Robust Non-Rigid Registration with Reweighted Position and Transformation Sparsity

Kun Li, Member, IEEE, Jingyu Yang, Senior Member, IEEE,
Yu-Kun Lai, Member, IEEE, and Daoliang Guo

Abstract—Non-rigid registration is challenging because it is ill-posed with high degrees of freedom and is thus sensitive to noise and outliers. We propose a robust non-rigid registration method using reweighted sparsities on position and transformation to estimate the deformations between 3-D shapes. We formulate the energy function with position and transformation sparsity on both the data term and the smoothness term, and define the smoothness constraint using local rigidity. The double sparsity based non-rigid registration model is enhanced with a reweighting scheme, and solved by transferring the model into four alternately-optimized subproblems which have exact solutions and guaranteed convergence. Experimental results on both public datasets and real scanned datasets show that our method outperforms the state-of-the-art methods and is more robust to noise and outliers than conventional non-rigid registration methods.

Index Terms—Non-rigid registration, noise and outliers, deformation, position sparsity, transformation sparsity

1 INTRODUCTION

Non-rigid registration is an active research area in computer graphics and computer vision [17], [24], [32], [37], and is a key technique for dynamic 3-D reconstruction using a depth camera. Commodity depth sensors, e.g., Microsoft Kinect, become cheaper and more widely used, but depth images and reconstructed point clouds captured by such devices contain much noise. Hence, non-rigid registration methods robust to noise and outliers are highly desirable to scan dynamic scenes with deformable objects.

Given two input 3-D shapes, one as the template shape and the other as the target shape, non-rigid registration aims to find a suitable transformation that aligns the template shape to the target shape. Non-rigid registration is often formulated as an optimization problem. Most methods formulate some energy functional with both position and transformation constraints. The position constraint measures the closeness of the transformed template shape and the target shape, and the transformation constraint measures the fitness to model, which might include the smoothness, namely the total energy of transformation differences of all the local neighbors. Most work uses the classic squared \( \ell_2 \)-norm in the position constraint and the transformation constraint [21], [3], [33]. However, the quadratic energy functional is more easily affected by noise and outliers. To address this problem, Yang et al. [40] propose a sparse non-rigid registration (SNR) method with an \( \ell_1 \)-norm regularized model for the transformation constraint. However, their position constraint is still based on the \( \ell_2 \)-norm. In practice, e.g. for near piece-wise rigid deformation, which is common for real-world deformable objects, the positional error tends to concentrate on small regions. This cannot be modeled well using the \( \ell_2 \)-norm.

In this paper, we propose a non-rigid registration method with sparsity-regularized position and transformation constraints. The distribution of positional errors and transformation differences for typical non-rigid deformation can be well modeled using the Laplacian distribution, or equivalently, the \( \ell_1 \)-norm should be used to measure both the positional errors and transformation differences. To promote the sparsity, we adopt a reweighted sparse model, which is solved by the alternating direction method of multipliers (ADMM). The proposed method is evaluated on public datasets [10], [38] and real datasets captured by a RGB-D depth sensor. The results demonstrate that the proposed method obtains better results than the state-of-the-art non-rigid registration methods.

The main contributions of this work are summarized as:

- We propose a non-rigid registration method on both position and transformation sparsity. The proposed model is robust against outliers as the sparsity terms allow a small fraction of regions with larger deviations.
- We incorporate orthogonality constraints in the sparsity-inducing non-rigid registration framework to promote locally rigid transformations.
- We equip the proposed non-rigid registration model with a reweighted scheme to iteratively enhance sparsity in the series of alternating optimization subproblems.

2 RELATED WORK

3-D shape registration consists of rigid registration and non-rigid registration. Rigid registration aims to find a global rigid-body transformation, while non-rigid registration needs to find a set of local transformations that align two shapes.

In rigid registration, the 3-D shapes are assumed to be aligned by a Euclidean transformation, including rotation and translation. Iterative Closest Point (ICP) and its variants [5] are the dominant algorithms for rigid registration. This kind of methods alternates between two steps: 1) finding closest points and 2) solving the optimal transformation. As an improved method of ICP, Chen et
al. [12] minimize the shortest distance between a point in the template and the tangent plane of the closest point on the target. Pottmann et al. [28] propose a registration method with quadratic convergence, which gives faster and more stable convergence than the standard ICP [27]. Bouaziz et al. [8] propose a new variant of the ICP algorithm, which uses sparsity-inducing norms to represent the positional constraint and achieves better results for the situation with noise and outliers. Their work focuses on rigid registration with low degrees of freedom, and hence regularization is not necessary.

When shapes have large deformations from template to target, automatic non-rigid registration is necessary. It is more challenging due to its high degrees of freedom, and an appropriate deformation model is the key for an efficient and robust algorithm.

Some methods compute global rigid transformations for bones and local non-rigid transformations near joints, which is essentially a piecewise rigid transformation model. Allen et al. [1] place markers on the object to help reconstruct the pose of scan and use it as a basis for modeling deformation. Pekelny et al. [26] use predefined bone information to find bone transformations.

Some models take more generic deformations into consideration. Chui et al. [14] use the thin-plate spline (TPS) as the non-rigid transformation model. Papazov et al. [25] allow points to move freely and use an additional uniform distribution to limit noise and outliers, and propose an ordinary differential equation (ODE) model. Local affine transformations [2] are also frequently used in non-rigid registration. Liao et al. [22] use differential coordinates as local affine transformations with smoothness constraints. Amberg et al. [3] use a stiffness term to ensure similarity of adjacent transformations. Rouhani et al. [29] model non-rigid deformation as an integration of locally rigid transformations. In our work, we use local affine transformations with an orthogonality constraint as it allows more flexibility to capture fine surface details while keeping local shapes.

Non-rigid registration is often formulated as an energy functional with data and regularization terms. Most of the non-rigid registration work models the data term in the \( \ell_2 \)-norm in a least-squares sense [34], [3].

Regularization terms help to preserve smoothness, making the optimization more robust to noise and outliers, and \( \ell_2 \)-norm is also widely used in regularization terms. Sühnuth et al. [35] use a generalized as-rigid-as-possible energy [33] to promote smoothness. Liao et al. [22] define a transformation model using the TPS [14], and use graduated assignment for non-rigid registration and optimization. Wand et al. [39] take a set of time-varying point data as input, and reconstruct a single shape and a deformation field that fit the data. To improve robustness, Li et al. [21] solve correspondences, confidence weights, and a deformation field within a single optimization framework using \( \ell_2 \)-norm. Their method however requires adjacent frames to be sufficiently close to work effectively. Hontani et al. [18] propose a statistical shape model (SSM) which is incorporated into the nonrigid ICP (NICP), and outliers can be detected based on their sparseness. Based on the observation that many deformable objects, in particular human bodies, have near articulated motions, Guo et al. [17] introduce \( \ell_0 \) regularization for motions which provide more accurate and robust tracking in dynamic 3D reconstruction. However, since their method is based on a tracking pipeline, adjacent frames are required to have high similarity. Moreover, their sparse regularization is only applied to motions. Yang et al. [40] propose a sparse non-rigid registration (SNR) method with an \( \ell_2 \)-norm regularized model for the smoothness. However, their \( \ell_2 \)-norm position constraint cannot model the concentration of positional errors well.

Non-rigid registration is also related to and often an important component in dynamic 3D (or 4D) reconstruction. Li et al. [20] propose a pioneering solution to dynamic reconstruction from a sequence of depth images captured by a single depth camera. The method produces impressive results but requires to capture the coarse 3D template of the deforming object. It also assumes adjacent scans are reasonably close. More recently, the work [41] achieves real-time reconstruction with GPU acceleration. However, the method still requires the complete template model to be scanned in advance. To reconstruct dynamic 3D deforming objects without a template prior is still challenging, and state-of-the-art techniques such as [15] utilize multi-camera systems (24 cameras producing 8 depth streams are used in [15]) to achieve real-time 4D performance capture. Our work considers general non-rigid registration where scans can have substantial deformation and no template prior is required.
In this paper, based on the observation that the deformations of 3-D surfaces vary smoothly and the positional distances and transformation differences are sparse, we propose a non-rigid registration method with sparse position and transformation constraints. The model is efficiently solved by the alternating direction method under the augmented Lagrangian multiplier framework.

3 Motivation

The data terms in previous shape registration models [17], [21], [40] are quadratic, which implicitly assumes the Gaussian distribution of positional errors. However, transformations in certain common scenarios, such as articulated motion of humans, are largely piecewise smooth signals residing on 3D surfaces, resulting in larger positional errors for geometric details and joints and smaller errors for the remaining surfaces. This suggests that the positional errors are sparse, and should be modeled by a heavy-tailed distribution, rather than being dense and modeled by a rapidly vanishing Gaussian distribution. This is verified in Fig. 1(b). We uniformly pick up 10% ground truth matchings (vertices) as correspondences, and solve for the transformations using the SNR method [40] which measures the positional errors in the standard quadratic term to avoid bias towards the \( \ell_1 \)-norm. The Laplacian distribution fits the histogram of positional errors significantly better than the Gaussian distribution, suggesting the use of sparsity-promoting \( \ell_1 \)-norm in the data term.

Let \( \mathbf{d}_i = (\tilde{x}_i, \tilde{y}_i, \tilde{z}_i) \) be the difference between the \( i \)th transformed template vertex position and the position of its corresponding vertex of coordinate differences in each dimension for all the corresponding vertices (i.e., \( E_1 = \sum_i \| \mathbf{d}_i \|_1 = \sum_i (|\tilde{x}_i| + |\tilde{y}_i| + |\tilde{z}_i|) \)). Another possibility is to use the sum of Euclidean distances (group sparsity) between corresponding points (i.e., \( E_g = \sum_i \| \mathbf{d}_i \|_2 = \sum_i \sqrt{\tilde{x}_i^2 + \tilde{y}_i^2 + \tilde{z}_i^2} \)), which also well fits the distribution of positional errors as shown in Fig. 1(c). The group sparsity advocates sparsity for each Euclidean distance as a whole, while the \( \ell_1 \)-norm allows a large distance along a particular dimension. In this sense, \( \ell_1 \)-norm is more flexible to preserve large non-rigid deformations along some dimensions. Such an advantage is also observed in the anisotropic total variation (TV) [16] that applies the \( \ell_1 \)-norm to the image gradient over the isotropic TV [30] that measures TV as the sum of \( \ell_2 \)-norm (not squared). Birkholz [6] showed that anisotropic TV achieves better denoising performance in preserving the geometries of corners in images. We choose the \( \ell_1 \)-norm to measure the positional errors for its potential flexibility, and also for its easier and faster implementation with an element-wise shrinkage (cf. Table 5 for statistics of running times).

In this paper, we make the assumption that the surfaces to be registered undergo transformations which are near piecewise smooth. This covers a broad range of practical scenarios ranging from common (near) articulated deformations such as human bodies to certain non-articulated deformations such as facial expressions. In such cases, substantial changes of transformations or large registration errors occur in relatively local areas. Note that our model does not require such consistency to satisfy entirely, and can well cope with situations such as muscle bulge, change of local shapes at joints, etc. Our assumptions also fit well with inaccurate correspondences and sparse noise/outliers, as they also induce sparse distributions of errors.

4 The Proposed Method

4.1 Iterative Framework

We iteratively compute the deformation between the template shape and the target shape. Each iteration consists of two steps. In the first step, the correspondences between template and target are estimated using the registration result obtained from the last iteration. At the beginning of the iteration, we use a technique based on local geometric similarity and diffusion pruning of inconsistent correspondences [36] as it often provides reliable correspondences. Alternative correspondence techniques or manual specification of a few correspondences may instead be used (an example is shown in Fig. 5). These computed correspondences are used to initialize the correspondence mapping, referred to as \( \mathbf{f} \). Then, during the iterative process, we update \( \mathbf{f} \) by using the closest points between template and target shapes to find additional correspondences similar to ICP. In the second step (Sec. 4.2), we propose an energy-minimization approach based on double sparsity representation to estimate the non-rigid transformations using the correspondences obtained from the first step.

4.2 Deformation Estimation

Let \( \mathbf{v}_i \triangleq [x_i, y_i, z_i, 1]^T \) be a 3D point in the homogenous coordinates. Denote by \( \mathcal{V} \triangleq \{ \mathbf{v}_1, \cdots, \mathbf{v}_N \} \) a template set of 3D points and by \( \mathcal{U} \triangleq \{ \mathbf{u}_1, \cdots, \mathbf{u}_M \} \) a target set of 3D points, where \( N \) and \( M \) are the numbers of points. Denote by \( \mathbf{u}_{f(i)} \in \mathcal{U} \) the correspondence of \( \mathbf{v}_i \in \mathcal{V} \). Define \( f : \{ 1, \cdots, N \} \mapsto \{ 0, 1, \cdots, M \} \) as the index mapping from the template points to the target points, where \( f(i) = 0 \) means the corresponding vertex cannot be found for the \( i \)th vertex. Denote by \( \mathbf{X}_f \) the \( 3 \times 4 \) transformation matrix for point \( \mathbf{v}_i \). Define \( X \triangleq \{ \mathbf{X}_1, \cdots, \mathbf{X}_N \} \) as the set of non-rigid transformations. For compact notation, we define \( \mathbf{X} \triangleq [\mathbf{X}_1, \cdots, \mathbf{X}_N]^T \) as a matrix containing the \( N \) transformation matrices to be solved. The proposed method is to find non-rigid transformations \( \mathbf{X} \) that transform the template \( \mathcal{V} \) into the target \( \mathcal{U} \) as accurately as possible, given a correspondence mapping \( f \).

The non-rigid registration is formulated as the minimization of the following energy function:

\[
E(\mathbf{X}; f) = E_{\text{data}}(\mathbf{X}; f) + \alpha E_{\text{smooth}}(\mathbf{X}) + \beta E_{\text{orth}}(\mathbf{X}),
\]

where \( E_{\text{data}}(\mathbf{X}), E_{\text{smooth}}(\mathbf{X}) \) and \( E_{\text{orth}}(\mathbf{X}) \) are data term, smoothness term, and orthogonality constraint, respectively. \( \alpha \) and \( \beta \) adjust the importance of different terms. The data term measures the position accuracy, the smoothness term imposes a smoothness constraint so that the original ill-posed problem (defined by only the data term) is now well-posed, and the orthogonality constraint promotes locally rigid transformations, which is particularly needed for underconstrained scenarios such as partial meshes.

Data term: We measure the accuracy of deformation as the closeness of the transformed points to their corresponding target points. We assign a weight, denoted by \( w_i \), for each point. The weight \( w_i \) is one if there is a corresponding point on the target shape for \( \mathbf{v}_i \), and zero otherwise. Hence, we propose the following data term

\[
E_{\text{data}}(\mathbf{X}; f) = \sum_{\mathbf{v}_i \in \mathcal{V}} w_i \| \mathbf{X}_f \mathbf{v}_i - \hat{\mathbf{u}}_{f(i)} \|_1,
\]

where \( \hat{\mathbf{u}}_{f(i)} \) is the Cartesian coordinate of \( \mathbf{u}_{f(i)} \).
Therefore, the cost of transformation smoothness is rewritten as

\[
E_{\text{data}}(X; f) = \|W \left( VX - \tilde{U} \right) \|_1.
\]

**Smoothness term:** In the smoothness term, local rigidity is assumed: for vertex \( v_i \), the transformations of neighboring vertices \( v_j \in N_i \) should have very close transformed positions when applied to \( v_i \). Therefore, we define the following smoothness term:

\[
E_{\text{smooth}}(X) = \sum_{v_i \in V} \sum_{v_j \in N_i} \|X_i v_i - X_j v_j\|_1.
\]

Define a graph \( G \triangleq (V, E) \), where the vertices of the graph are the 3D points in \( V \), and the edges of the graph are denoted by \( E \). For a 3D mesh, edges of the graph are simply defined by connecting each vertex with its \( K \)-nearest neighbors (\( K \) is typically set to \( 6 \)). Denote the neighborhood of vertex \( v_i \) by \( N_i \), and an edge \( e_{ij} \) is defined between each neighboring vertex \( v_j \) and \( v_i \). So, we have \( E = \{ e_{ij} \mid v_j \in N_i, v_i \in V \} \). To further the data term, we define a differential matrix \( K \in \{-1, 1\}^{\|E\times|V|} \) on the graph \( G \) for concise presentation. Concretely, each row of \( K \) corresponds to an edge in \( E \) and each column corresponds to a vertex in \( V \). Each row in \( K \) has only two nonzero entries. For example, assuming the \( r \)th row in \( K \) is associated with edge \( e_{ij} \), then the entry related to the reference vertex \( v_i \) is set at \( 1 \), while the one related to the neighboring vertex \( v_j \) is set at \(-1\), i.e., \( k_{ri} = 1 \) and \( k_{rj} = -1 \). Let \( K_r \) denote the \( r \)th row of \( K \). We introduce a matrix \( B \in R^{\|E\times|V|} \), where the \( i \)th row of \( B \) is defined as \( b_i := k_i \circ v_i^T \) and \( \circ \) denotes the operator of Kronecker product. Therefore, the cost of transformation smoothness is rewritten as

\[
E_{\text{smooth}}(X) = \|BX\|_1.
\]

**Orthogonality constraint:** Especially for partial meshes with large motions, the problem may be underconstrained leading to large distortions. In this case, the orthogonality constraint as defined below is effective in better preserving local shapes and making the solution more reasonable.

\[
E_{\text{orth}}(X; R_i) = \sum_{i=1}^{N} \|SX_i - R_i\|_F^2,
\]

s.t. \( R_i^T R_i = I \), \( \det(R_i) > 0 \),

where \( R_i \) is a \( 3 \times 3 \) rotation matrix, and \( S = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \) is a constant \( 3 \times 4 \) matrix that extracts the rotation component of \( X_i \), \( \det(R_i) > 0 \) ensures that \( R_i \) is a rotation matrix, not a reflection matrix.

The final energy function has the following compact form with matrix-vector notations:

\[
\begin{aligned}
\min_{X; R_i} & \left\| W \left( VX - \tilde{U} \right) \right\|_1 + \alpha \|BX\|_1 + \beta \sum_{i=1}^{N} \|SX_i - R_i\|_F^2, \\
\text{s.t.} & \quad R_i^T R_i = I, \quad \det(R_i) > 0.
\end{aligned}
\]

**Reweighting:** In a sparse representation, the gap between the convex \( \ell_1 \)-norm and the nonconvex \( \ell_0 \)-norm in measuring sparsity could be filled by reweighting the \( \ell_1 \)-norm [11]. To further promote sparsity, both the data term and the smoothness term are weighted, and the weighting matrices are updated at each iteration of non-rigid registration. The weighted version of the double sparsity model (8) is defined as follows:

\[
\begin{aligned}
\min_{X, R_i, R_j} & \left\| W_D \left( VX - \tilde{U} \right) \right\|_1 + \alpha \|W_S BX\|_1 + \beta \sum_{i=1}^{N} \|SX_i - R_i\|_F^2, \\
\text{s.t.} & \quad R_i^T R_i = I, \quad \det(R_i) > 0.
\end{aligned}
\]

where \( W_D \) and \( W_S \) are diagonal weighting matrices for the data term and smoothness term, respectively. The weighting matrices are updated according to the \( \ell_1 \)-norm of the corresponding entries. For the data term, the weights are updated as

\[
W_D^{(l)}(i, i) = \begin{cases} \frac{1}{\|X_{(l-1)i} - X_{(l-1)i}^s\|_1 + \epsilon_D}, & f(i) \neq 0, \\
0, & f(i) = 0, \end{cases}
\]

where \( l \) represents the index of iteration, \( \epsilon_D \) is a constant to avoid the division-by-zero issue, and is set as \( 0.01 \) in the experiments. Similarly, the weights for the smoothness term are updated as

\[
W_S^{(l)}(i, i) = \frac{1}{\|X_i^{(l-1)} - X_j^{(l-1)}\|_1 + \epsilon_S},
\]

where \( \epsilon_S \) is a constant which is set as \( 0.01 \) in the experiments, and the \( p \)th row of matrix \( BX \) is associated with edge \( e_{ij} \) between \( v_i \) and \( v_j \). The reweighting scheme is incorporated into the iterative registration framework, which only slightly increases the computation to calculate the weights. The reweighted \( \ell_1 \)-norm is also related to robust kernels [41] in suppressing the influence of outliers, although we propose a different formulation that works well in our sparse non-rigid registration framework.
Figure 2. (a) Template (top) and target (bottom) shapes, (b)-(d): Comparison results (top) and registration errors (bottom) of (b) $\ell_2$-norm method, (c) SNR method [40] and (d) Our method on Cat dataset.

Figure 3. (a) Template (top) and target (bottom) shapes, (b)-(d): Comparison results (top) and registration errors (bottom) of (b) $\ell_2$-norm method, (c) SNR method [40] and (d) Our method on Jumping dataset.

Figure 4. Comparison results on Bouncing dataset: (a) Template and target, (b) The method in [21], (c) SNR method [40], and (d) Our method.

Figure 5. Comparison results on Jumping dataset with 35 manually-specified correspondences: (a) Given correspondences, (b) $\ell_2$-norm method, (c) SNR method [40], and (d) Our method

5 Experimental Results

In this section, we evaluate the performance of the proposed method on clean datasets (Section 5.1), noisy datasets (Section 5.2), and real scans (Section 5.3). Running times of our method are reported in Section 5.4. All the experiments are performed on a desktop computer with an Intel i5 3.2GHz CPU and 8GB RAM. The numbers of inner and outer iterations of our method are both set as 20.

5.1 Results on Clean Datasets

We evaluate the proposed method on two datasets: TOSCA high-resolution dataset [10] and a human motion dataset [38]. Fig. 2 and Fig. 3 give the registration results on a particular pair of cat and jumping datasets, compared with the classic $\ell_2$-norm regularized non-rigid ICP method and the SNR method [40]. The results are shown as the overlap of the deformed template shape (blue) and the target shape (gray) and the registration errors are color-coded on the reconstructed mesh for visual inspection. Denote $g_i$ as the ground-truth correspondence of $v_i$. For a vertex $v_i$, the registration error is defined as $\|X_i v_i - g_i\|_2^2$. The compared classic $\ell_2$-norm based non-rigid ICP method [7] is formulated as
optimizing:

$$\min_{X} \| W(VX - \hat{U}_f) \|_F^2 + \alpha \| BX \|_F^2. \quad (14)$$

The smoothness constraint of this kind of methods is imposed on the transformation differences. To ensure fair comparison, we adjust the weight $\alpha$ until we get the most accurate registration without loss of smoothness for each method. The result shows that our method achieves the best results with less registration errors in the areas with intensive deformations than the SNR method [40] and the classic $\ell_2$-norm regularized non-rigid ICP method, such as the tail of the cat and the wrinkles around the waist of the person highlighted in rectangles.

We compare our method with a state-of-the-art non-rigid registration method [21] in Fig. 4. Obvious registration errors can be seen in the result of the method in [21], especially in the right foot (top) and head (bottom), while the methods with
Figure 9. Comparison results with and without reweighting scheme on Bouncing dataset: (a) Template, (b) Target, (c) Registration result without reweighting scheme, (d) Registration result with reweighting scheme, (e) Registration errors without reweighting scheme, and (f) Registration errors with reweighting scheme.

Figure 10. (a) Template (left) and target (right) shapes, (b)-(d): Comparison results (left) and registration errors (right) of (b) $\ell_2$-norm method, (c) SNR method [40] and (d) our method on Bouncing whole-to-part dataset with frames 1, 21, 51, 84 as templates and frames 87, 107, 137, 171 as targets.

Figure 8. Fraction of correspondences within the error threshold. The graph shows the fraction of correspondences (y-axis) within the error threshold (x-axis).

sparse representation (SNR [40] and our method) achieve better registration results. The method in [21] works effectively when the template and target shapes are close so that good initial correspondences can be obtained, but the pose changes substantially in this example. Moreover, our result is more accurate and better-distributed for the whole body than the SNR method [40], due to the sparse constraint on the position.

To evaluate the robustness of the proposed method, we manually assign 35 correspondences on Jumping dataset, and compare the result of our method with the SNR method [40] and the $\ell_2$-regularized method. As shown in Fig. 5, our method achieves the best result, especially around the places with substantial deformation, e.g., the right knee.

We also compare our method with a state-of-the-art method [17] that uses $\ell_0$ norm in Fig. 6. The code from the authors is used. The original method [17] uses ICP correspondences. When registering scans with large deformations, correspondences...
derived from intrinsic geometric properties can be more effective. To ensure fair comparison, we compare our method with two versions of the \( \ell_0 \)-norm method: ICP + \( \ell_0 \)-norm method that uses ICP to compute correspondences, and Diffusion pruning (DP) + \( \ell_0 \)-norm method that computes correspondences using [36] as initialization like our method. Fig. 6 gives the registration results for three different degrees of deformation (increasing from top to bottom). It can be seen that with moderate deformation (top row), both variants of [17] work reasonably well. However, when the deformation is large, the method fails to align the two surfaces, resulting in large errors. On the contrary, our method achieves more accurate results than the other methods. The quantitative
Figure 13. Comparison results on Jumping dataset with partially incorrect correspondences: (a) Template and target, (b) SNR method [40] result with one third SHOT correspondences, (c) Our method result with one third SHOT correspondences, (d) SNR method [40] result with all SHOT correspondences, and (e) Our method result with all SHOT correspondences.

Figure 14. Comparison results on Bouncing with noise ($\sigma = 0.3, 0.7, 1$). (a) Template and target, (b) Curves of registration errors vs. normalized noise levels, (c) Target with noise, (d) $\ell_2$-norm method, (e) SNR method [40], and (f) Our method.

evaluation (the mean of registration errors over all the vertices) is shown in Table 1. While diffusion pruning helps reduce the errors of [17] in most cases, our method has significantly smaller errors than both variants and the SNR method [40].

Table 1
Quantitative evaluation for Fig. 6.

<table>
<thead>
<tr>
<th>Deformation</th>
<th>Mean Error</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ICP + $\ell_0$-norm</td>
</tr>
<tr>
<td>Small</td>
<td>0.0148</td>
</tr>
<tr>
<td>Median</td>
<td>0.1707</td>
</tr>
<tr>
<td>Large</td>
<td>0.2011</td>
</tr>
</tbody>
</table>

To more comprehensively evaluate the proposed method, we test our method on five sets of shapes from the TOSCA high-resolution dataset [10], compared with the Diffusion pruning (DP) + $\ell_0$-norm method [17] and the SNR method [40]. The initial correspondences are uniformly selected 5% ground-truth correspondences. We test every pair of models in each set (treating one as template and the other as target), and the quantitative evaluation (the mean of registration errors over all the vertices for all the models) is shown in Table 2. Some examples are given in Fig. 7. The average errors are shown in the subfigures. Fig. 8 shows the fraction of correspondences (y-axis) within the error threshold (x-axis) [19]. Our method (green curve) detects nearly 100% correct correspondences for a small threshold of 0.05. The results show that our method achieves the most accurate and robust non-rigid registration.

To evaluate the effectiveness of the proposed reweighting scheme, we compare the registration results with and without reweighting on Bouncing dataset in Fig. 9. The parameters $\epsilon_D$ and $\epsilon_S$ are set as 0.01. As shown in the figure, the reweighting scheme significantly improves the registration results.

We also evaluate our method on the harder whole-to-part and part-to-part registration problems. Since the 3D models in the
Figure 15. Comparison results on Bouncing with 1%, 2%, 5% outliers. (a) Template and target, (b) Curves of registration errors vs. normalized noise levels, (c) Target with noise, (d) $\ell_2$-norm method, (e) SNR method [40], and (f) Our method.

Figure 16. Comparison results with and without reweighting scheme on Bouncing dataset with noise ($\sigma = 1$): (a) Template, (b) Target, (c) Registration result without reweighting scheme, (d) Registration result with reweighting scheme, (e) Registration errors without reweighting scheme, and (f) Registration errors with reweighting scheme.

Figure 17. Comparison results with and without reweighting scheme on Bouncing dataset with 50% outliers: (a) Template, (b) Target, (c) Registration result without reweighting scheme, (d) Registration result with reweighting scheme, (e) Registration errors without reweighting scheme, and (f) Registration errors with reweighting scheme.

Table 2
Quantitative evaluation for five sets of shapes from the TOSCA dataset.

<table>
<thead>
<tr>
<th>Method</th>
<th>Cat</th>
<th>Centaur</th>
<th>Gorilla</th>
<th>Horse</th>
<th>Wolf</th>
</tr>
</thead>
<tbody>
<tr>
<td>DP + $\ell_0$-norm</td>
<td>2.6366</td>
<td>0.7311</td>
<td>22.1740</td>
<td>6.6398</td>
<td>0.0545</td>
</tr>
<tr>
<td>SNR</td>
<td>0.0902</td>
<td>0.1799</td>
<td>5.364</td>
<td>0.0865</td>
<td>0.0359</td>
</tr>
<tr>
<td>Ours</td>
<td><strong>0.0090</strong></td>
<td><strong>0.0297</strong></td>
<td><strong>0.5315</strong></td>
<td><strong>0.0089</strong></td>
<td><strong>0.0012</strong></td>
</tr>
</tbody>
</table>

public dataset are complete, we obtain partial models by extracting the visible part of each complete model with a virtual depth camera rotating around the model, while keeping the ground truth correspondences. Fig. 10 and Fig. 11 give the whole-to-part and part-to-part registration results on the Bouncing dataset which has a total of 172 models. We use a systematic approach that takes the models at frames 1-86 as templates and the models at frames 87-172 as targets such that frame $t$ is registered to frame $t + 86$. We have compared the methods on all the models in this dataset, and Fig. 10 and Fig. 11 show the registration results for 4 frames. The mean of registration errors over all the vertices in the overlapping regions for all the models in the entire dataset of $\ell_2$-norm method, SNR method [40] and our method in the whole-to-part case are 0.0323, 0.0324 and 0.0262, respectively. The mean of registration errors over all the vertices in the overlapping regions for all the models in the entire dataset of $\ell_2$-norm method, SNR method [40] and our method in the part-to-part case are 0.0380, 0.0381 and 0.0318, respectively. It can be seen that our method achieves the best registration with the smallest errors among these methods. For part-to-part registration, our method reduces errors by more than a half, compared with state-of-the-art sparse non-rigid registration.
method [40]. We also compare with the \( \ell_0 \)-norm method [17] for the whole-to-part case in Fig. 12. The quantitative evaluation (the mean of registration errors over the whole model) is shown in Table 3. The \( \ell_0 \)-norm method [17] has similar limitations to handle large deformations, while our method gives more accurate and robust results.

<table>
<thead>
<tr>
<th>Deformation</th>
<th>Mean Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICP + ( \ell_0 \text{-norm} )</td>
<td>DP + ( \ell_0 \text{-norm} )</td>
</tr>
<tr>
<td>Small</td>
<td>0.0279</td>
</tr>
<tr>
<td>Median</td>
<td>0.0413</td>
</tr>
<tr>
<td>Large</td>
<td>0.3390</td>
</tr>
</tbody>
</table>

5.2 Results on Noisy Datasets

1) Correspondences with partially incorrect matchings:

It is common to include incorrect correspondences using established methods. We simulate this in two cases. In the first case, we obtain two thirds of correspondences using diffusion pruning [36] and the remaining one third using local geometric feature matching based on SHOT signatures [31]. The majority of correspondences from the former are correct while many correspondences from the latter are incorrect due to the ambiguity of local features. In the second case, we generate all the correspondences using SHOT signatures. Fig. 13 gives the results for the two cases in a difficult situation which involves very complex transformations from template to target. As shown in the figure, our method is significantly more robust than the SNR method [40] with respect to incorrect correspondences. The mean of registration errors over all the vertices for the four cases in Fig. 13 (b–e) are 0.038, 0.012, 0.047, and 0.032, respectively.

2) Target shapes with noise or outliers:

In the first case, 3-D shapes of targets are polluted with dense noise along the normal directions of the associated vertices. All the target vertices are perturbed with Gaussian noise. The standard deviation of the noise \( \sigma \) is normalized by \( \bar{l} \), where \( \bar{l} \) is the average length of triangle edges on the associated target mesh, and chosen in the range of \([0.1, 1]\). Fig. 14 gives the registration results compared with the SNR method [40] and the \( \ell_2 \)-norm regularization method. The results show that our method is more robust to noise, performing significantly better for models with high noise levels.

In the second case, 3-D shapes of targets are polluted with sparse outliers along the normal directions of the associated vertices. Fig. 15 gives the results for the situations when 1%, 2%, 5% of target vertices are perturbed with Gaussian noise. The results show that our method is more robust than the other two methods, particularly for cases with larger proportion of outliers.

To evaluate the effectiveness of the proposed reweighting scheme, we also compare the registration results with and without reweighting for noise and outlier cases on Bouncing dataset in Fig. 16 and Fig. 17. The parameters \( \epsilon_D \) and \( \epsilon_S \) are set as 0.01. The standard deviation of the noise \( \sigma \) is set as 1, and the percentage of outliers is set as 50%. It can be seen that the reweighting scheme contributes significantly to improving the registration results for the dataset with noise and outliers.

We compare the registration results with different parameter settings for the reweighting scheme on Bouncing dataset with 50% outliers in Fig. 18 to evaluate the influence of the parameters \( \epsilon_D \) and \( \epsilon_S \). To make experiments more tractable, we adjust both parameters consistently (i.e. \( \epsilon_D = \epsilon_S = \epsilon \)). It can be seen that the best setting is 0.006 for this case, which has the smallest registration errors. However, the performance is quite close, and 0.01 is a generally good choice (found in experiments).

5.3 Results on Real Scans

Fig. 19 presents the results on real scans generated by Kinect Fusion [23] using Kinect V2.0. The real scans are very challenging, because they have much noise and a large number of outliers. Moreover, each mesh is incomplete and the topology between the template and the target is inconsistent. Hence, it is difficult to obtain sufficient and reliable correspondences. The overlap of the deformed template and the target shows that the \( \ell_2 \)-norm regularization method and the SNR method present misalignments around the hands, arms and some other joints which have large deformations, while the result of our method is well-distributed and better registered.

We also compare our method with the \( \ell_0 \)-norm method [17] on the Kongfu dataset [17] in Fig. 20, where pairs of (non-adjacent)
frames with increasing degree of deformation are used as input. Our method clearly outperforms both variants of [17], especially for the hands where significant movements exist between scans. The quantitative evaluation (the mean of registration errors over the whole model) is shown in Table 4. Our method has the smallest errors.

Table 4
Quantitative evaluation for Fig. 20.

<table>
<thead>
<tr>
<th>Deformation</th>
<th>Mean Error</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ICP + $\ell_0$-norm</td>
</tr>
<tr>
<td>Small</td>
<td>0.0035</td>
</tr>
<tr>
<td>Large</td>
<td>0.0039</td>
</tr>
</tbody>
</table>

Fig. 21 gives an example of generating a complete color mesh for a human head. A base mesh is scanned by Kinect Fusion using Kinect V2.0, and four partial color meshes are registered to the base mesh using our method. The textures are blended by solving the Poisson equation over the surface of mesh [13]. As shown in the figure, our method correctly registers the input view surfaces with better registration than alternative methods, and successfully generates a watertight color mesh.

5.4 Running times
We compare the running times of the proposed method with the $\ell_2$-norm regularized method, SNR method, and group sparsity method on Crane dataset. We downsample the meshes into smaller meshes with 1K to 10K vertices. The number of NICP registration iterations for each method is set as 20. The comparison results are shown in Table 5. Our method has similar time complexity as SNR.

<table>
<thead>
<tr>
<th>Num. vertices</th>
<th>$\ell_2$-norm</th>
<th>SNR</th>
<th>Group sparsity</th>
<th>Ours</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000</td>
<td>1.23s</td>
<td>8.05s</td>
<td>7.39s</td>
<td>7.17s</td>
</tr>
<tr>
<td>2000</td>
<td>3.51s</td>
<td>17.36s</td>
<td>24.83s</td>
<td>22.13s</td>
</tr>
<tr>
<td>5000</td>
<td>12.88s</td>
<td>52.48s</td>
<td>59.96s</td>
<td>55.68s</td>
</tr>
<tr>
<td>10000</td>
<td>29.78s</td>
<td>119.06s</td>
<td>126.58s</td>
<td>122.85s</td>
</tr>
</tbody>
</table>

6 Conclusions
This paper proposes a non-rigid registration method with reweighted sparse position and transformation constraints. We formulate the energy function with position and transformation sparsity on both the data term and the smoothness term, and define the smoothness constraint using local rigidity. The double sparsity based non-rigid registration model is equipped with a reweighting scheme, and solved by the alternating direction method under the augmented Lagrangian multiplier framework which has exact solutions and guaranteed convergence. Experimental results on both public datasets and real scans show that our method provides significantly improved results over alternative methods, especially for more challenging cases, and is more robust to noise and outliers.

Acknowledgements
The authors thank Ke Li for her help with some experiments, and thank Shuai Lin for help with comparative experiments with [21]. The authors are also grateful to the Associate Editor and anonymous reviews for their helps in improving this paper.
This work was supported in part by the National Natural Science Foundation of China (Grant 61571322 and Grant 61771339).

**APPENDIX A**
**DERIVATION OF THE ADMM ALGORITHM**

Under the ADMM framework, the augmented Lagrangian function (13) is optimized with respect to the variables alternately, yielding the following subproblems to optimize:

\[
C^{(k+1)} = \arg\min_C \|C\|_1 + \frac{\mu_1}{2} \|C - W_D (VX^{(k)} - \bar{U}_f)\|_F^2,
\]

\[
A^{(k+1)} = \arg\min_A \frac{\mu_2}{2} \|A - W_S B X^{(k+1)}\|_F^2 + \beta \|\sum_{i=1}^N S_i^T R_i^{(k+1)}\|_F^2,
\]

\[
R_i^{(k+1)} = \arg\min_{R_i} \beta \|S X_i^{(k)} - R_i^{(k+1)}\|_F^2 \quad \text{s.t.} \quad R_i^T R_i = I, \det(R_i) > 0,
\]

\[
X^{(k+1)} = \arg\min_X \|Y_1^{(k)} + (C^{(k+1)} - W_D (VX^{(k+1)} - \bar{U}_f))\|_F^2 + \frac{\mu_1}{2} \|C^{(k+1)} - W_D (VX^{(k+1)} - \bar{U}_f)\|_F^2 + \frac{\mu_2}{2} \|A^{(k+1)} - W_S B X^{(k+1)}\|_F^2 + \beta \|\sum_{i=1}^N S_i^T R_i^{(k+1)}\|_F^2.
\]

The C-subproblem has the following closed solution:

\[
C^{(k+1)} = \text{shrink}\left(W_D (VX^{(k)} - \bar{U}_f) - \frac{1}{\mu_1} Y_1^{(k)}, \frac{1}{\mu_1} Y_1^{(k)}\right),
\]

where \(\text{shrink}(\cdot, \tau)\) is the shrinkage function applied on the matrix element-wise:

\[
\text{shrink}(x, \tau) = \text{sign}(x) \max(|x| - \tau, 0).
\]

The A-subproblem is solved in a similar way:

\[
A^{(k+1)} = \text{shrink}\left(W_S B X^{(k)} - \frac{1}{\mu_2} Y_2^{(k)}, \frac{1}{\mu_2} Y_2^{(k)}\right).
\]

The \(R_i\)-subproblem can be explicitly solved using Procrustes projection:

\[
(U, D, V^T) = \text{svd}(S X_i^{(k)}),
\]

\[
R_i^{(k+1)} = U V^T,
\]

where \(\text{svd}(\cdot)\) is the singular value decomposition. If the obtained matrix has a negative determinant, take \(R_i\) with the opposite sign to turn the matrix into a rotation matrix. This step is similar to [33] for minimizing as-rigid-as-possible energy, although our overall alternating optimization is different and more complicated.

Being quadratic, the \(X\)-subproblem can be readily solved by using the first-order optimality condition:

\[
\begin{aligned}
&\left(\mu_1^{(k)} V^T W_D^T W D V + \mu_2^{(k)} B^T W_S^T W S B + \beta \sum_{i=1}^N S_i^T S_i\right) \hat{X} \\
= &B^T W_S^T \left(Y_1^{(k)} + \mu_1^{(k)} (C^{(k+1)} - W_D \bar{U}_f)\right) \\
+ &\mu_2^{(k)} A^{(k+1)} + \beta \sum_{i=1}^N S_i^T R_i^{(k+1)}.
\end{aligned}
\]

However, the straightforward matrix inversion in solving (20) is inefficient or even practically impossible for large-scale problems, e.g., registration of tens of thousands of points. This can be relieved by using the LDL decomposition:

\[
(L, D) = \text{ldl}\left(\mu_1^{(k)} V^T W_D^T W D V + \mu_2^{(k)} B^T W_S^T W S B + \beta \sum_{i=1}^N S_i^T S_i\right),
\]

where \(L\) and \(D\) are the lower triangular matrix and the diagonal matrix of the LDL decomposition. Then, the linear equations in (20) is solved by solving the following much easier linear systems:

\[
LQ = V^T W_D^T \left(Y_1^{(k)} + \mu_1^{(k)} (C^{(k+1)} - W_D \bar{U}_f)\right),
\]

\[
B^T W_S^T \left(Y_2^{(k)} + \mu_2^{(k)} A^{(k+1)}\right) + \beta \sum_{i=1}^N S_i^T R_i^{(k+1)}.
\]

\[
DZ = Q,
\]

\[
L^T X = Z.
\]

**REFERENCES**


