By incorporating global information, our method can potentially enhance its performance, robustness, and ability to handle larger deformations. Exploring the integration of global information represents a valuable direction for future improvements to our approach. Additionally, our method does not currently account for changes in topology during deformation. For example, when a surface deforms and splits into two or more disconnected parts (as shown in Fig. 10), our current approach is unable to handle such scenarios. A valuable direction for future work will be to develop techniques that can dynamically adjust the deformation graph and local connection relationships in response to these topological changes.

VI. CONCLUSION

We developed a novel optimization-based method for non-rigid surface registration, which offers several key advantages over existing approaches. Our method leveraged a symmetrized point-to-plane distance metric, resulting in a more precise alignment of geometric surfaces. By incorporating adaptive robust weights, our method effectively handles data with defects such as noise, outliers, or partial overlaps. To address the complexity of the objective function, we employed an alternating optimization scheme and designed a surrogate function that is easy to solve. Additionally, we introduced a graph-based deformation technique for coarse alignment to improve both the accuracy and solution speed. Experimental results demonstrated the superiority of our proposed method compared to state-of-the-art techniques in most examples. Our method achieved the best performance while maintaining a relatively fast solution speed.

REFERENCES

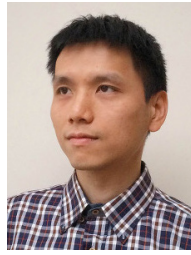
- [1] P. Besl and N. D. McKay, “A method for registration of 3-D shapes,” *IEEE TPAMI*, vol. 14, no. 2, pp. 239–256, 1992. 1, 3
- [2] B. Amberg, S. Romdhani, and T. Vetter, “Optimal step nonrigid ICP algorithms for surface registration,” in *CVPR*, 2007. 1, 2, 3, 7, 8, 10, 11, 12
- [3] K. Li, J. Yang, Y.-K. Lai, and D. Guo, “Robust non-rigid registration with reweighted position and transformation sparsity,” *IEEE TVCG*, vol. 25, no. 6, pp. 2255–2269, 2019. 1, 2
- [4] Y. Yao, B. Deng, W. Xu, and J. Zhang, “Fast and robust non-rigid registration using accelerated majorization-minimization,” *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 45, no. 08, pp. 9681–9698, 2023. 1, 6, 7, 8, 10, 11, 12, 15
- [5] B. Allen, B. Curless, and Z. Popović, “The space of human body shapes: reconstruction and parameterization from range scans,” *ACM TOG*, vol. 22, no. 3, pp. 587–594, 2003. 1
- [6] H. Li, R. W. Sumner, and M. Pauly, “Global correspondence optimization for non-rigid registration of depth scans,” *Comput. Graph. Forum*, vol. 27, pp. 1421–1430, 2008. 1, 2
- [7] W. Chang and M. Zwicker, “Global registration of dynamic range scans for articulated model reconstruction,” *ACM TOG*, vol. 30, no. 3, pp. 26:1–26:15, 2011. 1, 2
- [8] Y. Yao, B. Deng, W. Xu, and J. Zhang, “Quasi-newton solver for robust non-rigid registration,” in *CVPR*, 2020, pp. 7597–7606. 1, 2, 7

- [9] H. Li, B. Adams, L. J. Guibas, and M. Pauly, "Robust single-view geometry and motion reconstruction," *ACM TOG*, vol. 28, no. 5, pp. 175:1–175:10, 2009. 1, 2
- [10] L. Xu, Y. Liu, W. Cheng, K. Guo, G. Zhou, Q. Dai, and L. Fang, "FlyCap: Markerless motion capture using multiple autonomous flying cameras," *IEEE TVCG*, vol. 24, no. 8, pp. 2284–2297, 2018. 1
- [11] M. Dou, S. Khamis, Y. Degtyarev, P. Davidson, S. R. Fanello, A. Kowdle, S. O. Escolano, C. Rhemann, D. Kim, J. Taylor, P. Kohli, V. Tankovich, and S. Izadi, "Fusion4D: Real-time performance capture of challenging scenes," *ACM TOG*, vol. 35, no. 4, pp. 114:1–114:13, 2016. 1, 2
- [12] T. Yu, Z. Zheng, K. Guo, J. Zhao, Q. Dai, H. Li, G. Pons-Moll, and Y. Liu, "DoubleFusion: Real-time capture of human performances with inner body shapes from a single depth sensor," in *CVPR*, 2018, pp. 7287–7296. 1, 2
- [13] Z. Li, T. Yu, C. Pan, Z. Zheng, and Y. Liu, "Robust 3D self-portraits in seconds," in *CVPR*, 2020, pp. 1344–1353. 1, 2
- [14] N. J. Mitra, N. Gelfand, H. Pottmann, and L. Guibas, "Registration of point cloud data from a geometric optimization perspective," in *Proceedings of the 2004 Eurographics/ACM SIGGRAPH Symposium on Geometry Processing*, 2004, p. 22–31. 1, 4
- [15] H. Pottmann, Q.-X. Huang, Y.-L. Yang, and S.-M. Hu, "Geometry and convergence analysis of algorithms for registration of 3D shapes," *IJCV*, vol. 67, no. 3, pp. 277–296, 2006. 1, 3, 4
- [16] S. Rusinkiewicz, "A symmetric objective function for ICP," *ACM TOG*, vol. 38, no. 4, jul 2019. 1, 2, 4, 5
- [17] K. Lange, *MM Optimization Algorithms*. SIAM, 2016. 2, 5
- [18] R. W. Sumner, J. Schmid, and M. Pauly, "Embedded deformation for shape manipulation," *TOG*, vol. 26, no. 3, pp. 80:1–80:7, 2007. 2, 6
- [19] Y. Sahillioglu, "Recent advances in shape correspondence," *The Vis. Comput.*, vol. 36, no. 8, p. 1705–1721, 2020. 2
- [20] B. Deng, Y. Yao, R. M. Dyke, and J. Zhang, "A survey of non-rigid 3D registration," *Comput. Graph. Forum*, vol. 41, no. 2, pp. 559–589, 2022. 2
- [21] M. Liao, Q. Zhang, H. Wang, R. Yang, and M. Gong, "Modeling deformable objects from a single depth camera," in *ICCV*, 2009, pp. 167–174. 2
- [22] H. Hontani, T. Matsuno, and Y. Sawada, "Robust nonrigid ICP using outlier-sparsity regularization," in *CVPR*, 2012, pp. 174–181. 2
- [23] T. Yu, K. Guo, F. Xu, Y. Dong, Z. Su, J. Zhao, J. Li, Q. Dai, and Y. Liu, "BodyFusion: Real-time capture of human motion and surface geometry using a single depth camera," in *ICCV*, 2017, pp. 910–919. 2
- [24] M. Dou, P. Davidson, S. R. Fanello, S. Khamis, A. Kowdle, C. Rhemann, V. Tankovich, and S. Izadi, "Motion2Fusion: Real-time volumetric performance capture," *ACM TOG*, vol. 36, no. 6, pp. 246:1–246:16, 2017. 2
- [25] C. Li, Z. Zhao, and X. Guo, "ArticulatedFusion: Real-time reconstruction of motion, geometry and segmentation using a single depth camera," in *ECCV*, 2018, pp. 317–332. 2
- [26] H. Zhang, Z.-H. Bo, J.-H. Yong, and F. Xu, "InteractionFusion: Real-time reconstruction of hand poses and deformable objects in hand-object interactions," *ACM TOG*, vol. 38, no. 4, pp. 48:1–48:11, 2019. 2
- [27] K. Guo, F. Xu, Y. Wang, Y. Liu, and Q. Dai, "Robust non-rigid motion tracking and surface reconstruction using L0 regularization," in *ICCV*, 2015, pp. 3083–3091. 2, 7, 13, 14
- [28] D. Thomas and R.-I. Taniguchi, "Augmented blendshapes for real-time simultaneous 3D head modeling and facial motion capture," in *CVPR*, 2016, pp. 3299–3308. 2
- [29] C. Li and X. Guo, "Topology-change-aware volumetric fusion for dynamic scene reconstruction," in *ECCV*, 2020, pp. 258–274. 2
- [30] Z. Su, L. Xu, Z. Zheng, T. Yu, Y. Liu, and L. Fang, "RobustFusion: Human volumetric capture with data-driven visual cues using a RGBD camera," in *ECCV*, 2020, pp. 246–264. 2
- [31] K. Zampogiannis, C. Fermüller, and Y. Aloimonos, "Topology-aware non-rigid point cloud registration," *IEEE TPAMI*, vol. 43, no. 3, pp. 1056–1069, 2021. 2
- [32] A. Myronenko and X. Song, "Point set registration: Coherent point drift," *IEEE TPAMI*, vol. 32, no. 12, pp. 2262–2275, 2010. 2
- [33] B. Jian and B. C. Vemuri, "Robust point set registration using gaussian mixture models," *IEEE TPAMI*, vol. 33, no. 8, pp. 1633–1645, 2011. 2
- [34] J. Ma, J. Zhao, J. Tian, Z. Tu, and A. L. Yuille, "Robust estimation of nonrigid transformation for point set registration," in *CVPR*, 2013, pp. 2147–2154. 2
- [35] S. Ge, G. Fan, and M. Ding, "Non-rigid point set registration with global-local topology preservation," in *CVPR*, 2014, pp. 245–251. 2
- [36] O. Hirose, "A Bayesian formulation of coherent point drift," *IEEE TPAMI*, vol. 43, no. 7, pp. 2269–2286, 2021. 2, 7, 8, 10, 11, 12
- [37] —, "Acceleration of non-rigid point set registration with downsampling and Gaussian process regression," *IEEE TPAMI*, vol. 43, no. 8, pp. 2858–2865, 2020. 2, 7, 8, 10, 11, 12
- [38] —, "Geodesic-based bayesian coherent point drift," *IEEE TPAMI*, pp. 1–18, 2022. 2, 7, 8, 10, 11, 12
- [39] A. Fan, J. Ma, X. Tian, X. Mei, and W. Liu, "Coherent point drift revisited for non-rigid shape matching and registration," in *CVPR*, 2022, pp. 1424–1434. 2
- [40] R. A. Newcombe, D. Fox, and S. M. Seitz, "DynamicFusion: Reconstruction and tracking of non-rigid scenes in real-time," in *CVPR*, 2015, pp. 343–352. 2
- [41] L. Xu, Z. Su, L. Han, T. Yu, Y. Liu, and L. Fang, "UnstructuredFusion: realtime 4D geometry and texture reconstruction using commercial RGBD cameras," *IEEE TPAMI*, vol. 42, no. 10, pp. 2508–2522, 2020. 2
- [42] S. Ren, J. Hou, X. Chen, H. Xiong, and W. Wang, "Ddm: A metric for comparing 3d shapes using directional distance fields," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–16, 2025. 2
- [43] P. Huang, C. Budd, and A. Hilton, "Global temporal registration of multiple non-rigid surface sequences," in *CVPR*. IEEE, 2011, pp. 3473–3480. 2
- [44] Z. Li, Y. Ji, W. Yang, J. Ye, and J. Yu, "Robust 3D human motion reconstruction via dynamic template construction," in *International Conference on 3D Vision*. IEEE, 2017, pp. 496–505. 2
- [45] F. Bogo, J. Romero, M. Loper, and M. J. Black, "Faust: Dataset and evaluation for 3d mesh registration," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 3794–3801. 2
- [46] F. Bogo, J. Romero, G. Pons-Moll, and M. J. Black, "Dynamic FAUST: Registering human bodies in motion," in *CVPR*, Jul. 2017. 2, 7, 12, 13, 14
- [47] M. Eisenberger, D. Novotny, G. Kerchenbaum, P. Labatut, N. Neverova, D. Cremers, and A. Vedaldi, "Neuromorph: Unsupervised shape interpolation and correspondence in one go," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 7473–7483. 2
- [48] S. Prokudin, Q. Ma, M. Raafat, J. Valentin, and S. Tang, "Dynamic point fields," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 7964–7976. 2
- [49] R. S. Sundararaman, R. Marin, E. Rodola, and M. Ovsjanikov, "Reduced representation of deformation fields for effective non-rigid shape matching," *Advances in Neural Information Processing Systems*, vol. 35, pp. 10405–10420, 2022. 2
- [50] G. Yang, M. Vo, N. Neverova, D. Ramanan, A. Vedaldi, and H. Joo, "BANMo: Building animatable 3d neural models from many casual videos," in *CVPR*, 2022. 2
- [51] Y. Yao, S. Ren, J. Hou, Z. Deng, J. Zhang, and W. Wang, "Dynosurf: Neural deformation-based temporally consistent dynamic surface reconstruction," in *European Conference on Computer Vision*. Springer, 2024, pp. 271–288. 2
- [52] E. Corona, G. Pons-Moll, G. Alenya, and F. Moreno-Noguer, "Learned vertex descent: A new direction for 3d human model fitting," in *European Conference on Computer Vision*. Springer, 2022, pp. 146–165. 2
- [53] R. Marin, E. Corona, and G. Pons-Moll, "Nicip: neural icp for 3d human registration at scale," in *European Conference on Computer Vision*. Springer, 2024, pp. 265–285. 2, 3
- [54] S. Prokudin, C. Lassner, and J. Romero, "Efficient learning on point clouds with basis point sets," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 4332–4341. 2
- [55] W. Feng, J. Zhang, H. Cai, H. Xu, J. Hou, and H. Bao, "Recurrent multi-view alignment network for unsupervised surface registration," in *CVPR*, 2021, pp. 10297–10307. 3
- [56] J. Lei and K. Daniilidis, "Cadex: Learning canonical deformation coordinate space for dynamic surface representation via neural homeomorphism," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 6624–6634. 3
- [57] H. Cai, W. Feng, X. Feng, Y. Wang, and J. Zhang, "Neural surface reconstruction of dynamic scenes with monocular rgb-d camera," in *NeurIPS*, 2022. 3, 14
- [58] T. Groueix, M. Fisher, V. G. Kim, B. C. Russell, and M. Aubry, "3d-coded: 3d correspondences by deep deformation," in *Proceedings of the european conference on computer vision (ECCV)*, 2018, pp. 230–246. 3
- [59] L. Wei, Q. Huang, D. Ceylan, E. Vouga, and H. Li, "Dense human body correspondences using convolutional networks," in *CVPR*, 2016, pp. 1544–1553. 3
- [60] A. Bozic, M. Zollhöfer, C. Theobalt, and M. Nießner, "DeepDeform: Learning non-rigid RGB-D reconstruction with semi-supervised data," in *CVPR*, 2020, pp. 7002–7012. 3, 7, 13, 14

- [61] Y. Li and T. Harada, "Lepard: Learning partial point cloud matching in rigid and deformable scenes," *CVPR*, 2022. [3](#)
- [62] G. Trappolini, L. Cosmo, L. Moschella, R. Marin, S. Melzi, and E. Rodolà, "Shape registration in the time of transformers," *NeurIPS*, 2021. [3](#)
- [63] Z. Qin, H. Yu, C. Wang, Y. Peng, and K. Xu, "Deep graph-based spatial consistency for robust non-rigid point cloud registration," in *CVPR*, June 2023, pp. 5394–5403. [3](#), [7](#), [8](#), [10](#), [11](#), [12](#)
- [64] B. L. Bhatnagar, C. Sminchisescu, C. Theobalt, and G. Pons-Moll, "Loopreg: Self-supervised learning of implicit surface correspondences, pose and shape for 3d human mesh registration," *Advances in Neural Information Processing Systems*, vol. 33, pp. 12 909–12 922, 2020. [3](#)
- [65] Y. Wang and J. M. Solomon, "Deep closest point: Learning representations for point cloud registration," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 3523–3532. [3](#)
- [66] K. Wang, J. Xie, G. Zhang, L. Liu, and J. Yang, "Sequential 3D human pose and shape estimation from point clouds," in *CVPR*, 2020, pp. 7275–7284. [3](#)
- [67] A. Božić, P. Palafox, M. Zollhöfer, A. Dai, J. Thies, and M. Nießner, "Neural non-rigid tracking," in *NeurIPS*, H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, Eds., vol. 33. Curran Associates, Inc., 2020, pp. 18 727–18 737. [3](#)
- [68] Y. Zeng, Y. Qian, Z. Zhu, J. Hou, H. Yuan, and Y. He, "CorrNet3D: Unsupervised end-to-end learning of dense correspondence for 3D point clouds," in *CVPR*, 2021, pp. 6052–6061. [3](#)
- [69] W. Wang, D. Ceylan, R. Mech, and U. Neumann, "3DN: 3D deformation network," in *CVPR*, 2019, pp. 1038–1046. [3](#)
- [70] Y. Chen and G. Medioni, "Object modelling by registration of multiple range images," *Image and Vision Computing*, vol. 10, no. 3, pp. 145–155, 1992. [3](#)
- [71] S. Rusinkiewicz and M. Levoy, "Efficient variants of the icp algorithm," in *Proceedings Third International Conference on 3-D Digital Imaging and Modeling*, 2001, pp. 145–152. [4](#)
- [72] O. Sorkine and M. Alexa, "As-rigid-as-possible surface modeling," in *Proceedings of the Fifth Eurographics Symposium on Geometry Processing*, ser. SGP '07, 2007, p. 109–116. [4](#)
- [73] O. Sorkine-Hornung and M. Rabinovich, "Least-squares rigid motion using svd," *Computing*, vol. 1, no. 1, pp. 1–5, 2017. [6](#)
- [74] Y. Li and T. Harada, "Non-rigid point cloud registration with neural deformation pyramid," *NeurIPS*, 2022. [7](#), [8](#), [10](#), [11](#), [12](#)
- [75] J. Huang, T. Birdal, Z. Gojcic, L. J. Guibas, and S. Hu, "Multiway non-rigid point cloud registration via learned functional map synchronization," *IEEE TPAMI*, vol. 45, no. 2, pp. 2038–2053, 2023. [7](#), [8](#), [10](#), [11](#), [12](#)
- [76] Y. Li, H. Takehara, T. Taketomi, B. Zheng, and M. Nießner, "4dcomplete: Non-rigid motion estimation beyond the observable surface," *ICCV*, 2021. [7](#), [12](#), [15](#), [16](#)
- [77] D. Vlasic, I. Baran, W. Matusik, and J. Popović, "Articulated mesh animation from multi-view silhouettes," *ACM TOG*, vol. 27, no. 3, pp. 97:1–97:9, Aug. 2008. [7](#), [8](#), [9](#), [10](#), [14](#), [15](#)
- [78] R. M. Dyke, F. Zhou, Y.-K. Lai, P. L. Rosin, D. Guo, K. Li, R. Marin, and J. Yang, "SHREC 2020 Track: Non-rigid shape correspondence of physically-based deformations," in *Eurographics Workshop on 3D Object Retrieval*, T. Schreck, T. Theoharis, I. Pratikakis, M. Spagnuolo, and R. C. Veltkamp, Eds. The Eurographics Association, 2020, pp. 19–26. [7](#), [10](#), [11](#), [15](#)
- [79] B. L. Bhatnagar, X. Xie, I. Petrov, C. Sminchisescu, C. Theobalt, and G. Pons-Moll, "Behave: Dataset and method for tracking human object interactions," in *CVPR*. IEEE, jun 2022. [7](#), [11](#), [12](#)
- [80] O. Litany, T. Remez, E. Rodola, A. Bronstein, and M. Bronstein, "Deep functional maps: Structured prediction for dense shape correspondence," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 5659–5667. [7](#)
- [81] D. Cao, P. Roetzer, and F. Bernard, "Unsupervised learning of robust spectral shape matching," *ACM Trans. Graph.*, vol. 42, no. 4, Jul. 2023. [Online]. Available: <https://doi.org/10.1145/3592107> [7](#)
- [82] M. Loper, N. Mahmood, J. Romero, G. Pons-Moll, and M. J. Black, "SMPL: A skinned multi-person linear model," *ACM TOG*, vol. 34, no. 6, pp. 248:1–248:16, 2015. [12](#), [13](#)
- [83] P. Bergström and O. Edlund, "Robust registration of point sets using iteratively reweighted least squares," *Computational optimization and applications*, vol. 58, pp. 543–561, 2014. [15](#)
- [84] M. Adlerstein, J. C. V. Soares, A. Bratta, and C. Semini, "Sandro: a robust solver with a splitting strategy for point cloud registration," 2025. [15](#)



Yuxin Yao is currently a post-doctoral researcher with the Department of Computer Science, City University of Hong Kong. She received the BE degree from the University of Electronic Science and Technology of China, Chengdu, China, in 2018, and the PhD degree from the University of Science and Technology of China, Hefei, China, in 2023. Her research interests include computer vision, computer graphics, and 3D reconstruction.



Applications.

Bailin Deng is a Senior Lecturer in the School of Computer Science and Informatics at Cardiff University. He received the BEng degree in computer software (2005) and the MSc degree in computer science (2008) from Tsinghua University (China), and the PhD degree in technical mathematics (2011) from Vienna University of Technology (Austria). His research interests include geometry processing, numerical optimization, computational design, and digital fabrication. He is a member of IEEE, and an associate editor of IEEE Computer Graphics and



visual computing.

Junhui Hou is an Associate Professor with the Department of Computer Science, City University of Hong Kong. He holds a B.Eng. degree in information engineering (Talented Students Program) from the South China University of Technology, Guangzhou, China (2009), an M.Eng. degree in signal and information processing from Northwestern Polytechnical University, Xi'an, China (2012), and a Ph.D. degree from the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore (2016). His research interests are multi-dimensional

Dr. Hou received the Early Career Award (3/381) from the Hong Kong Research Grants Council in 2018 and the NSFC Excellent Young Scientists Fund in 2024. He has served or is serving as an Associate Editor for *IEEE Transactions on Visualization and Computer Graphics*, *IEEE Transactions on Image Processing*, *IEEE Transactions on Multimedia*, and *IEEE Transactions on Circuits and Systems for Video Technology*.



Juyong Zhang is a professor in the School of Mathematical Sciences at University of Science and Technology of China. He received the BS degree from the University of Science and Technology of China in 2006, and the PhD degree from Nanyang Technological University, Singapore. His research interests include computer graphics, computer vision, and numerical optimization. He is an associate editor of IEEE Transactions on Mobile Computing and IEEE Computer Graphics and Applications.