

# Quasi-Newton Solver for Robust Non-Rigid Registration

Yuxin Yao<sup>1</sup> Bailin Deng<sup>2</sup> Weiwei Xu<sup>3</sup> Juyong Zhang<sup>1\*</sup>

<sup>1</sup>University of Science and Technology of China <sup>2</sup>Cardiff University <sup>3</sup>Zhejiang University

yaoyuxin@mail.ustc.edu.cn DengB3@cardiff.ac.uk xww@cad.zju.edu.cn juyong@ustc.edu.cn

## Abstract

*Imperfect data (noise, outliers and partial overlap) and high degrees of freedom make non-rigid registration a classical challenging problem in computer vision. Existing methods typically adopt the  $\ell_p$  type robust estimator to regularize the fitting and smoothness, and the proximal operator is used to solve the resulting non-smooth problem. However, the slow convergence of these algorithms limits its wide applications. In this paper, we propose a formulation for robust non-rigid registration based on a globally smooth robust estimator for data fitting and regularization, which can handle outliers and partial overlaps. We apply the majorization-minimization algorithm to the problem, which reduces each iteration to solving a simple least-squares problem with L-BFGS. Extensive experiments demonstrate the effectiveness of our method for non-rigid alignment between two shapes with outliers and partial overlap, with quantitative evaluation showing that it outperforms state-of-the-art methods in terms of registration accuracy and computational speed. The source code is available at [https://github.com/Juyong/Fast\\_RNRR](https://github.com/Juyong/Fast_RNRR).*

## 1. Introduction

With the popularity of depth acquisition devices such as Kinect, PrimeSense and the depth sensors on smartphones, techniques for 3D object tracking and reconstruction from point clouds have enabled various applications. Non-rigid registration is a fundamental problem for such techniques, especially for the reconstruction of dynamic objects. Since depth maps obtained from structured light or time-of-flight cameras often contain outliers and holes, a robust non-rigid registration algorithm is needed to handle such data. Moreover, real-time applications require high computational efficiency for non-rigid registration.

Given two point clouds sampled from a source surface and a target surface respectively, the aim of non-rigid registration is to find a transformation field for the source point

cloud to align it with the target point cloud. This problem is typically solved via optimization. The objective energy often includes alignment terms that measure the deviation between the two point clouds after the transformation, as well as regularization terms that enforce smoothness of the transformation field. Existing methods often formulate these terms using the  $\ell_2$ -norm, which penalizes alignment and smoothness errors across the whole surface [1, 21, 20]. On the other hand, the ground-truth alignment can potentially induce large errors for these terms in some localized regions of the point clouds, due to noises, outliers, partial overlaps, or articulated motions between the point clouds. Such large localized errors will be inhibited by the  $\ell_2$  formulations, which can lead to erroneous alignment results. To improve the alignment accuracy, recent works have utilized sparsity-promoting norms for these terms, such as the  $\ell_1$ -norm [36, 22, 17] and the  $\ell_0$ -norm [12]. The sparsity optimization enforces small error metrics on most parts of the point cloud while allowing for large errors in some local regions, improving the robustness of the registration process. However, these sparsity terms can lead to non-smooth problems that are more challenging to solve. Existing methods often employ first-order solvers such as the alternating direction method of multipliers (ADMM), which can suffer from slow convergence to high-accuracy solutions [6].

In this paper, we propose a new approach for robust non-rigid registration with fast convergence. The key idea is to enforce sparsity using the Welsch's function [14], which has been utilized for robust processing of images [13] and meshes [37]. We formulate an optimization that applies the Welsch's function to both the alignment error and the regularization, to achieve robust registration. Unlike the  $\ell_p$ -norms, the Welsch's function is smooth and does not induce non-smooth optimization. We solve the optimization problem using the majorization-minimization (MM) algorithm [19]. It iteratively constructs a surrogate function for the target energy based on the current variable values and minimizes the surrogate function to update the variables, and is guaranteed to converge to a local minimum. The Welsch's function enables us to derive a surrogate function in a simple least-squares form, which can be solved efficiently using L-BFGS.

\*Corresponding author

Experimental results verify the robustness of our method as well as its superior performance compared to existing robust registration approaches.

In summary, the main contributions of this paper include:

- We formulate an optimization problem for non-rigid registration, using the Welsch’s function to induce sparsity for alignment error and transformation smoothness. The proposed formulation effectively improves the robustness and accuracy of the results.
- We propose an MM algorithm to solve the optimization problem, using L-BFGS to tackle the sub-problems. The combination of MM and L-BFGS greatly improves the computational efficiency of robust non-rigid registration compared to existing approaches.

## 2. Related work

Non-rigid registration has been widely studied in computer vision and image processing. The reader is referred to [31] for a recent survey on rigid and non-rigid registration of 3D point clouds and meshes. In the following, we focus on works that are closely related to our method.

Various optimization-based methods have been proposed for non-rigid registration. Chui et al. [8] utilized a Thin Plate Spline (TPS) model to represent non-rigid mappings, and alternately update the correspondence and TPS parameters to find an optimized alignment. Following this approach, Brown et al. [7] used a weighted variant of iterative closest point (ICP) to obtain sparse correspondences, and warped scans to global optimal position by TPS. Extending the classical ICP algorithm for rigid registration, Amberg et al. [1] proposed a non-rigid ICP algorithm that gradually decreases the stiffness of regularization and incrementally deforms the source model to the target model. Li et al. [21] adopted an embedded deformation approach [29] to express a non-rigid deformation using transformations defined on a deformation graph, and simultaneously optimized the correspondences between source and target scans, confidence weights for measuring the reliability of correspondence, and a warping field that aligns the source with the target. Later, Li et al. [20] combined point-to-point and point-to-plane metrics for the more accurate measure of correspondence.

Other methods tackle the problem from a statistical perspective. Considering the fitting of two point clouds as a probability density estimation problem, Myronenko et al. [2] proposed the Coherent Point Drift algorithm which encourages displacement vectors to point into similar directions to improve the coherence of the transformation. Hontani et al. [15] incorporated a statistical shape model and a noise model into the non-rigid ICP framework, and detected outliers based on their sparsity. Jian et al. [16] represented each point set as a mixture of Gaussians treated point set registration as a problem of aligning two mixtures. Also

with a statistical framework, Wand et al. [34] used a meshless deformation model to perform the pairwise alignment. Ma et al. [24] proposed an  $L_2E$  estimator to build more reliable sparse and dense correspondences.

Many non-rigid registration algorithms are based on  $\ell_2$ -norm metrics for the deviation between source and target models and the smoothness of transformation fields, which can lead to erroneous results when the ground-truth alignment induces large deviation or non-smooth transformation due to noises, partial overlaps, or articulate motions. To address this issue, various sparsity-based methods have been proposed. Yang et al. [36] utilized the  $\ell_1$ -norm to promote regularity of the transformation. Li et al. [22] additionally introduced position sparsity to improve robustness against noises and outliers. For robust motion tracking and surface reconstruction, Guo et al. [12] proposed an  $\ell_0$ -based motion regularizer to accommodate articulated motions.

Depending on the application, different regularizations for the transformation have been proposed to improve the robustness of the results. In [10], an as-rigid-as-possible energy was introduced to avoid shrinkage and keep local rigidity. Wu et al. [35] introduced an as-conformal-as-possible energy to avoid mesh distortion. Jiang et al. [17] applied a Huber-norm regularization to induce piecewise smooth transformation.

## 3. Problem Formulation

Let  $\mathcal{S} = \{\mathcal{V}, \mathcal{E}\}$  be a source surface consisting of sample points  $\mathcal{V} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n \in \mathbb{R}^3\}$  connected by a set of edges  $\mathcal{E}$ . Let  $\mathcal{T}$  be a target surface with sample points  $\mathcal{U} = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m \in \mathbb{R}^3\}$ . We seek to apply affine transformations to the source points  $\mathcal{V}$  to align them with the target surface. In this paper, we adopt the embedded deformation approach proposed in [29] to model the transformations. Specifically, we construct a *deformation graph*  $\mathcal{G}$  with its vertices  $\mathcal{V}_{\mathcal{G}} = \{\mathbf{p}_1, \dots, \mathbf{p}_r\}$  being a subset of the source points  $\mathcal{V}$ , and with its edges  $\mathcal{E}_{\mathcal{G}}$  connecting vertices that are nearby on the target surface. On each deformation graph vertex  $\mathbf{p}_j$  we define an affine transformation represented with a transformation matrix  $\mathbf{A}_j \in \mathbb{R}^{3 \times 3}$  and a displacement vector  $\mathbf{t}_j \in \mathbb{R}^3$ . Each vertex  $\mathbf{p}_j$  influences a localized region that contains any source point  $\mathbf{v}_i$  whose geodesic distance  $D(\mathbf{v}_i, \mathbf{p}_j)$  to  $\mathbf{p}_j$  on the source surface is smaller than a user-specified radius  $R$ . If  $\mathbf{p}_j$  influences a source point  $\mathbf{v}_i$ , the affine transformation  $(\mathbf{A}_j, \mathbf{t}_j)$  associated with  $\mathbf{p}_j$  induces a transformed position  $\mathbf{A}_j(\mathbf{v}_i - \mathbf{p}_j) + \mathbf{p}_j + \mathbf{t}_j$  for  $\mathbf{v}_i$ . Then the final transformed position  $\hat{\mathbf{v}}_i$  for  $\mathbf{v}_i$  is a convex combination of all positions induced by the vertices in graph  $\mathcal{G}$  [20]:

$$\hat{\mathbf{v}}_i = \sum_{\mathbf{p}_j \in \mathcal{I}(\mathbf{v}_i)} w_{ij} \cdot (\mathbf{A}_j(\mathbf{v}_i - \mathbf{p}_j) + \mathbf{p}_j + \mathbf{t}_j), \quad (1)$$

where  $\mathcal{I}(\mathbf{v}_i) = \{\mathbf{p}_j \mid D(\mathbf{v}_i, \mathbf{p}_j) < R\}$  denotes the set of vertices that influence  $\mathbf{v}_i$ , and  $w_{ij}$  is a distance-dependent

normalized weight:

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

Compared to formulations that attach a different transformation to each source point such as [22], a major benefit of using the deformation graph is that the deformation of the target surface can be defined with a much smaller number of variables, enabling more efficient optimization.

In the following, we first present an optimization formulation to determine the affine transformations associated with the deformation graph, and then explain how to construct the deformation graph. For the ease of presentation, we use  $\mathbf{X}_j$  to denote the affine transformation  $(\mathbf{A}_j, \mathbf{t}_j)$  at vertex  $\mathbf{p}_j$ , whereas  $\mathbf{X}$  denotes the set of all transformations.

### 3.1. Optimization Formulation

For robust alignment between the source and the target surfaces, we determine the affine transformations  $\mathbf{X}$  via the following optimization:

$$\min_{\mathbf{X}} E_{\text{align}}(\mathbf{X}) + \alpha E_{\text{reg}}(\mathbf{X}) + \beta E_{\text{rot}}(\mathbf{X}). \quad (2)$$

Here the terms  $E_{\text{align}}$ ,  $E_{\text{smooth}}$ , and  $E_{\text{rot}}$  measures the alignment error, the regularization of transformations across the surface, and the deviation between the transformation matrices and rotation matrices, respectively.  $\alpha$  and  $\beta$  are positive weights that control the tradeoff between these terms. The definition for each term is explained in the following.

**Alignment Term.** For each transformed source point  $\hat{\mathbf{v}}_i$ , we can find the closest target point  $\mathbf{u}_{\rho(i)} \in \mathcal{U}$ . The alignment term should penalize the deviation between  $\hat{\mathbf{v}}_i$  and  $\mathbf{u}_{\rho(i)}$ . A simple approach is to define it as the sum of squared distances between all such pairs of points. This is indeed the alignment error metric used in the classical ICP algorithm [3] for rigid registration [25]. On the other hand, such  $\ell_2$ -norm of pointwise distance can lead to erroneous alignment on real-world data, where the two surfaces might only overlap partially and their point positions might be noisy. This is because partial overlaps and noisy data can induce large distance from some source points to their corresponding target points under the ground-truth alignment, which would be prohibited by the  $\ell_2$ -norm minimization.

Some previous work, such as the Sparse ICP algorithm from [5], adopts the  $\ell_p$ -norm ( $0 < p < 1$ ) as the error metric. It is less sensitive to noises and partial overlaps, since  $\ell_p$ -norm minimization allows for large distances at some points. However, numerical minimization of the  $\ell_p$ -norm can be much more expensive than the  $\ell_2$ -norm. For example, the problem is solved in [5] with an iterative algorithm that alternately updates the point correspondence and the transformation which is similar with classical ICP. However, its

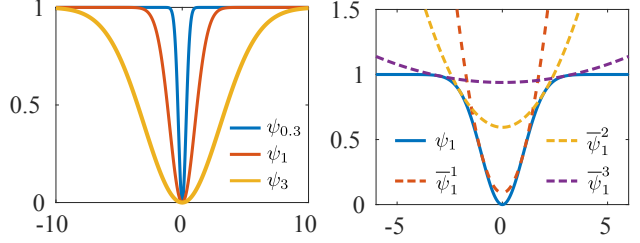


Figure 1. Left: the Welsch's function with different  $\nu$  values. Right: different surrogate functions for the Welsch's function with  $\nu = 1$ .

transformation update has to be done with an inner ADMM solver which is much slower than the closed-form update in classical ICP, and also lack convergence guarantee due to the non-convexity of the problem.

Inspired by the recent work from [13] on robust image filtering and [37] on robust mesh filtering, we adopt the following robust metric for the alignment error:

$$E_{\text{align}}(\mathbf{X}) = \sum_{i=1}^n \psi_{\nu_a}(\|\hat{\mathbf{v}}_i - \mathbf{u}_{\rho(i)}\|), \quad (3)$$

where  $\psi_{\nu_a}(\cdot)$  is the Welsch's function [14] (see Fig. 1 left):

$$\psi_{\nu_a}(x) = 1 - \exp\left(-\frac{x^2}{2\nu_a^2}\right),$$

and  $\nu_a > 0$  is a user-specified parameter.  $\psi_{\nu_a}$  is monotonically increasing on  $[0, +\infty)$ , thus  $\psi_{\nu_a}(\|\hat{\mathbf{v}}_i - \mathbf{u}_{\rho(i)}\|)$  penalizes the deviation between  $\hat{\mathbf{v}}_i$  and  $\mathbf{u}_{\rho(i)}$ . On the other hand, as  $\psi_{\nu_a} \leq 1$ , the deviation only induces a bounded influence on the metric  $E_{\text{align}}$ . Moreover, if  $\nu_a \rightarrow 0$ , then  $E_{\text{align}}$  approaches the  $\ell_0$ -norm of the pointwise distance from the source points to their closest points on the target surface. Therefore, this error metric is insensitive to noisy data and partial overlaps.

**Regularization Term.** Ideally, the transformation induced by two neighboring vertices  $\mathbf{p}_i, \mathbf{p}_j$  of the deformation graph should be consistent on their overlapping influenced regions. In [29], such consistency is measured at  $\mathbf{p}_i$  using the difference of deformation induced by the transformation  $\mathbf{X}_i$  at  $\mathbf{p}_i$  and the transformation  $\mathbf{X}_j$  at  $\mathbf{p}_j$ :

$$\mathbf{D}_{ij} = \mathbf{A}_j(\mathbf{p}_i - \mathbf{p}_j) + \mathbf{p}_j + \mathbf{t}_j - (\mathbf{p}_i + \mathbf{t}_i). \quad (4)$$

Ideally,  $\mathbf{D}_{ij}$  should be small across the deformation graph. On the other hand, in some cases the optimal deformation may induce large values of  $\mathbf{D}_{ij}$  in some regions, such as the joints of a human body. To reduce the magnitudes of  $\mathbf{D}_{ij}$  across the graph while allowing for large magnitudes in some regions, we define the regularization term using the Welsch's function on  $\|\mathbf{D}_{ij}\|$ :

$$E_{\text{reg}}(\mathbf{X}) = \sum_{i=1}^r \sum_{\mathbf{p}_j \in \mathcal{N}(\mathbf{p}_i)} \psi_{\nu_r}(\|\mathbf{D}_{ij}\|), \quad (5)$$

where  $r$  is the number of nodes in  $\mathcal{G}$ ,  $\nu_r > 0$  is a user-specified parameter, and  $\mathcal{N}(\mathbf{p}_i)$  denotes the set of neighboring vertices for  $\mathbf{p}_i$  in  $\mathcal{G}$ .

**Rotation Matrix Term.** To preserve rigidity on local surface regions during registration, we would like each transformation  $\mathbf{X}_i$  to be as close to a rigid transformation as possible. We measure this property using the deviation between the transformation matrix  $\mathbf{A}_i$  and its closest projection onto the rotation matrix group  $\mathcal{R} = \{\mathbf{R} \in \mathbb{R}^{3 \times 3} \mid \mathbf{R}^T \mathbf{R} = \mathbf{I}, \det(\mathbf{R}) > 0\}$ , and define  $E_{\text{rot}}$  as

$$E_{\text{rot}}(\mathbf{X}) = \sum_{i=1}^r \|\mathbf{A}_i - \text{proj}_{\mathcal{R}}(\mathbf{A}_i)\|_F^2, \quad (6)$$

where  $\text{proj}$  is the projection operator, and  $\|\cdot\|_F$  is the Frobenius norm.

### 3.2. Construction of Deformation Graph

To construct the deformation graph  $\mathcal{G}$ , we first extract its vertices  $\mathcal{V}_{\mathcal{G}}$  by iteratively adding source points as follows.  $\mathcal{V}_{\mathcal{G}}$  is initialized as empty. Then we perform PCA on all source points, and sort them based on their projections onto the axis with the largest eigenvalue of the covariance matrix. According to the sorted order, we go through each source point  $\mathbf{v}_i$ , and add it to  $\mathcal{V}_{\mathcal{G}}$  if  $\mathcal{V}_{\mathcal{G}}$  is empty or the shortest geodesic distance from  $\mathbf{v}_i$  to the points in  $\mathcal{V}_{\mathcal{G}}$  is no smaller than the radius parameter  $R$ . After determining the vertex set  $\mathcal{V}_{\mathcal{G}}$ , we construct the edge set  $\mathcal{E}_{\mathcal{G}}$  by connecting any two vertices whose geodesic distance is smaller than  $R$ . We compute the geodesic distance using the fast marching method [18].  $R$  is set to  $5\bar{l}$  by default, where  $\bar{l}$  is the average edge length on the source surface. Compared with alternative sampling strategies such as the farthest point sampling method [26], our approach results in fewer sample points for the same radius parameter, which reduces the computational cost while achieving similar accuracy. A comparison example is provided in the supplementary material.

## 4. Numerical Algorithm

The target function  $E$  for the optimization problem (2) is non-linear and non-convex. Thanks to the use of the Welsch's function, it can be solved efficiently using the MM algorithm [19]. Specifically, given the variable values  $\mathbf{X}^{(k)}$  in the current iteration, the MM algorithm constructs a surrogate function  $\bar{E}^{\mathbf{X}^{(k)}}$  for the target function  $E$  such that

$$\begin{aligned} \bar{E}^{\mathbf{X}^{(k)}}(\mathbf{X}^{(k)}) &= E(\mathbf{X}^{(k)}), \\ \bar{E}^{\mathbf{X}^{(k)}}(\mathbf{X}) &\geq E(\mathbf{X}), \quad \forall \mathbf{X} \neq \mathbf{X}^{(k)}. \end{aligned} \quad (7)$$

Then the variables are updated by minimizing the surrogate function

$$\mathbf{X}^{(k+1)} = \arg \min_{\mathbf{X}} \bar{E}^{\mathbf{X}^{(k)}}(\mathbf{X}). \quad (8)$$

In this way, each iteration is guaranteed to decrease the target function, and the iterations are guaranteed to converge to a local minimum regardless of the initialization [19]. In comparison, existing solvers for minimizing the non-convex  $\ell_p$ -norm such as ADMM [5] or iteratively reweighted least squares [9] either lack convergence guarantee or rely on strong assumptions for convergence. In the following, we explain the construction of the surrogate function, and present a numerical algorithm for its minimization in each iteration.

### 4.1. Surrogate Function

To construct the surrogate function  $\bar{E}^{\mathbf{X}^{(k)}}$ , we note that there is a convex quadratic surrogate function for the Welsch's function  $\psi_{\nu}$  at  $y$  [13] (see Fig. 1 right):

$$\bar{\psi}_{\nu}^y(x) = \psi_{\nu}(y) + \left( \frac{1 - \psi_{\nu}(y)}{2\nu^2} \right) (x^2 - y^2).$$

This function bounds the Welsch's function from above, and the two function graphs touch at  $x = y$ . Applying this surrogate to all relevant terms, we obtain a surrogate function for  $E_{\text{reg}}$  at  $\mathbf{X}^{(k)}$ . Moreover, we can ignore all constant terms as they do not affect the minimum, resulting in the following convex quadratic function

$$\bar{E}_{\text{reg}}^{\mathbf{X}^{(k)}} = \frac{1}{2\nu_r^2} \sum_{i=1}^r \sum_{\mathbf{p}_j \in \mathcal{N}(\mathbf{p}_i)} \exp\left(-\frac{\|\mathbf{D}_{ij}^{(k)}\|^2}{2\nu_r^2}\right) \|\mathbf{D}_{ij}\|^2, \quad (9)$$

where  $\mathbf{D}_{ij}^{(k)}$  is evaluated using Eq. (4) at  $\mathbf{X}^{(k)}$ . Similarly, we can obtain the following function as a surrogate for  $E_{\text{align}}$  at  $\mathbf{X}^{(k)}$  up to a constant:

$$\frac{1}{2\nu_a^2} \sum_{i=1}^n \exp\left(-\frac{\|\hat{\mathbf{v}}_i^{(k)} - \mathbf{u}_{\rho(i)}^{(k)}\|^2}{2\nu_a^2}\right) \|\hat{\mathbf{v}}_i - \mathbf{u}_{\rho(i)}\|^2, \quad (10)$$

where  $\hat{\mathbf{v}}_i^{(k)}$  is the transformed position of  $\mathbf{v}_i$  according to  $\mathbf{X}^{(k)}$ , and  $\mathbf{u}_{\rho(i)}^{(k)}$  is the closest target point for  $\hat{\mathbf{v}}_i^{(k)}$ . Eq. (10) is not a quadratic function of  $\mathbf{X}$ , as  $\mathbf{u}_{\rho(i)}$  depends non-linearly on  $\mathbf{X}$ . To obtain a more simple form, we note that the term  $\|\hat{\mathbf{v}}_i - \mathbf{u}_{\rho(i)}\|^2$  has a quadratic surrogate function  $\|\hat{\mathbf{v}}_i - \mathbf{u}_{\rho(i)}^{(k)}\|^2$  at  $\mathbf{X}^{(k)}$ . Applying it to Eq. (10), we obtain the following convex quadratic surrogate function for  $E_{\text{align}}$  at  $\mathbf{X}^{(k)}$  up to a constant:

$$\bar{E}_{\text{align}}^{\mathbf{X}^{(k)}} = \frac{1}{2\nu_a^2} \sum_{i=1}^n \exp\left(-\frac{\|\hat{\mathbf{v}}_i^{(k)} - \mathbf{u}_{\rho(i)}^{(k)}\|^2}{2\nu_a^2}\right) \|\hat{\mathbf{v}}_i - \mathbf{u}_{\rho(i)}^{(k)}\|^2. \quad (11)$$



Replacing  $E_{\text{align}}$  and  $E_{\text{reg}}$  in the target function with their surrogates, we arrive at the following MM iteration scheme:

$$\mathbf{X}^{(k+1)} = \min_{\mathbf{X}} \overline{E}^{\mathbf{X}^{(k)}}(\mathbf{X}), \quad (12)$$

where  $\overline{E}^{\mathbf{X}^{(k)}} = \overline{E}_{\text{align}}^{\mathbf{X}^{(k)}} + \alpha \overline{E}_{\text{reg}}^{\mathbf{X}^{(k)}} + \beta E_{\text{rot}}$ .

## 4.2. Numerical Minimization

The target function  $\overline{E}^{\mathbf{X}^{(k)}}$  in Eq. (12) still contains a non-linear term  $E_{\text{rot}}$ , as the projection onto the rotation matrix group depends non-linearly on  $\mathbf{A}_i$ . On the other hand, as explained later, the special structure of  $E_{\text{rot}}$  leads to a simple form of its gradient, allowing us to evaluate the gradient of  $\overline{E}^{\mathbf{X}^{(k)}}$  efficiently. Therefore, we solve the sub-problem (12) using an L-BFGS solver for fast convergence. In each iteration, L-BFGS utilizes the gradients of  $\overline{E}^{\mathbf{X}^{(k)}}$  at the latest  $m+1$  iterates  $\mathbf{X}_{(j)}$ ,  $\mathbf{X}_{(j-1)}$ ,  $\dots$ ,  $\mathbf{X}_{(j-m)}$  to implicitly approximate its inverse Hessian and derive a descent direction  $\mathbf{d}_{(j)}$ , followed by a line search along  $\mathbf{d}_{(j)}$  for a new iterate  $\mathbf{X}_{(j+1)}$  with sufficient decrease of the target function [27]. In the following, we present the details of the solver.

**Gradient Computation.** For the function  $E_{\text{rot}}$  defined in Eq. (6), each term  $\|\mathbf{A}_i - \text{proj}_{\mathcal{R}}(\mathbf{A}_i)\|_F^2$  is the squared Euclidean distance from the matrix  $\mathbf{A}_i$  to the manifold  $\mathcal{R}$  of rotation matrices. Even though  $\text{proj}_{\mathcal{R}}(\mathbf{A}_i)$  depends non-linearly on  $\mathbf{A}_i$ , it can be shown that the squared distance has a simple form of gradient [11]:

$$\frac{\partial \|\mathbf{A}_i - \text{proj}_{\mathcal{R}}(\mathbf{A}_i)\|_F^2}{\partial \mathbf{A}_i} = 2(\mathbf{A}_i - \text{proj}_{\mathcal{R}}(\mathbf{A}_i)).$$

Thus we can write the gradient of  $E_{\text{rot}}$  as:

$$\frac{\partial E_{\text{rot}}}{\partial \mathbf{A}_i} = 2(\mathbf{A}_i - \text{proj}_{\mathcal{R}}(\mathbf{A}_i)), \quad \frac{\partial E_{\text{rot}}}{\partial \mathbf{t}_i} = \mathbf{0}.$$

Since  $\overline{E}_{\text{align}}^{\mathbf{X}^{(k)}}$  and  $\overline{E}_{\text{reg}}^{\mathbf{X}^{(k)}}$  are quadratic functions, their gradients have simple linear forms. To facilitate presentation, we first rewrite the functions in matrix forms. In the following, we assume all 3D vectors are  $3 \times 1$  matrices. The transformation at  $\mathbf{p}_i$  is represented as  $\mathbf{X}_i = [\mathbf{A}_i, \mathbf{t}_i]^T \in \mathbb{R}^{4 \times 3}$ , and  $\mathbf{X} = [\mathbf{X}_1^T, \dots, \mathbf{X}_r^T]^T \in \mathbb{R}^{4r \times 3}$ . Then we have

$$\overline{E}_{\text{align}}^{\mathbf{X}^{(k)}} = \|\mathbf{W}_a(\mathbf{F}\mathbf{X} + \mathbf{P} - \mathbf{U})\|_F^2, \quad (13)$$

where  $\mathbf{W}_a = \text{diag}(\sqrt{w_1^a}, \dots, \sqrt{w_n^a}) \in \mathbb{R}^{n \times n}$  with  $w_i^a = \frac{1}{2\nu_i^2} \exp\left(-\frac{\|\hat{\mathbf{v}}_i^{(k)} - \mathbf{u}_{\rho(i)}^{(k)}\|^2}{2\nu_i^2}\right)$ ,  $\mathbf{F}$  is a block matrix  $\{\mathbf{F}_{ij}\}_{\substack{1 \leq i \leq n \\ 1 \leq j \leq r}} \in \mathbb{R}^{n \times 4r}$  with

$$\mathbf{F}_{ij} = \begin{cases} w_{ij} \cdot [\mathbf{v}_i^T - \mathbf{p}_j^T, 1] & \text{if } \mathbf{p}_j \in \mathcal{I}(\mathbf{v}_i) \\ \mathbf{0} & \text{otherwise} \end{cases},$$

---

**Algorithm 1:** Two-loop recursion for computing descent direction  $\mathbf{d}_{(j)}$

---

```

Q = -G(X(j));
for  $i = j - 1, \dots, j - m$  do
    S $i$  = X(i+1) - X $i$ ; T $i$  = G(X(i+1)) - G(X $i$ );
     $\rho_i$  =  $\text{Tr}(\mathbf{T}_i^T \mathbf{S}_i)$ ;
     $\xi_i$  =  $\text{Tr}(\mathbf{S}_i^T \mathbf{Q}) / \rho_i$ ;
    Q = Q -  $\xi_i \mathbf{T}_i$ ;
end
R = H0-1Q;
for  $i = j - m, \dots, j - 1$  do
     $\eta$  =  $\text{Tr}(\mathbf{T}_i^T \mathbf{R}) / \rho_i$ ;
    R = R + S $i$ ( $\xi_i - \eta$ );
end
 $\mathbf{d}_{(j)}$  = R;

```

---

and  $\mathbf{P} = \left[ \sum_{\mathbf{p}_j \in \mathcal{I}(\mathbf{v}_1)} \mathbf{p}_j, \dots, \sum_{\mathbf{p}_j \in \mathcal{I}(\mathbf{v}_n)} \mathbf{p}_j \right]^T \in \mathbb{R}^{n \times 3}$ ,  $\mathbf{U} = \left[ \mathbf{u}_{\rho(1)}^{(k)}, \dots, \mathbf{u}_{\rho(n)}^{(k)} \right]^T \in \mathbb{R}^{n \times 3}$ . Similarly, we have

$$\overline{E}_{\text{reg}}^{\mathbf{X}^{(k)}} = \|\mathbf{W}_r(\mathbf{B}\mathbf{X} - \mathbf{Y})\|_F^2, \quad (14)$$

where the matrices  $\mathbf{B} \in \mathbb{R}^{2|\mathcal{E}_{\mathcal{G}}| \times 4r}$  and  $\mathbf{Y} \in \mathbb{R}^{2|\mathcal{E}_{\mathcal{G}}| \times 3}$  encode the computation of a term  $\mathbf{D}_{ij}$  in each row, and the diagonal matrix  $\mathbf{W}_r$  stores the weight  $\sqrt{\frac{1}{2\nu_r^2} \exp\left(-\frac{\|\mathbf{D}_{ij}^{(k)}\|^2}{2\nu_r^2}\right)}$  at the corresponding row. In the row corresponding to  $\mathbf{D}_{ij}$ , the elements in  $\mathbf{Y}$  are  $[\mathbf{p}_j^T - \mathbf{p}_i^T]$ , whereas the elements in  $\mathbf{B}$  for  $\mathbf{X}_i$  and  $\mathbf{X}_j$  are  $[\mathbf{p}_i^T - \mathbf{p}_j^T, 1]$  and  $[0, 0, 0, 1]$ , respectively. We can further write the gradient of  $E_{\text{rot}}$  in matrix form as

$$\frac{\partial E_{\text{rot}}}{\partial \mathbf{X}} = 2(\mathbf{J}\mathbf{X} - \mathbf{Z}),$$

where  $\mathbf{Z} = [\text{proj}_{\mathcal{R}}(\mathbf{A}_1), \mathbf{0}, \dots, \text{proj}_{\mathcal{R}}(\mathbf{A}_r), \mathbf{0}]^T \in \mathbb{R}^{4r \times 3}$ , and  $\mathbf{J} = \text{diag}(1, 1, 1, 0, 1, 1, 1, 0, \dots, 1, 1, 1, 0) \in \mathbb{R}^{4r \times 4r}$ .

We can then derive the gradient function of  $\overline{E}^{\mathbf{X}^{(k)}}$  as

$$\mathbf{G}(\mathbf{X}) = 2[\mathbf{F}^T \mathbf{W}_a^2 (\mathbf{F}\mathbf{X} + \mathbf{P} - \mathbf{U}) + \alpha \mathbf{B}^T \mathbf{W}_r^2 (\mathbf{B}\mathbf{X} - \mathbf{Y}) + \beta (\mathbf{J}\mathbf{X} - \mathbf{Z})]. \quad (15)$$

**Computing  $\mathbf{d}_{(j)}$  and  $\mathbf{X}_{(j+1)}$ .** We adopt the L-BFGS implementation from [23] to utilize the special structure of  $\overline{E}^{\mathbf{X}^{(k)}}$ . Specifically, given an initial approximation  $\mathbf{H}_0$  for the Hessian of  $\overline{E}^{\mathbf{X}^{(k)}}$  at  $\mathbf{X}_{(j)}$ , we compute the descent direction  $\mathbf{d}_{(j)}$  using a two-loop recursion explained in [23] (see Algorithm 1). Following [23], we derive  $\mathbf{H}_0$  by assuming a fixed projection  $\text{proj}_{\mathcal{R}}(\mathbf{A}_i)$  for the term  $E_{\text{rot}}$ , resulting in the following approximation:

$$\mathbf{H}_0 = 2(\mathbf{F}^T \mathbf{W}_a^2 \mathbf{F} + \alpha \mathbf{B}^T \mathbf{W}_r^2 \mathbf{B} + \beta \mathbf{J}). \quad (16)$$

The new iterate  $\mathbf{X}_{(j+1)}$  is then computed as

$$\mathbf{X}_{(j+1)} = \mathbf{X}_{(j)} + \lambda \mathbf{d}_{(j)},$$

using a line search to determine the step size  $\lambda > 0$  that achieve sufficient decrease of  $\overline{E}^{\mathbf{X}^{(k)}}$ :

$$\overline{E}^{\mathbf{X}^{(k)}}(\mathbf{X}_{(j+1)}) \leq \overline{E}^{\mathbf{X}^{(k)}}(\mathbf{X}_{(j)}) + \gamma \lambda \text{Tr}((\mathbf{G}(\mathbf{X}_{(j)}))^T \mathbf{d}_{(j)}), \quad (17)$$

with  $\gamma \in (0, 1)$ . The iterative L-BFGS solver is terminated if  $\overline{E}^{\mathbf{X}^{(k)}}(\mathbf{X}_{(j)}) - \overline{E}^{\mathbf{X}^{(k)}}(\mathbf{X}_{(j+1)}) < \epsilon_1$  where  $\epsilon_1$  is a user-specified threshold. And the outer MM iteration is terminated if  $\max_i \|\widehat{\mathbf{v}}_i^{(k+1)} - \widehat{\mathbf{v}}_i^{(k)}\| < \epsilon_2$  with a user-specified threshold  $\epsilon_2$  or the number of outer iterations reaches a threshold  $I_{\max}$ . In all experiments, we set  $m = 5$ ,  $\gamma = 0.3$ ,  $\epsilon_1 = 10^{-3}$ ,  $\epsilon_2 = 10^{-3}$ , and  $I_{\max} = 100$ .

The matrix  $\mathbf{H}_0$  is sparse symmetric positive definite and remains unchanged in each iteration of an L-BFGS solver. To solve the linear equation  $\mathbf{R} = \mathbf{H}_0^{-1} \mathbf{Q}$  efficiently in the two-loop recursion, we compute a sparse Cholesky factorization for  $\mathbf{H}_0$  at the beginning of an L-BFGS run, and reuse it in each iteration to solve the linear system with different right-hand-sides. In addition, the columns of the right-hand-side matrix  $\mathbf{Q}$  are independent, thus we solve them in parallel. Moreover, although the matrix  $\mathbf{H}_0$  may change between different L-BFGS runs, its sparsity pattern remains the same. Therefore, we pre-compute a symbolic factorization of  $\mathbf{H}_0$  and only perform numerical factorization subsequently.

**Choosing  $\nu_a$  and  $\nu_r$ .** To achieve robust registration, the values of  $\nu_a$  and  $\nu_r$  play an important role. Each of them acts as the standard deviation parameter for the Gaussian weight in the surrogate function (9) or (11). The Gaussian weight becomes effectively zero for error terms whose current magnitude is much larger than the standard deviation. Both  $\nu_a$  and  $\nu_r$  need to be sufficiently small at the final stage of the solver, to exclude the influence of large error terms that arise from partial overlaps etc. On the other hand, at the initial stage of the solver, their values need to be larger in order to accommodate more error terms and achieve coarse alignment. Therefore, we initialize the variables  $\mathbf{X}$  with a rigid transformation (see Sec. 5 for details), and solve the problem (2) with relatively large values  $\nu_a = \nu_a^{\max}$  and  $\nu_r = \nu_r^{\max}$ . The solution is then used as initial variable values to re-solve the problem (2) with the values of  $\nu_a$  and  $\nu_r$  decreased by half. We repeat this process until the value of  $\nu_a$  reaches a lower bound  $\nu_a^{\min}$ , and take the final solution as the registration result. Algorithm 2 summarizes our overall registration algorithm with gradually decreased values of  $\nu_a$  and  $\nu_r$ . By default, we set  $\nu_r^{\max} = 40\bar{l}$ ,  $\nu_a^{\max} = 10\bar{l}$  and  $\nu_a^{\min} = 0.5\bar{l}$ , where  $\bar{l}$  is the medium distance from source points to their corresponding target points under the initial rigid transformation, and  $\bar{l}$  is the average edge length on the

---

### Algorithm 2: Non-rigid registration

---

```

 $\nu_a = \nu_a^{\max}; \nu_r = \nu_r^{\max};$ 
while TRUE do
     $k = 0;$ 
    repeat
        Find corresponding point  $\mathbf{u}_{\rho(i)}^{(k)}$  for each  $\widehat{\mathbf{v}}_i^{(k)}$ ;
        Update weight matrices  $\mathbf{W}_a$  and  $\mathbf{W}_r$ ;
         $j = -1; \mathbf{X}_{(0)} = \mathbf{X}^{(k)};$ 
        repeat
             $j = j + 1;$ 
            Compute gradient  $\mathbf{G}(\mathbf{X}_{(j)})$  with (15);
            Compute direction  $\mathbf{d}_{(j)}$  with Alg. 1;
            Perform line search for  $\mathbf{X}_{(j+1)}$  that
                satisfies condition (17);
        until  $\overline{E}^{\mathbf{X}^{(k)}}(\mathbf{X}_{(j)}) - \overline{E}^{\mathbf{X}^{(k)}}(\mathbf{X}_{(j+1)}) < \epsilon_1;$ 
         $\mathbf{X}^{(k+1)} = \mathbf{X}_{(j+1)};$ 
         $k = k + 1;$ 
    until  $\max_i \|\widehat{\mathbf{v}}_i^{(k+1)} - \widehat{\mathbf{v}}_i^{(k)}\| < \epsilon_2$  OR  $k = I_{\max};$ 
    if  $\nu_a = \nu_a^{\min}$  then
        | return  $\mathbf{X}^{(k+1)};$ 
    end
     $\nu_a = \max(0.5 \cdot \nu_a, \nu_a^{\min}); \nu_r = 0.5 \cdot \nu_r;$ 
end

```

---

source surface. In the supplementary material, we compare our strategy of gradually decreasing  $\nu_a$  and  $\nu_r$  with registration using fixed  $\nu_a$  and  $\nu_r$ , which shows that our strategy achieves a more accurate result.

**Setting of  $\alpha$  and  $\beta$ .** To account for the number of terms in each component of the target function in Eq. (2), we set the weights as  $\alpha = k_\alpha (|\mathcal{V}|/|\mathcal{E}_{\mathcal{G}}|)$  and  $\beta = k_\beta (|\mathcal{V}|/|\mathcal{V}_{\mathcal{G}}|)$ , with  $k_\alpha = k_\beta = 1$  by default. We decrease the values of  $k_\alpha$  and  $k_\beta$  on problem instances with reliable initialization to achieve better alignment, and increase their values on problem instances with less reliable initialization to avoid converging to an undesirable local minimum. Moreover, the value of  $k_\alpha$  can be increased for smoother deformation of the source surface, or decreased for faster convergence.

## 5. Results

In this section, we evaluate the effectiveness of our approach, and compare its performance with state-of-the-art methods including N-ICP [1], RPTS [22], and SVR- $\ell_0$  [12]. The evaluations are performed on the MPI Faust dataset [4] and the human motion datasets from [33]. We only show representative results in this section, with more comparisons provided in the supplementary material.

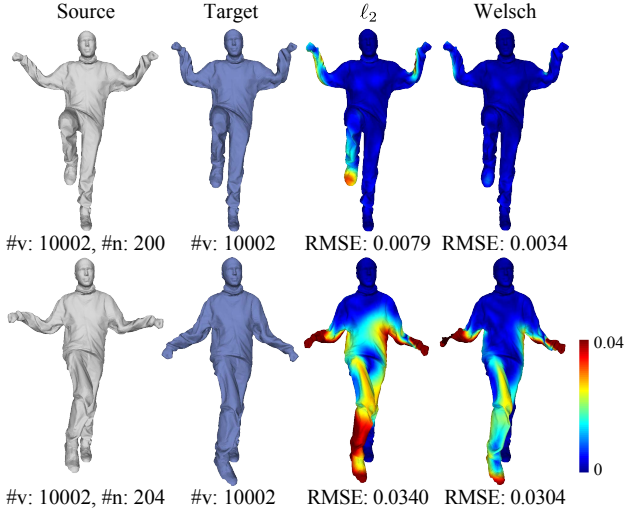


Figure 2. Comparison between our formulation and an alternative formulation using the  $\ell_2$ -norm instead of the Welsch’s function, tested on two pairs of meshes from the “crane” dataset of [33]. We set with  $k_\alpha = 0.1, k_\beta = 100$  for the  $\ell_2$ -norm formulation, and  $k_\alpha = 1, k_\beta = 10^4$  for our formulation.

**Implementation Details** We implement our method in C++, and all results and comparisons are tested on a PC with 16GB of RAM and a 6-core CPU at 3.60GHz. Each pair of surfaces are pre-processed by aligning their centroids and scaling them to achieve unit-length diagonal for their combined bounding box. We initialize the transformation variables using a rigid transformation computed via 15 iterations of the ICP algorithm to align the source and target points, with rejection of point pairs whose distance is larger than a threshold  $\epsilon_d$  or whose normal vectors deviate by more than an angle  $\theta$  [28]. We set  $\epsilon_d = 0.3$  and  $\theta = 60^\circ$  by default. The ICP iterations are initialized by aligning corresponding feature points, determined either using the closest point pairs between the pre-processed surfaces coupled with the above rejection criteria, or using the SHOT feature [32] with diffusion pruning [30], or through manual labeling. In the figures, we visualize the initial correspondence using blue lines connecting the corresponding points. We evaluate the registration accuracy via the root mean square error compared with the ground truth:

$$\text{RMSE} = \sqrt{\frac{\sum_{\mathbf{v}_i \in \mathcal{V}} e_i^2}{|\mathcal{V}|}}, \quad (18)$$

where  $e_i = \|\mathbf{v}_i^* - \mathbf{v}_i^{\text{gt}}\|$  is the deviation between the transformed positions  $\mathbf{v}_i^*$ ,  $\mathbf{v}_i^{\text{gt}}$  of a source point  $\mathbf{v}_i$  under the computed and the ground-truth transformations respectively. In the figures, we show the RMSE for each registration result, and use color-coding to visualize the error values  $\{e_i\}$  across the surface. Unless stated otherwise, all RMSE values and color-coded registration errors are in the unit of meters. We use #v to denote the number of sample points on a surface,

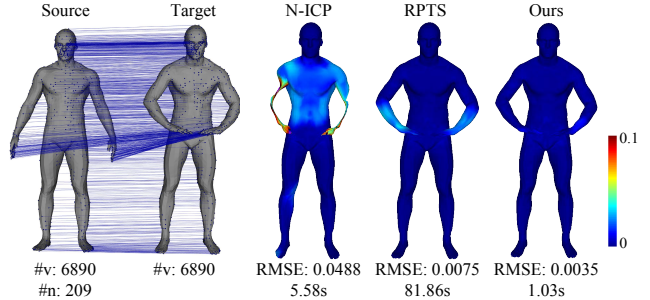


Figure 3. Comparison of registration results on an example from the MPI Faust dataset.

Table 1. Average computation time and RMSE (in millimeters) on the MPI Faust dataset. We set  $\alpha = 5$  for N-ICP,  $\alpha = 100, \beta = 0.1$  for RPTS, and  $k_\alpha = k_\beta = 0.001$  for our method.

Pose Pair	N-ICP		RPTS		Ours	
	Time (s)	RMSE	Time (s)	RMSE	Time (s)	RMSE
1	7.43	59.7	61.83	9.16	<b>1.24</b>	<b>5.46</b>
2	4.46	20.8	46.52	3.22	<b>0.86</b>	<b>1.35</b>
3	7.95	80.1	35.25	<b>7.97</b>	<b>1.36</b>	8.99
4	5.45	45.4	22.99	5.69	<b>1.27</b>	<b>4.26</b>
5	4.13	14.2	14.04	<b>0.717</b>	<b>0.81</b>	1.06
6	8.61	59.6	28.72	6.29	<b>1.30</b>	<b>3.89</b>
7	12.57	208	43.26	61.6	<b>2.23</b>	<b>49.3</b>
8	6.05	70.4	26.27	<b>6.58</b>	<b>1.34</b>	7.54
Mean	7.08	69.7	34.86	12.7	<b>1.30</b>	<b>10.2</b>
Median	6.74	59.6	31.99	6.44	<b>1.28</b>	<b>4.86</b>

and #n for the number of deformation graph vertices. Similar to our formulation (2), the optimization target functions of N-ICP, RPTS, and SVR- $\ell_0$  all involve a regularization term and/or a term for the rigidity of transformations. To ease presentation, for all methods we use  $\alpha$  and  $\beta$  to denote the weights for the regularization term and the rigidity term, respectively. We tune the weights for each method to obtain their best results for comparison. In addition, N-ICP and RPTS require dropping unreliable correspondence between source and target points in each iteration, and we adopt the distance and normal deviation thresholds  $\epsilon_d$  and  $\theta$  as used in the ICP iterations for initialization.

## 5.1. Effectiveness of the Welsch’s Function

We perform the optimization (2) with the Welsch’s functions in  $E_{\text{align}}$  and  $E_{\text{reg}}$  replaced by square functions. Fig. 2 compares the registration accuracy of such  $\ell_2$ -norm formulation with our approach, on two problem instances from the “crane” dataset of [33]. we can see that the Welsch’s function leads to more accurate results than the  $\ell_2$ -norm.

## 5.2. Comparison with Other Methods

In Tab. 1 and Fig. 3, we compare our method with N-ICP and RPTS on the MPI Faust dataset. We select 10 subjects with 9 poses, and use the first pose of each subject as the

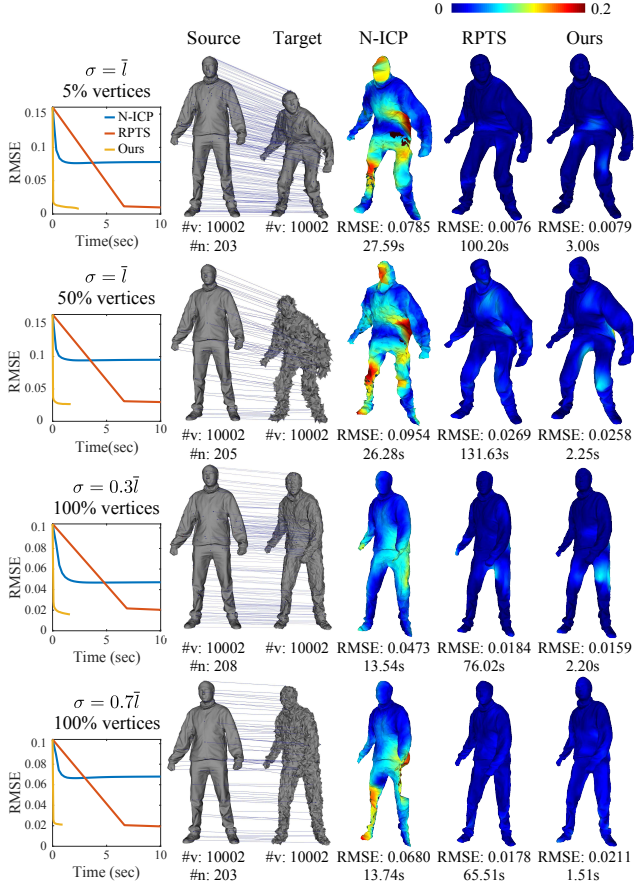


Figure 4. Comparison with N-ICP and RPTS on noisy models constructed by adding Gaussian noise to some vertices on the target surface. The original models are taken from the “jumping” dataset of [33]. The left column shows how the RMSE of registration results changes over time during optimization. We set  $\alpha = 5$  for N-ICP,  $\alpha = 10, \beta = 1$  for RPTS,  $k_\alpha = 0.01, k_\beta = 1$  for our method, and  $\theta = 180^\circ$  for all methods.

source surface the other poses of the subject as the targets. We evaluate the average RMSE among all subjects for the same pose pair, and list them in Tab. 1. Fig. 3 shows the results from different methods on a pose pair for a subject. We can see that our method requires significantly less computational time while achieving similar or better accuracy. A major reason for the efficiency of our method is the adoption of a deformation graph, which only requires optimizing one affine transformation per graph vertex. In comparison, N-ICP and RPTS require optimizing one affine transformation per mesh vertex, which significantly increases the number of variables as well as computational time.

Fig. 4 compares our method with N-ICP and RPTS on models from the “jumping” dataset of [33], with added Gaussian noise on vertex positions of the target surface along their normal directions. In the first two rows of Fig. 4, we add noise with standard deviation  $\sigma = \bar{l}$  to 5% or 50% of the vertices on the target surface respectively, where  $\bar{l}$  is the

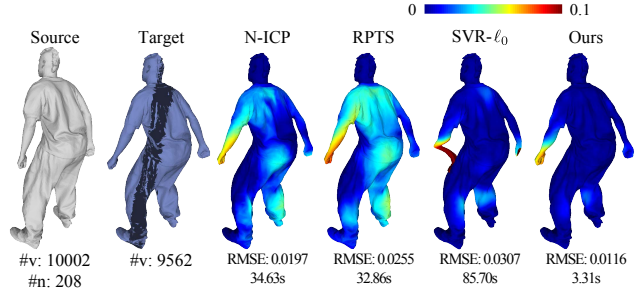


Figure 5. Comparison with N-ICP, RPTS and SVR- $\ell_0$  on models with partial overlaps, constructed by removing some vertices from the target surface. The original models are taken from the “bouncing” dataset of [33]. We set  $\alpha = 10$  for N-ICP,  $\alpha = 1, \beta = 100$  for RPTS,  $\alpha = 0.1, \beta = 100$  for SVR- $\ell_0$ ,  $k_\alpha = 1, k_\beta = 100, \nu_a^{\max} = 30\bar{d}, \nu_r^{\max} = 100\bar{l}$  for our method, and  $\theta = 45^\circ$  for all methods.

average edge length of the target surface. In the last two rows, we add noise to all vertices of the target surface with stand deviation  $\sigma = 0.3\bar{l}$  and  $\sigma = 0.7\bar{l}$ , respectively. The comparison shows that our method is more robust to noisy data.

In Fig. 11, we compare our method with N-ICP, RPTS and SVR- $\ell_0$  on partially overlapping data, which is synthesized from the “bouncing” dataset of [33] by removing some vertices from the target surface. For SVR- $\ell_0$ , we use the squared distance from source points to their closest target points as the data fitting term. We can see that our method is significantly faster than other methods, and is more robust to such partial overlapping model.

## 6. Conclusion

In this paper, we proposed a robust non-rigid registration model based on the Welsch’s function. Applying the Welsch’s function to the alignment term and the regularization term makes the formulation robust to the noise and partial overlap. To efficiently solve this problem, we apply majorization-minimization to transform the nonlinear and non-convex problem into a sequence of simple sub-problems that are efficiently solved with L-BFGS. Extensive experiments demonstrate the effectiveness of our method and its efficiency compared to existing approaches.

**Acknowledgement** We thank the authors of [12] for providing their implementation. This work was supported by the National Natural Science Foundation of China (No. 61672481), and Youth Innovation Promotion Association CAS (No. 2018495).

## References

- [1] Brian Amberg, Sami Romdhani, and Thomas Vetter. Optimal step nonrigid ICP algorithms for surface registration. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8. IEEE, 2007. 1, 2, 6



- [2] Myronenko Andriy and Song Xubo. Point set registration: coherent point drift. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(12):2262–2275, 2010. 2
- [3] Paul J. Besl and Neil D. McKay. A method for registration of 3-d shapes. *IEEE Trans. Pattern Anal. Mach. Intell.*, 14(2):239–256, 1992. 3
- [4] Federica Bogo, Javier Romero, Matthew Loper, and Michael J. Black. FAUST: Dataset and evaluation for 3D mesh registration. In *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Piscataway, NJ, USA, June 2014. IEEE. 6
- [5] Sofien Bouaziz, Andrea Tagliasacchi, and Mark Pauly. Sparse iterative closest point. *Computer Graphics Forum*, 32(5):113–123, 2013. 3, 4
- [6] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, Jonathan Eckstein, et al. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3(1):1–122, 2011. 1
- [7] Benedict J Brown and Szymon Rusinkiewicz. Global non-rigid alignment of 3-d scans. *ACM Transactions on Graphics (TOG)*, 26(3), 2007. 2
- [8] Haili Chui and A. Rangarajan. A new algorithm for non-rigid point matching. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 44–51. IEEE, 2000. 2
- [9] Ingrid Daubechies, Ronald DeVore, Massimo Fornasier, and C. Sinan Güntürk. Iteratively reweighted least squares minimization for sparse recovery. *Communications on Pure and Applied Mathematics*, 63(1):1–38, 2010. 4
- [10] Roberto M Dyke, Yu-Kun Lai, Paul L Rosin, and Gary KL Tam. Non-rigid registration under anisotropic deformations. *Computer Aided Geometric Design*, 71:142–156, 2019. 2
- [11] José Gomes and Olivier Faugeras. The vector distance functions. *International Journal of Computer Vision*, 52(2):161–187, 2003. 5
- [12] Kaiwen Guo, Feng Xu, Yangang Wang, Yebin Liu, and Qionghai Dai. Robust non-rigid motion tracking and surface reconstruction using L0 regularization. In *IEEE International Conference on Computer Vision (ICCV)*, pages 3083–3091, 2015. 1, 2, 6, 8
- [13] Bumsub Ham, Minsu Cho, and Jean Ponce. Robust image filtering using joint static and dynamic guidance. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4823–4831, 2015. 1, 3, 4
- [14] Paul W Holland and Roy E Welsch. Robust regression using iteratively reweighted least-squares. *Communications in Statistics-theory and Methods*, 6(9):813–827, 1977. 1, 3
- [15] Hidekata Hontani, Takamiti Matsuno, and Yoshihide Sawada. Robust nonrigid icp using outlier-sparsity regularization. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 174–181. IEEE, 2012. 2
- [16] Bing Jian and Baba C Vemuri. A robust algorithm for point set registration using mixture of gaussians. In *IEEE International Conference on Computer Vision (ICCV)*, volume 2, pages 1246–1251. IEEE, 2005. 2
- [17] Tao Jiang, Xiaosong Yang, Jianjun Zhang, Feng Tian, Shuang Liu, Nan Xiang, and Kun Qian. Huber-  $l_1$ -based non-isometric surface registration. *The Visual Computer*, 35(6-8):935–948, 2019. 1, 2
- [18] Ron Kimmel and James A Sethian. Computing geodesic paths on manifolds. *Proceedings of the national academy of Sciences*, 95(15):8431–8435, 1998. 4
- [19] Kenneth Lange. *Optimization*, chapter The MM Algorithm, pages 119–136. Springer New York, 2004. 1, 4
- [20] Hao Li, Bart Adams, Leonidas J Guibas, and Mark Pauly. Robust single-view geometry and motion reconstruction. *ACM Transactions on Graphics (ToG)*, 28(5):175, 2009. 1, 2
- [21] Hao Li, Robert W Sumner, and Mark Pauly. Global correspondence optimization for non-rigid registration of depth scans. *Computer Graphics Forum*, 27(5):1421–1430, 2008. 1, 2
- [22] Kun Li, Jingyu Yang, Yu-Kun Lai, and Daoliang Guo. Robust non-rigid registration with reweighted position and transformation sparsity. *IEEE Transactions on Visualization and Computer Graphics*, 25(6):2255–2269, 2018. 1, 2, 3, 6
- [23] Tiantian Liu, Sofien Bouaziz, and Ladislav Kavan. Quasi-newton methods for real-time simulation of hyperelastic materials. *ACM Transactions on Graphics (TOG)*, 36(3):23, 2017. 5
- [24] Jiayi Ma, Weichao Qiu, Ji Zhao, Yong Ma, Alan L. Yuille, and Zhuowen Tu. Robust  $L_2E$  estimation of transformation for non-rigid registration. *IEEE Transactions on Signal Processing*, 63(5):1115–1129, 2015. 2
- [25] Niloy J. Mitra, Natasha Gelfand, Helmut Pottmann, and Leonidas Guibas. Registration of point cloud data from a geometric optimization perspective. In *Proceedings of the 2004 Eurographics/ACM SIGGRAPH Symposium on Geometry Processing*, SGP '04, pages 22–31. ACM, 2004. 3
- [26] Carsten Moenning and Neil A Dodgson. Fast marching farthest point sampling. Technical report, University of Cambridge, Computer Laboratory, 2003. 4, 10
- [27] Jorge Nocedal and Stephen J. Wright. *Numerical Optimization*. Springer, New York, 2nd edition, 2006. 5
- [28] S. Rusinkiewicz and M. Levoy. Efficient variants of the icp algorithm. In *Proceedings Third International Conference on 3-D Digital Imaging and Modeling*, pages 145–152, 2001. 7
- [29] Robert W Sumner, Johannes Schmid, and Mark Pauly. Embedded deformation for shape manipulation. *ACM Transactions on Graphics (TOG)*, 26(3):80, 2007. 2, 3
- [30] Gary KL Tam, Ralph Robert Martin, Paul L Rosin, and Yukun Lai. Diffusion pruning for rapidly and robustly selecting global correspondences using local isometry. *ACM Transactions on Graphics*, 33(1), 2014. 7
- [31] Gary K. L. Tam, Zhi-Quan Cheng, Yu-Kun Lai, Frank C. Langbein, Yonghuai Liu, A. David Marshall, Ralph R. Martin, Xianfang Sun, and Paul L. Rosin. Registration of 3d point clouds and meshes: A survey from rigid to nonrigid. *IEEE Transactions on Visualization and Computer Graphics*, 19(7):1199–1217, 2013. 2
- [32] Federico Tombari, Samuele Salti, and Luigi Di Stefano. Unique signatures of histograms for local surface description. In *European conference on computer vision*, pages 356–369. Springer, 2010. 7

- [33] Daniel Vlasic, Ilya Baran, Wojciech Matusik, and Jovan Popović. Articulated mesh animation from multi-view silhouettes. In *ACM Transactions on Graphics (TOG)*, volume 27, page 97. ACM, 2008. 6, 7, 8
- [34] Michael Wand, Bart Adams, Maksim Ovsjanikov, Alexander Berner, Martin Bokeloh, Philipp Jenke, Leonidas Guibas, Hans Peter Seidel, and Andreas Schilling. Efficient reconstruction of nonrigid shape and motion from real-time 3d scanner data. *ACM Transactions on Graphics (TOG)*, 28(2):1–15, 2009. 2
- [35] Zhenchao Wu, Kun Li, Yu-Kun Lai, and Jingyu Yang. Global as-conformal-as-possible non-rigid registration of multi-view scans. In *IEEE International Conference on Multimedia and Expo (ICME)*, pages 308–313. IEEE, 2019. 2
- [36] Jingyu Yang, Ke Li, Kun Li, and Yu-Kun Lai. Sparse non-rigid registration of 3d shapes. *Computer Graphics Forum*, 34(5):89–99, 2015. 1, 2
- [37] Juyong Zhang, Bailin Deng, Yang Hong, Yue Peng, Wenjie Qin, and Ligang Liu. Static/dynamic filtering for mesh geometry. *IEEE transactions on visualization and computer graphics*, 25(4):1774–1787, 2018. 1, 3

## Appendix

### The choice of sampling radius

The number of graph nodes and edges will influence the memory footprint and computational cost for the solver. The farthest point sampling method [26] is to repeatedly add the farthest point to the graph until the geodesic distance between graph nodes and the farthest point is smaller than the given radius parameter  $R$ . Compared with farthest point sampling method (Fig. 6), our adopt method can obtain fewer nodes and is faster to converge with the similar accuracy. In our method, the radius  $R$  can be used to balance the speed and accuracy. A smaller  $R$  leads to more nodes in the deformation graph, which increases the number of variables and accuracy while requires more computational time. We show the comparison in Fig. 7. Our method does not vary the sampling density based on curvature. In our experiments, such uniform density is sufficient to generate good results and curvature-adaptive sampling can be a future work.

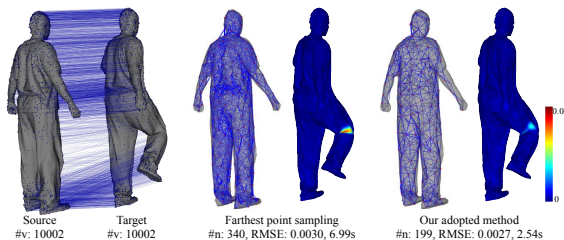


Figure 6. Comparison with the farthest sampling method for the given radius  $R = 5\bar{l}$ , where  $\bar{l}$  is the average edge length on the source surface. The farthest point sampling method can obtain more graph nodes. ( $k_\alpha = 0.001, k_\beta = 0.1$ ). The RMSE and the color-coded registration errors are in the unit of meters.

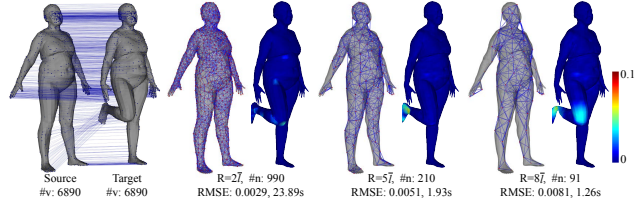


Figure 7. Comparison with different sampling radius with  $k_\alpha = 0.001, k_\beta = 0.1$ . More nodes can get more accurate results. The RMSE and the color-coded registration errors are in the unit of meters.

### The comparison with fixed parameters

In our method, the  $\nu_a$  and  $\nu_r$  values will influence the registered result, and discussion on this part is given in “Choosing  $\nu_a$  and  $\nu_r$ ” (Sec 4.2) of the paper. In Fig. 8, we show the comparison between our dynamic adjustment strategy and the strategy by fixing  $\nu_a, \nu_r$ , and we can see our method can get higher accuracy.

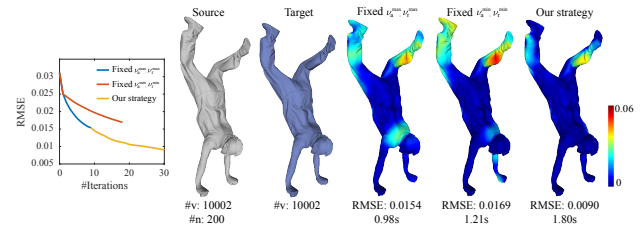


Figure 8. Comparison with fixed  $\nu_a, \nu_r$  on 42-th to 40-th mesh in “handstand” with  $k_\alpha = 100$  and  $k_\beta = 50$ . Here  $\nu_r^{\min}$  is the value when  $\nu_a$  reaches  $\nu_a^{\min}$ . The RMSE and the color-coded registration errors are in the unit of meters.

### Experiment on clean data

We show more results on five models “crane”, “march1”, “samba”, “squat1” and “swing” in Human-motion datasets. For each model, we use the closest points to construct the correspondences for small deformation, and use the SHOT with diffusion pruning method for big deformation. For each method, we search the parameter setting for best performance. The results are shown in Fig. 9 and Fig. 10, and we can see that our method is faster than other methods and achieves similar or better accuracy.

### Experiment on partially overlapping data

We show more results and comparisons on partially overlapping data from “bouncing” datasets. We choose  $\alpha = 10$  for N-ICP,  $\alpha = 1, \beta = 100$  for RPTS,  $\alpha = 0.1, \beta = 100$  for SVR- $\ell_0$ , and  $k_\alpha = 1, k_\beta = 100, \nu_a^{\max} = 30\bar{d}, \nu_r^{\max} = 100\bar{l}$  for our method and  $\theta = 45^\circ$  for all method in these examples. The results are shown in Fig. 11, and we can see that our methods is robust to partially overlapping data, and faster than other methods.

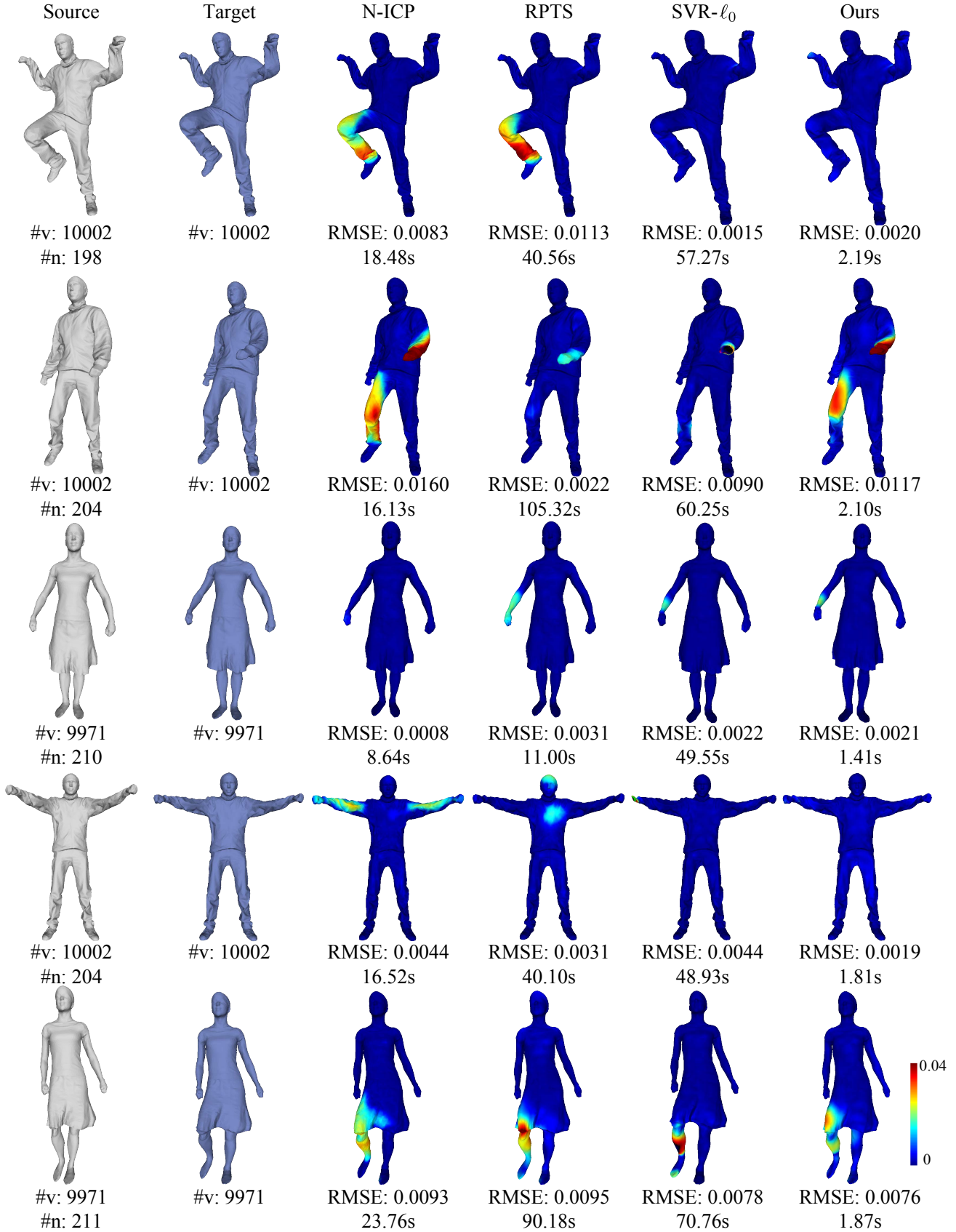


Figure 9. Comparison with N-ICP, RPTS and SVR- $\ell_0$  on “crane”, “march1”, “samba”, “squat1” and “swing” datasets with small deformation. We set  $\alpha = 10$  for N-ICP,  $\alpha = 10$  and  $\beta = 1$  for RPTS,  $\alpha = 0.1$  and  $\beta = 100$  for SVR- $\ell_0$ , and  $k_\alpha = 1, k_\beta = 10^3, \nu_a = 30\bar{j}$  for our method in these examples. The RMSE and the color-coded registration errors are in the unit of meters.

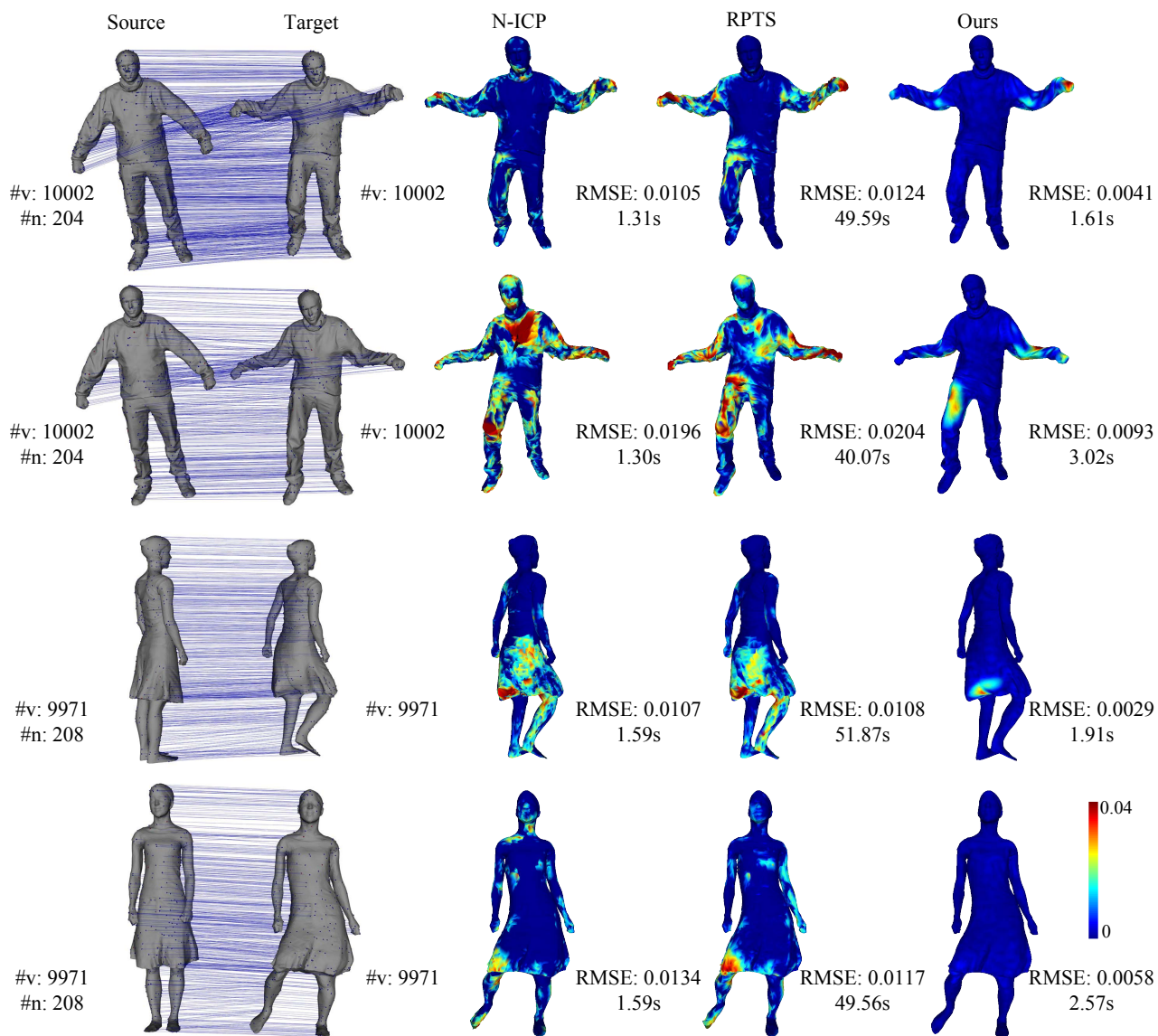


Figure 10. Comparison with N-ICP, RPTS and SVR- $\ell_0$  on “crane”, and “swing” datasets with big deformation. We set  $\alpha = 0.01$  for N-ICP,  $\alpha = 0.01$  and  $\beta = 1$  for RPTS, and  $k_\alpha = 0.01$  and  $k_\beta = 1$  for our method in these examples. The RMSE and the color-coded registration errors are in the unit of meters.



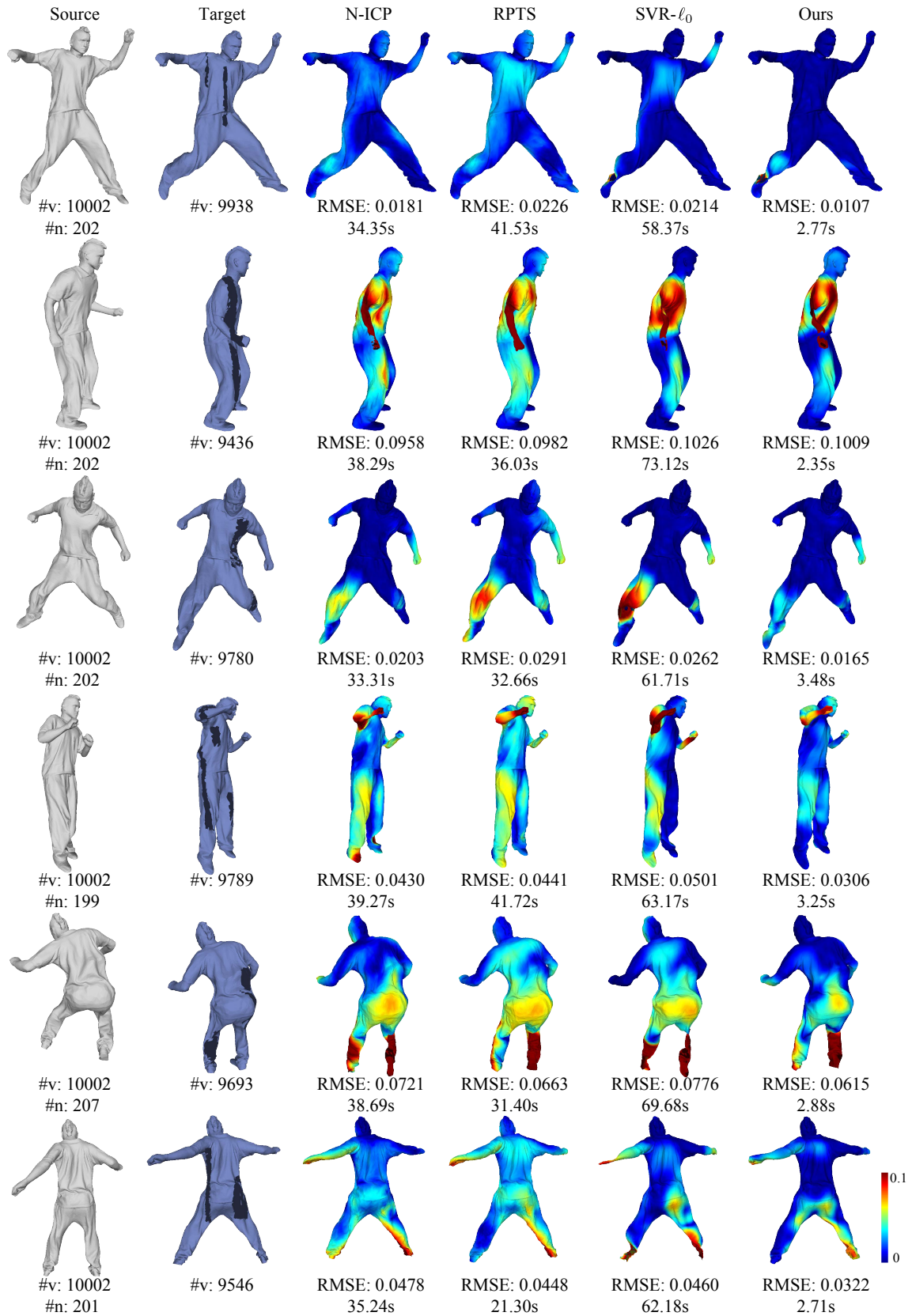


Figure 11. Comparison with N-ICP, RPTS and SVR- $\ell_0$  on “bouncing” datasets with partially overlapping data. The RMSE and the color-coded registration errors are in the unit of meters.