GLOBAL AS-CONFORMAL-AS-POSSIBLE NON-RIGID REGISTRATION OF MULTI-VIEW SCANS

Zhenchao Wu [1*], *Kun Li* [1*], *Yu-Kun Lai* [2], *Jingyu Yang* [1]

[1] Tianjin University, Tianjin, China
[2] Cardiff University, Wales, UK

ABSTRACT

In this paper, we present a novel framework for global non-rigid registration of multi-view scans captured using consumer-level depth cameras. In our method, all scans from different viewpoints are allowed to undergo large non-rigid deformations and finally fused into a complete high quality model. To avoid the well-known loop closure problem, we simultaneously optimize a global alignment problem instead of pairwise non-rigid registration in succession. We employ a joint point-to-point and point-to-plane positional constraint to reduce the influence of wrong correspondences, and incorporate an as-conformal-as-possible constraint to avoid mesh distortions during deformation. We also design a reweighting scheme on position and transformation to reduce registration errors. Experimental results on both public datasets and real scanned datasets demonstrate that our approach outperforms state-of-the-art methods through extensive quantitative and qualitative evaluations.

Index Terms— 3D scanning, global non-rigid registration, large deformation, depth cameras, surface reconstruction

1. INTRODUCTION

3D reconstruction has been extensively studied for decades due to its wide applicability in 3D printing, games and movie production [1, 2, 3]. The availability of low-cost commodity depth cameras, such as Microsoft Kinect, has made the reconstruction of a 3D scene easier and cheaper than ever using a laser scanner. However, the captured depth suffers from the pollution of serious noise and outliers, which increases the challenges of 3D reconstruction, especially for dynamic non-rigid reconstruction with larger dimensionality and higher complexity of the deformation space.

Some previous work achieves dynamic 3D reconstruction based on multi-camera systems. Li *et al.* [1] at Tsinghua University build a dome system with 20 cameras to recover shape and capture motion from multi-view video sequences using a separating and merging strategy and a volumetric deformation method. Collet *et al.* [4] obtain a textured high-resolution mesh with different connectivity in each frame by a set of high-speed RGB and IR cameras. Unfortunately, such systems are not readily usable in many real-world applications due to high cost, complex maintenance and lack of portability. The Microsoft Kinect has been extensively employed with the advantage of low-cost and multi-sensing. Tong *et al.* [5] use three Kinects to capture different parts of a human body to generate a convincing 3D human model. Kinect Fusion [2] achieves the reconstruction of a 3D object in real-time using a single Kinect camera but requires the object to keep still. Li *et al.* [6] develop a pipeline in which ordinary users capture complete and fully textured 3D models of themselves using only a Kinect sensor, but the captured body must keep the same pose among various viewpoints. To reconstruct scenes with non-rigid deformations, Newcombe *et al.* [7] present a system that reconstructs non-rigidly deforming subjects in real-time. However, drifting problems exist in both rigid and non-rigid alignment where errors accumulate continuously. To handle this, Zollhöfer *et al.* [3] propose a template-based approach, which starts from a scanned template of the subject and gradually deforms it toward the input depth sequences. To remove the requirement of the template prior, Yu *et al.* [8] propose a method to reconstruct detailed geometry, non-rigid motion and the inner human body shape from a single Kinect camera with a double layer representation consisting of parametric body shapes and gradually fused surfaces. Much effort has been made to reconstruct small deformation between two neighboring viewpoints, but little work in the literature allows large motions of the subjects between different viewpoints. Li *et al.* [9] present a pair-wise non-rigid registration method for large deformation cases, but the loop closure problem tends to occur due to error accumulation. Guo *et al.* [10] propose a global non-rigid registration method to avoid the error accumulation, but the viewpoints need to be reasonably dense to avoid too large differences between neighboring views.

In this paper, we propose a method for global non-rigid registration of multi-view scans captured by a single depth camera, which allows large deformations of the objects between different viewpoints, *i.e.*, sparse viewpoints. To dis-

tribute errors in the deformation space, we compute the deformations of all the scans simultaneously by optimizing a global alignment problem. We employ a joint point-to-point and point-to-plane positional constraint to reduce the influence of wrong correspondences, and incorporate an as-conformal-as-possible (ACAP) constraint to avoid mesh distortions and maintain mesh structures during deformation, especially when the meshes undergo large deformations. Moreover, a reweighting scheme on position and transformation is designed in order to reduce the registration error. Experimental results on both public datasets and real scanned datasets demonstrate that our approach achieves the best reconstruction compared with state-of-the-art methods with quantitative and qualitative evaluations.

To sum up, we contribute a novel global optimization method for registration of multi-view scans with large deformations, which is robust to noise and outliers and avoids the loop closure problem. To achieve this, we use a joint point-to-point and point-to-plane weighting for positional constraint to reduce the influence of wrong correspondences. Moreover, we incorporate an ACAP constraint to avoid mesh distortions and maintain mesh structures by preserving the angles of triangles in the meshes during deformation. We also design a reweighting scheme on position and transformation with exponential functions to better approximate ℓ_0-norm in measuring sparseness, which helps reduce final registration errors.

2. THE PROPOSED METHOD

The aim of global non-rigid registration is to find out a set of non-rigid transformations $\mathbf{X}$ by the correspondences of pairwise scans. Given the transformations $\mathbf{X}$, we transform and align these scans consistently while keeping their shapes plausible. There are two alternating steps in each iterative procedure: (1) Find candidate correspondences in the overlapping areas between pairwise scans after deformation; (2) Solve a set of local affine transformations by minimizing a global energy function. Below we will discuss the energy function of global non-rigid registration in detail.

Given a captured scan sequence $\{\mathcal{U}^{(1)}, \mathcal{U}^{(2)}, \ldots, \mathcal{U}^{(n)}\}$, we compute a mapping between each neighboring pair of scans $\mathcal{U}^{(m)}$ and $\mathcal{U}^{(m+1)}$ (assuming $\mathcal{U}^{(n+1)} = \mathcal{U}^{(1)}$). Denote by $\mathcal{U}^{(m)} \triangleq \left\{\mathbf{u}_1^{(m)}, \mathbf{u}_2^{(m)}, \ldots, \mathbf{u}_{N_m}^{(m)}\right\}$ a set of 3D points in each scan, where $\mathbf{u}_i^{(m)} \triangleq (x_i^{(m)}, y_i^{(m)}, z_i^{(m)}, 1)$ represents the homogeneous coordinates of the ith vertex and N_m represents the number of vertices in scan $\mathcal{U}^{(m)}$. The mapping $f_{m \to m+1} : \{1, \cdots, N_m\} \mapsto \{1, \cdots, N_{m+1}\}$ is established by correspondence computation: $\mathbf{u}_{f_m(i)}^{(m+1)} \in \mathcal{U}^{(m+1)}$ is the correspondence point of $\mathbf{u}_i^{(m)} \in \mathcal{U}^{(m)}$. At the first iteration, we use a diffusion pruning method [11] for correspondence computation as it often provides reliable correspondences. Alternative correspondence techniques or manual specification of a few correspondences may instead be used. At the

other iterations, we update the correspondences by using the closest points between two shapes to find additional correspondences similar to iterative closest point (ICP).

For non-rigid registration, we solve an affine transformation for each point to cover a wide range of non-rigid deformations. Denote the set of non-rigid transformations for scan $\mathcal{U}^{(m)}$ by $\mathbf{X}^{(m)} \triangleq \left\{\mathbf{X}_1^{(m)}, \cdots, \mathbf{X}_{N_m}^{(m)}\right\}$, where $\mathbf{X}_i^{(m)}$ is the 4×3 transformation matrix for point $\mathbf{u}_i^{(m)}$. For convenience, denote by $\mathbf{X}^{(m)} \triangleq \left[\mathbf{X}_1^{(m)}, \cdots, \mathbf{X}_{N_m}^{(m)}\right]^T$ of size $4N_m \times 3$ the ensemble matrix containing N_m transformation matrices to be estimated. The global non-rigid registration is formulated as the minimization of the following energy function:

$$E(\mathbf{X}; f) = E_{data}(\mathbf{X}; f) + \alpha E_{smooth}(\mathbf{X}) + \lambda E_{rig}(\mathbf{X})$$
$$+ \beta E_{arap}(\mathbf{X}) + \gamma E_{acap}(\mathbf{X}), \quad (1)$$

which consists of data term and four regularizers weighted by the factors α, λ, β and γ. $E_{data}(\mathbf{X})$ is the data term to measure the position accuracy, $E_{smooth}(\mathbf{X})$ is the smoothness term to measure the smoothness of local transformations, $E_{rig}(\mathbf{X})$ is the orthogonality term to measure the rigidness of local transformations, $E_{arap}(\mathbf{X})$ is the as-rigid-as-possible term to ensure the length of each edge to be as close as possible before and after transformation and $E_{acap}(\mathbf{X})$ is the as-conformal-as-possible term to preserve angles of triangles in the meshes during deformation.

Data Term: Similar to the pairwise registration, we represent the mapping between a neighboring pair of overlapping scans $\mathcal{U}^{(m)}$ and $\mathcal{U}^{(m+1)}$ by f_m. As neighboring surfaces only have partial overlaps, not all the vertices have correspondences, so we define $\tilde{\mathcal{U}}^{(m)} \subset \mathcal{U}^{(m)}$ as the corresponding point set and suppose the vertex $\mathbf{u}_i^{(m)} \in \tilde{\mathcal{U}}^{(m)}$ has the kth correspondence between the neighboring scans $\mathcal{U}^{(m)}$ and $\mathcal{U}^{(m+1)}$. We design a reweighting scheme to update a weight $\mathbf{W}_{\mathbf{D}k}^{(m)}$ in each iteration, which is mentioned in detail later. Given the correspondence mapping f, the data term is designed to penalize the point-to-point and point-to-plane distances of each neighboring pair of overlapping scans $\mathcal{U}^{(m)}$ and $\mathcal{U}^{(m+1)}$ so that vertex $\mathbf{u}_i^{(m)}$ is forced to move to its corresponding point $\mathbf{u}_{f_m(i)}^{(m+1)}$ especially along the normal direction of $\mathbf{u}_{f_m(i)}^{(m+1)}$:

$$E_{data}(\mathbf{X}; f) \triangleq \sum_m \sum_{\mathbf{u}_i^{(m)} \in \tilde{\mathcal{U}}^{(m)}}$$
$$\mathbf{W}_{\mathbf{D}k}^{(m)}(\alpha_{point} \left\|\mathbf{u}_i^{(m)}\mathbf{X}_i^{(m)} - \mathbf{u}_{f_m(i)}^{(m+1)}\mathbf{X}_{f_m(i)}^{(m+1)}\right\|_1$$
$$+ \alpha_{plane}|\mathbf{n}_{f_m(i)}^{(m+1)}(\mathbf{u}_i^{(m)}\mathbf{X}_i^{(m)} - \mathbf{u}_{f_m(i)}^{(m+1)}\mathbf{X}_{f_m(i)}^{(m+1)})^\top|), \quad (2)$$

where $\|\cdot\|_1$ denotes ℓ_1 norm of a matrix considered as a long vector, and $\mathbf{n}_i^{(m)} \in \mathbb{R}^{1 \times 3}$ represents the normal of vertex $\mathbf{u}_i^{(m)}$. We define $\theta_{f_m(i)}^{(m+1)}$ as the angle between $\mathbf{n}_{f_m(i)}^{(m+1)}$ and

the opposite direction from vertex $\mathbf{u}_i^{(m)}$ to its corresponding point $\mathbf{u}_{f_m(i)}^{(m+1)}$. Let $\mathbf{X} := \left[\mathbf{X}^{(1)}, \dots, \mathbf{X}^{(n)} \right]^T$. For easier calculation, we have the following form of the data term:

$$E_{data}\left(\mathbf{X}; f\right) = \left\| \mathbf{W_D H X} \right\|_1, \tag{3}$$

$$\mathbf{W_D} = \mathrm{diag}\left(\mathbf{W_D}^{(1)}, \dots, \mathbf{W_D}^{(n)} \right),$$

$$\mathbf{W_D}^{(m)} = \mathrm{diag}\left(\mathbf{W_D}_1^{(m)}, \cdots, \mathbf{W_D}_{K_m}^{(m)} \right),$$

$$\mathbf{H} = \begin{bmatrix} \check{\mathbf{U}}^{(1)} & -\hat{\mathbf{U}}^{(2)} & & & \\ & \check{\mathbf{U}}^{(2)} & -\hat{\mathbf{U}}^{(3)} & & \\ & & \ddots & \ddots & \\ & & & \check{\mathbf{U}}^{(n-1)} & -\hat{\mathbf{U}}^{(n)} \\ -\hat{\mathbf{U}}^{(1)} & & & & \check{\mathbf{U}}^{(n)} \end{bmatrix},$$

where $\mathrm{diag}(\cdot)$ is the block-wise diagonal matrix of the input vectors, and K_m represents the number of the vertices in the point set $\tilde{\mathcal{U}}^{(m)}$. $\check{\mathbf{U}}^{(m)}$ and $\hat{\mathbf{U}}^{(m+1)}$ are of size $K_m \times 4N_m$ and $K_m \times 4N_{m+1}$, respectively. The kth row of $\check{\mathbf{U}}^{(m)}$ and $\hat{\mathbf{U}}^{(m+1)}$ are associated with the kth correspondence, with elements $\check{\mathbf{u}}_k^{(m)}$ and $\hat{\mathbf{u}}_k^{(m+1)}$ in corresponding columns, in which $\check{\mathbf{u}}_k^{(m)}$ and $\hat{\mathbf{u}}_k^{(m+1)}$ are defined as

$$\check{\mathbf{u}}_k^{(m)} = (\alpha_{point} + \cos\theta_{f_m(i)}^{(m+1)}\alpha_{plane})\mathbf{u}_i^{(m)},$$

$$\hat{\mathbf{u}}_k^{(m+1)} = (\alpha_{point} + \cos\theta_{f_m(i)}^{(m+1)}\alpha_{plane})\mathbf{u}_{f_m(i)}^{(m+1)}. \tag{4}$$

Smoothness Term: We assume that the vertex $\mathbf{u}_i^{(m)}$ should have very close transformed position with the vertices $\mathbf{u}_j^{(m)}$ in its neighborhood $\mathcal{N}_i^{(m)}$. Then we define $e_{ij}^{(m)}$ as the pth edge of scan $\mathcal{U}^{(m)}$, which is connected between the neighboring vertices $\mathbf{u}_i^{(m)}$ and $\mathbf{u}_j^{(m)}$. Hence, we have the edge set $\mathcal{E}^{(m)} = \left\{ e_{ij}^{(m)} \mid \mathbf{u}_j^{(m)} \in \mathcal{N}_i^{(m)}, \mathbf{u}_i^{(m)} \in \mathcal{U}^{(m)} \right\}$, and the smoothness term [9] is formulated as

$$E_{smooth}\left(\mathbf{X}\right) = \sum_m \sum_{e_{ij}^{(m)} \in \mathcal{E}^{(m)}} \mathbf{W_s}_p^{(m)}$$
$$\left\| \mathbf{u}_j^{(m)}\mathbf{X}_i^{(m)} - \mathbf{u}_j^{(m)}\mathbf{X}_j^{(m)} \right\|_1, \tag{5}$$

where $\mathbf{W_s}_p^{(m)}$ is our proposed reweighting weight introduced later. The smoothness term can further be rewritten in the matrix form as

$$E_{smooth}\left(\mathbf{X}\right) = \sum_m \left\| \mathbf{W_s}^{(m)} \mathbf{B}^{(m)}\mathbf{X}^{(m)} \right\|_1,$$

$$\mathbf{W_s}^{(m)} = \mathrm{diag}\left(\mathbf{W_s}_1^{(m)}, \dots, \mathbf{W_s}_{P_m}^{(m)} \right), \tag{6}$$

where P_m represents the number of edges in scan $\mathcal{U}^{(m)}$. Let $\mathbf{B} = \mathrm{diag}\left(\mathbf{B}^{(1)}, \dots, \mathbf{B}^{(n)} \right)$, then we have the final form of the smoothness term:

$$E_{smooth}\left(\mathbf{X}\right) = \left\| \mathbf{W_s B X} \right\|_1,$$

$$\mathbf{W_s} = \mathrm{diag}\left(\mathbf{W_s}^{(1)}, \dots, \mathbf{W_s}^{(n)} \right). \tag{7}$$

Orthogonality Term: For non-rigid registration, there are high degrees of freedom in deformation space, which may lead to an unreasonable shape, especially when partial meshes have large motions. Fortunately, the fact is observed and verified that the deformations of usual objects such as human bodies and animals are locally rigid. Therefore, we use an orthogonality constraint similar to [10]:

$$E_{rig}\left(\mathbf{X}\right) = \sum_m \sum_i \left\| \mathbf{D}\mathbf{X}_i^{(m)} - \mathbf{R}_i^{(m)} \right\|_F^2,$$
$$\text{s.t.} \quad \mathbf{R}_i^{(m)\top}\mathbf{R}_i^{(m)} = \mathbf{I}_3, \det(\mathbf{R}_i^{(m)}) > 0, \tag{8}$$

where $\mathbf{D} \in \mathbb{R}^{3\times4}$ is defined to extract the rotation transformation from $\mathbf{X}_i^{(m)}$. To ensure $\mathbf{R}_i^{(m)} > 0$, we enforce to multiply $\mathbf{R}_i^{(m)}$ with -1 if $\det(\mathbf{R}_i^{(m)}) < 0$.

ARAP Term: The shrinkage problem of the deformed models will occur when neighboring scans have less overlap. To address this, we use an as-rigid-as-possible constraint to maintain the lengths of all the edges before and after transformations as much as possible. Denote by $\mathbf{e}_{ij}^{(m)} = \mathbf{v}_i^{(m)} - \mathbf{v}_j^{(m)}$ the original edge and by $\mathbf{e}_{ij}^{'(m)} = \mathbf{v}_i^{'(m)} - \mathbf{v}_j^{'(m)}$ the transformed edge, where $\mathbf{v}_i^{(m)} \triangleq (x_i^{(m)}, y_i^{(m)}, z_i^{(m)})$ represents the Cartesian coordinates of the vertex $\mathbf{u}_i^{(m)} \in \tilde{\mathcal{U}}^{(m)}$, and $\mathbf{v}_i^{'(m)}$ is the transformed vertex of $\mathbf{v}_i^{(m)}$. The ARAP constraint is defined as [12, 13]:

$$E_{arap}\left(\mathbf{X}\right) = \min_{\mathbf{T}_i^{(m)}} \sum_m \sum_i \sum_{j\in\mathcal{N}(i)}$$
$$w_{ij}^{(m)}\left\| (\mathbf{e}_{ij}^{'(m)} - \mathbf{e}_{ij}^{(m)}\mathbf{T}_i^{(m)}) \right\|_F^2, \tag{9}$$

where $w_{ij}^{(m)} = \frac{1}{2}(\cot\alpha_{ij} + \cot\beta_{ij})$ is defined by cotangent weights, in which α_{ij} and β_{ij} are the angles opposite of the mesh edge (i, j) (for a boundary edge, only one such angle exists). $\mathbf{T}_i^{(m)}$ is a rotation matrix obtained by the singular value decomposition (SVD) of $\mathbf{S}_i^{(m)}$, where $\mathbf{S}_i^{(m)}$ is defined as $\mathbf{S}_i^{(m)} = \sum_m \sum_{j\in\mathcal{N}(i)} w_{ij}^{(m)}\mathbf{e}_{ij}^{(m)}\mathbf{e}_{ij}^{'(m)\top}$. Then, we obtain the following by setting the partial deviation $\frac{\partial E_{arap}}{\partial \mathbf{v}_i^{'(m)}}$ to zero:

$$\sum_{j\in\mathcal{N}(i)} w_{ij}^{(m)}\mathbf{e}_{ij}^{'(m)} = \sum_{j\in\mathcal{N}(i)} \frac{w_{ij}^{(m)}}{2}\mathbf{e}_{ij}^{(m)}(\mathbf{T}_i^{(m)} + \mathbf{T}_j^{(m)}). \tag{10}$$

Let $\mathbf{L}^{(m)}$ represent the linear combination on the left-hand side, which is the discrete Laplace-Beltrami operator and $\mathbf{b}^{(m)}$ represent an n-vector whose ith row contains the right-hand side expression. In the setting, the deformed edges have

positions determined by transformations $\mathbf{X}$, which are optimized as a whole, so when defining E_{arap}, only $\mathbf{T}_i^{(m)}$'s are optimized. Using matrix-vector notation, E_{arap} can be rewritten as

$$E_{arap}(\mathbf{X}) = \sum_m \left\| \mathbf{L}^{(m)} \mathbf{X}^{(m)} - \mathbf{b}^{(m)} \right\|_F^2. \qquad (11)$$

Denote by $\mathbf{L} = \operatorname{diag}\left(\mathbf{L}^{(1)}, \ldots, \mathbf{L}^{(n)}\right)$, and by $\mathbf{b} = [\mathbf{b}^{(1)}, \ldots, \mathbf{b}^{(n)}]^\top$, we have the following form of ARAP term:

$$E_{arap}(\mathbf{X}) = \left\| \mathbf{L}\mathbf{X} - \mathbf{b} \right\|_F^2. \qquad (12)$$

ACAP Term: We observe that the registered surfaces may have distortions and fold-overs for drastically different geometries. Hence, we incorporate conformal mapping into the global non-rigid registration framework to suppress distortions and fold-overs by preserving the angles of triangles in the meshes. The conformal mapping is defined by constraining local affine transformations as

$$(\mathbf{D}\mathbf{X}_i^{(m)})^\top (\mathbf{D}\mathbf{X}_i^{(m)}) = s^2 \mathbf{I_d}, \qquad (13)$$

where s is a scale and $\mathbf{I_d}$ is a 3×3 identity matrix. Since the original conformal mapping is nonlinear, we thus derive a linearized version inspired by [14], which proposes the linear approximation of a constrained transformation as:

$$\mathbf{D}\mathbf{X}_i^{(m)} = \begin{pmatrix} s & -h_3 & h_2 \\ h_3 & s & -h_1 \\ -h_2 & h_1 & s \end{pmatrix}. \qquad (14)$$

Therefore, we define the ACAP term as

$$E_{acap}(\mathbf{X}) = \sum_m \sum_i \left\| \mathbf{D}\mathbf{X}_i^{(m)} - f_{Lacap}(\mathbf{X}_i^{(m)}) \right\|_F^2, \quad (15)$$

where $f_{Lacap}(\cdot)$ is defined as

$$f_{Lacap}(\mathbf{X}_i^{(m)}) = \begin{pmatrix} \mathbf{X}_i^{(m)}{}_{22} & -\mathbf{X}_i^{(m)}{}_{21} & -\mathbf{X}_i^{(m)}{}_{31} \\ -\mathbf{X}_i^{(m)}{}_{12} & \mathbf{X}_i^{(m)}{}_{33} & -\mathbf{X}_i^{(m)}{}_{32} \\ -\mathbf{X}_i^{(m)}{}_{13} & -\mathbf{X}_i^{(m)}{}_{23} & \mathbf{X}_i^{(m)}{}_{11} \end{pmatrix}. \quad (16)$$

Reweighting: We design a reweighting scheme on both data term and smoothness term in the form of exponential function instead of inverse function used in the method [9]. For the data term, we design the weight $\mathbf{W_D}_k^{(m)}$ for the vertex $\mathbf{u}_i^{(m)} \in \tilde{\mathcal{U}}^{(m)}$ as

$$\mathbf{W_D}_k^{(m)} = e^{-\dfrac{\left\| \mathbf{u}_i^{(m)} \mathbf{X}_i^{(m)} - \mathbf{u}_{f_m(i)}^{(m+1)} \mathbf{X}_{f_m(i)}^{(m+1)} \right\|_2}{\sigma_D}}. \qquad (17)$$

The weight of the smoothness term is similar to that of the data term. For the kth edge in scan $\mathcal{U}^{(m)}$, it is defined as

$$\mathbf{W_S}_p^{(m)} = e^{-\dfrac{\left\| \mathbf{u}_j^{(m)} \mathbf{X}_i^{(m)} - \mathbf{u}_j^{(m)} \mathbf{X}_j^{(m)} \right\|_2}{\sigma_S}}, \qquad (18)$$

where σ_D and σ_S are the parameters that control the magnitude of $\mathbf{W_D}_k^{(m)}$ and $\mathbf{W_S}_p^{(m)}$.

To optimize the total energy, we set some boundary conditions where a scan *e.g.* $\mathcal{U}^{(1)}$ is fixed, *i.e.* with $\mathbf{X}_i^{(1)}$ being an identity transform for each vertex of the scan. We also introduce auxiliary variables $\mathbf{A}$ and $\mathbf{C}$ into the minimization problem:

$$\min_{\mathbf{X},\mathbf{C},\mathbf{A}} \left\| \mathbf{C} \right\|_1 + \alpha \left\| \mathbf{A} \right\|_1 + \lambda \sum_m \sum_i \left\| \mathbf{D}\mathbf{X}_i^{(m)} - \mathbf{R}_i^{(m)} \right\|_F^2$$
$$+ \beta \left\| \mathbf{L}\mathbf{X} - \mathbf{b} \right\|_F^2 + \gamma \sum_m \sum_i \left\| \mathbf{D}\mathbf{X}_i^{(m)} - f_{Lacap}(\mathbf{X}_i^{(m)}) \right\|_F^2,$$
$$\text{s.t.} \quad \mathbf{C} = \mathbf{W_D}\mathbf{H}\mathbf{X}, \mathbf{A} = \mathbf{W_S}\mathbf{B}\mathbf{X},$$
$$\mathbf{R}_i^{(m)\top} \mathbf{R}_i^{(m)} = \mathbf{I}_3, \det(\mathbf{R}_i^{(m)}) > 0. \qquad (19)$$

We solve the constrained minimization Eq. (19) using the augmented Lagrangian method (ALM).

Multi-Resolution Approach: Considering the high dimensionality and complexity of the deformation space in non-rigid registration, we use a coarse-to-fine approach [10] to reduce the degrees of freedom and deal with the large scale problems efficiently. Suppose that we decompose the shapes up to $S+1$ scales. For any shape $\mathcal{U}^{(m)}$, denote by $\mathcal{U}^{(m)(s)}$ the sth scale of the shape via standard downsampling [15]. Thus $\mathcal{U}^{(m)(0)} \equiv \mathcal{U}^{(m)}$ is at the full resolution while $\mathcal{U}^{(m)(S)}$ is at the coarsest resolution. The optimization Eq. (19) at scale s can be eventually rewritten as:

$$\min_{\mathbf{X},\mathbf{C},\mathbf{A}} \left\| \mathbf{C} \right\|_1 + \alpha \left\| \mathbf{A} \right\|_1 + \lambda \sum_m \sum_i \left\| \mathbf{D}\mathbf{X}_i^{(m)(s)} - \mathbf{R}_i^{(m)(s)} \right\|_F^2$$
$$+ \beta \left\| \mathbf{L}\mathbf{M}\mathbf{X}^{(s)} - \mathbf{b} \right\|_F^2 + \gamma \sum_m \sum_i \left\| \mathbf{D}\mathbf{X}_i^{(m)(s)} - f_{Lacap}(\mathbf{X}_i^{(m)(s)}) \right\|_F^2,$$
$$\text{s.t.} \quad \mathbf{C} = \mathbf{W_D}\mathbf{H}\mathbf{M}\mathbf{X}^{(s)}, \mathbf{A} = \mathbf{W_S}\mathbf{B}\mathbf{M}\mathbf{X}^{(s)},$$
$$\mathbf{R}_i^{(m)(s)\top} \mathbf{R}_i^{(m)(s)} = \mathbf{I}_3, \det(\mathbf{R}_i^{(m)(s)}) > 0, \qquad (20)$$

where $\mathbf{M}$ contains the mapping transformations from $\mathcal{U}^{(m)(s)}$ to $\mathcal{U}^{(m)(s-1)}$ for all scans, and $\mathbf{X}^{(s)}$ contains the transformations on all the $\mathcal{U}^{(m)(s)}$.

3. EXPERIMENTAL RESULTS

3.1. Results on Public Datasets

We first evaluate our method on *Jumping* dataset [16] which contains complete models with large deformations and known correspondences, compared with three state-of-the-art methods [17, 9, 10]. For pairwise registration methods

[17, 9], we register all the models in sequence with the previous registration result used as the next target model. Besides, we use the first model as the reference pose model for all the methods. Fig. 1 shows alignment results of all the methods for ten models. The registration errors between the deformed models and the reference model are color-coded on the deformed models for visual inspection. The distance errors are computed using the standard Metro tool [18]. It can be seen that our method achieves best registration results. Table 1 gives quantitative evaluation for clean and noisy datasets. Our method consistently has the smallest average fitting errors.
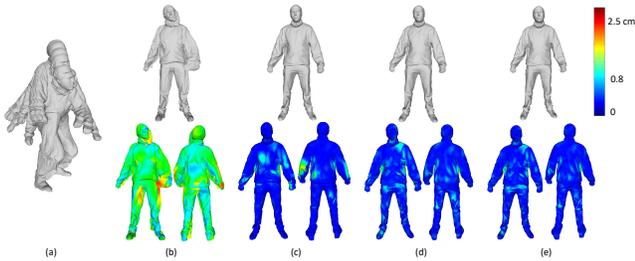
Fig. 1. Alignment results on *Jumping* dataset: (a) original models, (b) results of [17], (c) results of [9], (d) results of [10] and (e) our results.

Table 1. Quantitative evaluation for $Jumping$ dataset (cm).

Method	Clean Data	Noise Data	Outlier Data
[17]	1.8122	1.8646	1.9780
[9]	0.1497	0.1841	0.1844
[10]	0.1210	0.1550	0.1621
Ours	**0.0931**	**0.1444**	**0.1379**

We also evaluate all the methods on 35 partial models extracted from the complete models with different geometries and poses in *Bouncing* dataset [16]. To obtain these partial models, we only extract visible parts of each complete model under different viewpoints by rotating a virtual camera round the model. Then we downsample the original partial models with 3000-5000 vertices to 1/10 of the full resolution and get low-resolution models with 300-500 vertices using method [15] in order to use the multi-resolution approach. We solve the global registration problem from coarse to fine, and use standard Poisson reconstruction [19] to merge all the registered scans into a watertight mesh. As shown in Fig. 2, the method [17] has severe distortions on the left arm and right leg, and large fitting errors for the back. Visible misalignment on the left arm and head of the fused model can be found for the method [9] due to error accumulation. With the strategy of global registration, both the method [10] and our method avoid the well-known loop closure problem and obtain plausible reconstruction results, but our method has smaller registration error, especially on the left arm/shoulder. Table 2 gives quantitative evaluation for clean and noisy cases. Our method consistently has the smallest average fitting errors.

Table 2. Quantitative evaluation for partial models in $Bouncing$ dataset (cm).

| Method	Clean Data	Noise Data	Outlier Data
[17]	1.6989	1.7631	2.1537
[9]	1.0045	1.0093	1.0426
[10]	0.7396	0.9853	0.9580
Ours	**0.4950**	**0.5338**	**0.5440**
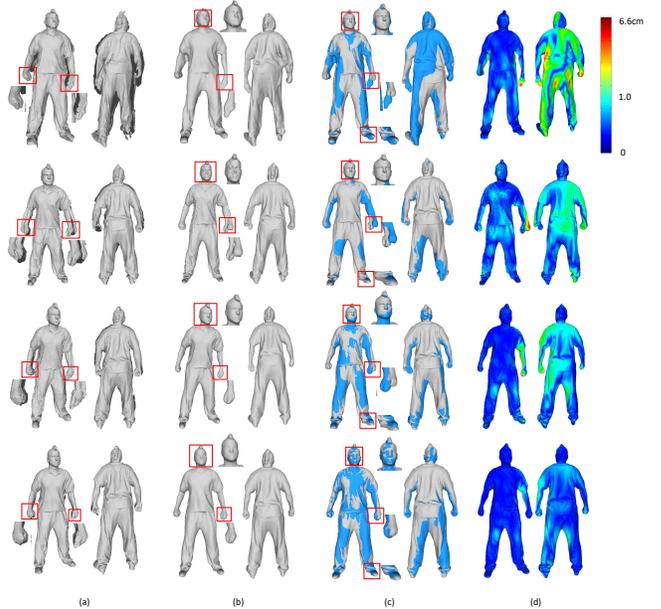
Fig. 2. Results on *Bouncing* dataset of method [17] (top row), method [9] (second row), method [10] (third row) and our method (bottom row): (a) alignment results of all the scans, (b) Possion reconstruction results for (a), (c) the overlap between (b) (blue) and the reference model (gray), and (d) the color-coded models by the fitting errors.

3.2. Results on Real Scans

We test our method on *Flying* dataset [10], a very challenging real dataset scanned using a Kinect v2.0 containing much noise and a large number of outliers. Fig. 3 illustrates the results when scans are accumulated gradually. From visual inspection, it is obvious that the results of [17] have severe distortions during pairwise non-rigid registration in succession. Besides, there is palpable misalignment in the results of [9] because of error accumulation, especially on the arms. The fused models of both the method [10] and our method look acceptable, but the deformed scans by our method are aligned more accurately, such as the hands.

4. CONCLUSIONS

This paper presents a novel approach which achieves global non-rigid registration of multi-view scans with drastically different geometries and motions. Our approach allows to preserve mesh structures during deformation by incorporat-
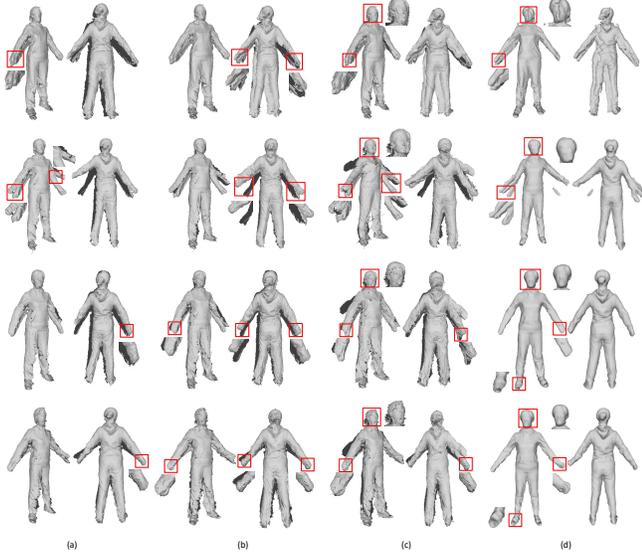
Fig. 3. Results using gradually accumulated scans on *Flying* dataset of method [17] (top row), method [9] (second row), method [10] (third row) and our method (bottom row): (a) results of scans 1-6, (b) results of scans 1-10, (c) results of all the scans, and (d) Poisson reconstruction results for (c).

ing ACAP constraint into the energy function. Moreover, with the reweighting scheme, the global alignment problem is optimized with smaller registration errors. The problem is solved by the alternating direction method under the augmented Lagrangian multiplier framework, which has exact solutions and guaranteed convergence. Experimental results on both public and real scanned datasets show that our method achieves better results compared with state-of-the-art methods. As future work, we will exploit potential overlaps between non-adjacent scans to find more candidate correspondences to enable accurate alignment and high fidelity reconstruction.

5. REFERENCES

[1] K. Li, Q. Dai, and W. Xu, "Markerless shape and motion capture from multi-view video sequences," *IEEE TCSVT*, vol. 21, no. 3, pp. 320–334, 2011.

[2] R. A. Newcombe *et al.*, "KinectFusion: Real-time dense surface mapping and tracking," in *IEEE ISMAR*, 2011.

[3] M. Zollhöfer, M. Nießner, S. Izadi, C. Rhemann, C. Zach, M. Fisher, C. Wu, A. Fitzgibbon, C. Loop, C. Theobalt, and M. Stamminger, "Real-time non-rigid reconstruction using an RGB-D camera," *ACM ToG*, vol. 33, no. 4, 2014.

[4] A. Collet, M. Chuang, P. Sweeney, D. Gillett, D. Evseev, D. Calabrese, Hu. Hoppe, A. Kirk, and S. Sullivan,
"High-quality streamable free-viewpoint video," *ACM ToG*, vol. 34, no. 4, 2015.

[5] J. Tong, J. Zhou, L. Liu, Z. Pan, and H. Yan, "Scanning 3D full human bodies using Kinects," *IEEE TVCG*, vol. 18, no. 4, pp. 643–650, 2017.

[6] H. Li, E. Vouga, A. Gudym, L. Luo, J. T. Barron, and G. Gusev, "3D self-portraits," *ACM ToG*, vol. 32, no. 6, pp. 2504–2507, 2013.

[7] R. A. Newcombe, D. Fox, and S. M. Seitz, "Dynamic-Fusion: Reconstruction and tracking of non-rigid scenes in real-time," in *CVPR*, 2015, pp. 343–352.

[8] T. Yu, Z. Zheng, K. Guo, J. Zhao, Q. Dai, H. Li, G. Pons-Moll, and Y. Liu, "DoubleFusion: Real-time capture of human performances with inner body shapes from a single depth sensor," in *CVPR*, 2018.

[9] J. Yang, K. Li, Y.-K. Lai, and D. Guo, "Robust non-rigid registration with reweighted position and transformation sparsity," *IEEE TVCG*, 2018.

[10] D. Guo, K. Li, Y. Lai, and J. Yang, "Global alignment of deformable objects captured by a single RGB-D camera," in *ICME*. IEEE, 2017.

[11] G. KL Tam, R. R. Martin, P. L. Rosin, and Y.-K. Lai, "Diffusion pruning for rapidly and robustly selecting global correspondences using local isometry," *ACM ToG*, vol. 33, no. 1, pp. 4, 2014.

[12] D. Guo, K. Li, Y.-K. Lai, and J. Yang, "Global alignment for deformable objects captured by a single RGB-D camera," in *ICME*, 2017.

[13] O. Sorkine and M. Alexa, "As-rigid-as-possible surface modeling," in *SGP*, 2007, pp. 109–116.

[14] O. Sorkine, D. Cohen-Or, Y. Lipman, M. Alexa, C. Rössl, and H. Seidel, "Laplacian surface editing," in *SGP*. ACM, 2004, pp. 175–184.

[15] M. Garland and P. S. Heckbert, "Surface simplification using quadric error metrics," in *ACM SIGGRAPH*, 1997.

[16] D. Vlasic, I. Baran, W. Matusik, and J. Popović, "Articulated mesh animation from multi-view silhouettes," *ACM ToG*, vol. 27, no. 3, pp. 97, 2008.

[17] J. Ma, J. Zhao, and A. L Yuille, "Non-rigid point set registration by preserving global and local structures," *IEEE TIP*, vol. 25, no. 1, pp. 53–64, 2016.

[18] P. Cignoni, C. Rocchini, and R. Scopigno, "METRO: Measuring error on simplified surfaces," *CGF*, vol. 17, no. 2, pp. 167–174, 1998.

[19] M. Kazhdan and H. Hoppe, "Screened poisson surface reconstruction," *ACM ToG*, vol. 32, no. 3, 2013.