GLOBAL ALIGNMENT OF DEFORMABLE OBJECTS CAPTURED BY A SINGLE RGB-D CAMERA

Daoliang Guo [1], *Kun Li* [1*], *Yu-Kun Lai* [2], *Jingyu Yang* [1]

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

ABSTRACT

We present a novel global registration method for deformable objects captured using a single RGB-D camera. Our algorithm allows objects to undergo large non-rigid deformations, and achieves high quality results without constraining the actor's pose or camera motion. We compute the deformations of all the scans simultaneously by optimizing a global alignment problem to avoid the well-known loop closure problem, and use an as-rigid-as-possible constraint to eliminate the shrinkage problem of the deformed model. To attack large scale problems, we design a coarse-to-fine multi-resolution scheme, which also avoids the optimization being trapped into local minima. The proposed method is evaluated on public datasets and real datasets captured by an RGB-D sensor. Experimental results demonstrate that the proposed method obtains better results than the state-of-the-art methods.

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

1. INTRODUCTION

Dynamic 3D reconstruction is an active research topic in computer graphics and computer vision [1, 2, 3], which tries to recover dynamic scenes by capturing videos using multiple cameras or a single camera. With the appearance of commodity depth cameras, *e.g.*, Microsoft Kinect, it is easier and cheaper to reconstruct the shape and texture of a 3D scene using a single RGB-D camera. This has wide applicability in 3D printing, gaming, and movie production. However, when deformable objects are concerned, reconstruction results obtained using KinectFusion [4] has serious drifting artifacts. Moreover, the captured depth point clouds usually contain a lot of noise and outliers. Hence, it remains a huge challenge to reconstruct dynamic 3D scenes using a single RGB-D camera.

To achieve dynamic 3D reconstruction, several groups have set up multi-camera systems. Li *et al.* [2] at Tsinghua University build a dome system with 20 cameras to synchronously capture and recover the dynamic shape and texture of arbitrary objects using a variational multi-view stereo method and a volumetric deformation method. Aguiar *et al.* [3] at MPI Informatik build a sparsely sampled system of 8 cameras to capture the shape and motion of 3D objects by effectively combining the power of surface- and volume-based shape deformation techniques. However, high cost, complex maintenance, and lack of portability limit the practical applications of such systems. The Microsoft Kinect camera has been widely used due to its low-cost and multi-sensing. Tong *et al.* [5] scan a 3D full human body model using three Kinect cameras, but the captured person must keep still. Guo *et al.* [1] achieve marker-less performance capture of interacting humans using three hand-held Kinect cameras. To be easier and more convenient, Li *et al.* [6] capture a complete 3D model using only a single Kinect sensor, but the poses of the person in various viewpoints must keep the same. Dou *et al.* [7] develop a 3D scanning system which allows considerable amount of deformations during scanning, but the deformation between two neighboring viewpoints (time instances) cannot be large. To our knowledge, little work in the literature allows large motions of the person between different viewpoints, which is common for snapshot or high speed motion capture.

In this paper, we propose a method for global non-rigid registration and reconstruction of deformable objects with a single RGB-D camera, building on a recent pairwise sparse non-rigid registration framework [8, 9]. The motion of the object between different viewpoints can be very large. Naive solution of applying pairwise non-rigid registration in succession leads to error accumulation and the well-known loop closure problem. To address this, we compute the deformations of all the scans simultaneously by optimizing a *global* alignment problem. We introduce an as-rigid-as-possible (ARAP) constraint to the sparse non-rigid registration framework to eliminate the shrinkage problem of the deformed models when overlapping regions are small and the problem would otherwise be underconstrained, and design a coarse-to-fine multi-resolution scheme to improve efficiency and robustness. The proposed method is evaluated on public datasets and real datasets captured by an RGB-D sensor. The results demon-

[*]Corresponding author: lik@tju.edu.cn. This work was supported in part by the National Natural Science Foundation of China (Grant 61571322 and 61372084).

978-1-5090-6067-2/17/$31.00 ©2017 IEEE

strate that the proposed method obtains better results than the state-of-the-art methods.

The main contributions of this work are summarized as:

- We propose a global optimization method for reconstruction of deformable objects with large motions, which is robust to noise and outliers, and avoids the loop closure problem.

- We design a coarse-to-fine multi-resolution scheme to avoid the optimization being trapped into local minima, which also helps to attack large scale problems that would otherwise be prohibitively expensive (in terms of computation and storage costs).

- We introduce an ARAP constraint to the sparse non-rigid registration framework, which eliminates the shrinkage problem of the deformed models.

2. RELATED WORK

We review recent relevant work in 3D object reconstruction. Firstly, for multi-camera systems, drifting is not a concern given that a relatively complete model is captured at each frame. Starck *et al.* [10] design a system to reconstruct a full human body using 16 cameras, which require careful positioning of cameras to obtain better raw data. Li *et al.* [2] and Aguiar *et al.* [3] similarly obtain full objects using 20 cameras and 8 cameras, respectively. Tong *et al.* [5] use a turntable to turn a person around to reconstruct a full body, but the method cannot handle large deformations. Other systems [1, 11, 12], either require complex lighting or cannot generate high quality results. Ye *et al.* [12] use 3 handheld Kinects to reconstruct human performance with deforming template models. Dou *et al.* [11] scan and track deforming objects using fusion of dynamic inputs from 8 Kinects.

Considering the high cost of multi-camera systems, the approach based on a single camera becomes more and more popular. Many systems [5] focusing on scanning humans require the person to stand still. However, even for rigid alignment, the problem of drifting occurs when aligning a sequence of partial scans consecutively, where the alignment error accumulates quickly and the scan does not close seamlessly. Drifting is more serious in the case of non-rigid alignment. To make the problem more tractable, some existing systems rely on exploiting specific poses or using parametric models. Zollhöfer *et al.* [13] propose an approach based on the template prior, which acquires a template of the object using KinectFusion and registers the template to the non-rigid sequences. Cui *et al.* [14] propose a method which limits the user to keep a 'T' shape. Li *et al.* [6] adapt a more general non-rigid registration framework which allows a wider range of poses and multiple actors. This system demonstrates compelling results, but it requires users to keep almost the same pose and follow a specific sequence of scanning. Dou *et al.* [7] develop a 3D scanning system which allows considerable amount of deformations during scanning and show fine results, but the deformation between two neighboring viewpoints (time instances) cannot be large, and the cost of computation and storage are high. Alternative methods [1, 15] are also based on tracking and fusion of RGB-D sequences of non-rigidly deforming objects, although with different formulations. They have similar limitations that neighboring views can only have mild deformation.

Our work uses a single Kinect sensor to capture several different noisy partial scans of a deformable object, and aligns the scans in a global framework without the drifting problem. Moreover, our method allows a large motions of the object during scanning. To achieve this, we propose a global sparse non-rigid registration iteration framework. Unlike most of existing non-rigid registration methods that are based on pairwise registration, we propose a global non-rigid framework based on sparse priors [8, 9], as they are robust to noise and outliers. We further introduce an ARAP constraint to the sparse non-rigid registration framework to eliminate the shrinking problem.

3. THE PROPOSED METHOD

3.1. Iterative Framework

The aim of global non-rigid registration is to find a set of non-rigid transformations $\mathbf{X}$ that transform scans so that they are consistently aligned while satisfying smoothness prior. For this end, an iterative procedure is applied with the following two alternating steps: 1) given the current transformations (and hence the vertex positions after deformation), refine the correspondences between each pair of scans as long as they overlap. In practice, if the scans are circularly distributed, it is sufficient to consider adjacent pairs. 2) given pairwise corresponding mappings, find a set of local affine transformations by minimizing a *global* energy function (details given later). Compared with straightforward successive pairwise registration, the benefit of global registration is to avoid the well-known loop closure problem where the misalignment accumulates and the surfaces do not match up when the last pair is to be registered.

3.2. Global Registration

Assume that we have M scans to be registered $\mathcal{U}^{(1)},\mathcal{U}^{(2)},\ldots,\mathcal{U}^{(M)}$. For each scan, $\mathcal{U}^{(m)} \triangleq \left\{\mathbf{u}_1^{(m)},\mathbf{u}_2^{(m)},\ldots,\mathbf{u}_{N_m}^{(m)}\right\}$, where N_m is the number of vertices in the scan $\mathcal{U}^{(m)}$. $\mathbf{u}_i^{(m)} \triangleq (x_i^{(m)},y_i^{(m)},z_i^{(m)},1)$ represents the homogeneous coordinates of vertex $\mathbf{u}_i^{(m)}$. For a neighboring pair of scans $\mathcal{U}^{(m)}$ and $\mathcal{U}^{(m+1)}$ (assuming $\mathcal{U}^{(M+1)} = \mathcal{U}^{(1)}$), let $f_{m\rightarrow m+1} : \{1,\cdots,N_m\} \mapsto \{1,\cdots,N_{m+1}\}$ be the index mapping from the points on $\mathcal{U}^{(m)}$ to the points on $\mathcal{U}^{(m+1)}$ established by correspondence computation: $\mathbf{u}_{f_m(i)}^{(m+1)} \in \mathcal{U}^{(m+1)}$ is the corresponding point of $\mathbf{u}_i^{(m)} \in \mathcal{U}^{(m)}$. For non-rigid registration, we allow 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 $\mathbf{X}^{(m)} \triangleq \left[\mathbf{X}_1^{(m)}; \cdots; \mathbf{X}_{N_m}^{(m)}\right]$ of size $4N_m \times 3$ the ensemble matrix containing N_m transformation matrices to be estimated.

Energy Function Formulation: The overall function to be minimized in Step 2 is given as follows:

$$\begin{aligned} E\left(\mathbf{X}; f\right) = E_{data}\left(\mathbf{X}; f\right) + \alpha E_{smooth}\left(\mathbf{X}\right) \\ + \lambda E_{rig}\left(\mathbf{X}\right) + \beta E_{arap}\left(\mathbf{X}\right), \end{aligned} \tag{1}$$

where $E_{data}\left(\mathbf{X}\right)$ is the data term to measure the registration accuracy, $E_{smooth}\left(\mathbf{X}\right)$ is the smoothness term to measure the smoothness of local transformations, $E_{rig}\left(\mathbf{X}\right)$ is the orthogonality term to measure the rigidness of local transformations and $E_{arap}\left(\mathbf{X}\right)$ is the as-rigid-as-possible constraint to ensure the length of each edge to be as close as possible before and after transformation; α, λ and β are weights to balance the relative importance of the terms. The four terms are defined as follows.

Data Term: A similar strategy as the pairwise registration is used to estimate the mapping between a neighboring pair of overlapping scans $\mathcal{U}^{(m)}$ and $\mathcal{U}^{(m+1)}$: $f_{m \to m+1}$, noted as f_m. As neighboring surfaces only have partial overlaps, not every point has a corresponding point, so we assume there are K_m corresponding points between $\mathcal{U}^{(m)}$ and $\mathcal{U}^{(m+1)}$, where $K_m \leq min(N_m, N_{m+1})$. Given the correspondence mapping f_m where $f_m(i,1)$ and $f_m(i,2)$ are the indexes of corresponding points on $\mathcal{U}^{(m)}$ and $\mathcal{U}^{(m+1)}$, respectively. The data term is defined by summing over each neighboring pair of overlapping scans $\mathcal{U}^{(m)}$ and $\mathcal{U}^{(m+1)}$:

$$E_{data}\left(\mathbf{X}; f\right) \triangleq \sum_m \sum_{\mathbf{u}_{f_m(i,1)}^{(m)} \in \mathcal{U}^{(m)}} \left\| \mathbf{u}_{f_m(i,1)}^{(m)} \mathbf{X}_{f_m(i,1)}^{(m)} - \mathbf{u}_{f_m(i,2)}^{(m+1)} \mathbf{X}_{f_m(i,2)}^{(m+1)} \right\|_1 \tag{2}$$

where $\|\cdot\|_1$ denotes ℓ_1 norm of a matrix considered as a long vector. Data term (2) can be rewritten as

$$E_{data}\left(\mathbf{X}; f\right) = \sum_m \left\| \left(\mathbf{U}^{(m)} \mathbf{X}^{(m)} - \mathbf{U}_{f_m}^{(m+1)} \mathbf{X}^{(m+1)} \right) \right\|_1, \tag{3}$$

$\mathbf{U}^{(m)}$ and $\mathbf{U}_{f_m}^{(m+1)}$ are $K_m \times 4N_m$ and $K_m \times 4N_{m+1}$ respectively. The ith row of $\mathbf{U}^{(m)}$ and $\mathbf{U}_{f_m}^{(m+1)}$ is associated with the ith correspondence,with elements $\mathbf{u}_{f_m(i,1)}^{(m)}$ and $\mathbf{u}_{f_m(i,2)}^{(m)}$ in relevant columns. Let $\mathbf{X} := \left[\mathbf{X}^{(1)}; \ldots; \mathbf{X}^{(M)}\right]$, using matrix notation, we have the following form of data term:

$$E_{data}\left(\mathbf{X}; f\right) = \left\|\mathbf{H}\mathbf{X}\right\|_1, \tag{4}$$

where $\mathbf{H}$ is determined according to the overlapping relationship:

$$\mathbf{H} = \begin{bmatrix} \mathbf{U}^{(1)} & -\mathbf{U}_{f_1}^{(2)} & & & & \\ & \mathbf{U}^{(2)} & -\mathbf{U}_{f_2}^{(3)} & & & \\ & & \ddots & \ddots & & \\ & & & & \mathbf{U}^{(M-1)} & -\mathbf{U}_{f_{M-1}}^{(M)} \\ -\mathbf{U}_{f_M}^{(1)} & & & & & \mathbf{U}^{(M)} \end{bmatrix}. \tag{5}$$

Smoothness Term: Similar to the pairwise registration, we use the edge set defined with a neighboring system. For scan $\mathcal{U}^{(m)}$, denote by $\mathcal{N}_i^{(m)}$ the neighborhood of vertex $\mathbf{u}_i^{(m)}$, and by $e_{ij}^{(m)}$ the edge defined between each pair of neighboring vertices $\mathbf{u}_j^{(m)}$ and $\mathbf{u}_i^{(m)}$. So, we have the edge set $\mathcal{E}^{(m)} = \left\{ e_{ij}^{(m)} \mid \mathbf{u}_j^{(m)} \in \mathcal{N}_i^{(m)}, \mathbf{u}_i^{(m)} \in \mathcal{U}^{(m)} \right\}$. Similar to the pairwise registration, the smoothness is regularized by the ℓ_1 norm of transformation differences on the neighboring system over all the scans $\mathcal{U}^{(m)}$ [9]:

$$E_{smooth}\left(\mathbf{X}\right) = \sum_m \sum_{e_{ij}^{(m)} \in \mathcal{E}^{(m)}} \left\| \mathbf{u}_j^{(m)} \mathbf{X}_i^{(m)} - \mathbf{u}_j^{(m)} \mathbf{X}_j^{(m)} \right\|_1. \tag{6}$$

E_{smooth} can be rewritten in the matrix form as:

$$E_{smooth}\left(\mathbf{X}\right) = \sum_m \left\| \mathbf{B}^{(m)} \mathbf{X}^{(m)} \right\|_1. \tag{7}$$

Let $\mathbf{B} = \text{diag}\left(\mathbf{B}^{(1)}, \ldots, \mathbf{B}^{(n)}\right)$, and we have the following form of the smoothness term:

$$E_{smooth}\left(\mathbf{X}\right) = \left\|\mathbf{B}\mathbf{X}\right\|_1. \tag{8}$$

Orthogonality Term: In non-rigid registration, each node is assigned an affine transformation, which provides sufficient flexibility to capture non-rigidness of deformable objects. However, even with smoothness regularization, the high degrees of freedom may also result in unreasonable deformation. Since the deformation of usual objects such as human bodies and animals are locally rigid, a local rigidness term is used to reduce the flexibility of the transformations. Specifically, the transformation $\mathbf{X}_i^{(m)}$ is assumed to be rigid, consisting of a rotation and a translation where the rotation is represented by an orthonormal matrix. To this end, the orthogonality term is defined as follows [9]:

$$\begin{aligned} E_{rig}\left(\mathbf{X}\right) = \sum_m \sum_i \left\| \mathbf{D}_i \mathbf{X}_i^{(m)} - \mathbf{R}_i^{(m)} \right\|_F^2, \\ \text{s.t.} \quad \mathbf{R}_i^{(m)^T} \mathbf{R}_i^{(m)} = \mathbf{I}_3, \det(\mathbf{R}_i^{(m)}) > 0, \end{aligned} \tag{9}$$

where $\mathbf{D}_i$ is a constant 3×4 matrix which is used to extract the rotation transformation from $\mathbf{X}_i^{(m)}$. To eliminate the case of reflection, we enforce a positive determinant of $\mathbf{R}_i^{(m)}$. If $\det(\mathbf{R}_i^{(m)}) < 0$, we multiply $\mathbf{R}_i^{(m)}$ with -1.

ARAP Term: We observe that some nodes of the registered surfaces may have inwards shrinking, especially when neighboring scans have less overlap. To avoid this artifact, we introduce an as-rigid-as-possible term to the sparse non-rigid registration framework to maintain the lengths of all the edges before and after transformations as much as possible. In the following, we denote the edge $\mathbf{e}_{ij}^{(m)} = \mathbf{p}_i^{(m)} - \mathbf{p}_j^{(m)}$, and similarly the transformed edge $\mathbf{e}_{ij}^{'(m)} = \mathbf{p}_i^{'(m)} - \mathbf{p}_j^{'(m)}$ for the deformed model, where the $\mathbf{p}_i^{(m)} \triangleq (x_i^{(m)}, y_i^{(m)}, z_i^{(m)})$ is the vertex position of $\mathcal{U}^{(m)}$. We define the ARAP term as follows, similar to [16, 13]:

$$E_{arap}\left(\mathbf{X}\right) = \min_{\mathbf{T}_i^{(m)}} \sum_m \sum_i w_i^{(m)} \sum_{\mathbf{j}\in\mathcal{N}(i)} w_{ij}^{(m)} \left\| \mathbf{e}_{ij}^{'(m)} - \mathbf{e}_{ij}^{(m)} \mathbf{T}_i^{(m)} \right\|^2 \tag{10}$$

where $w_i^{(m)} = 1$ for vertices with known correspondence and 0 otherwise, and $w_{ij}^{(m)}$ is defined by cotangent weights:

$$w_{ij}^{(m)} = \frac{1}{2}(\cot\alpha_{ij} + \cot\beta_{ij}), \tag{11}$$

where α_{ij} and β_{ij} are the angles opposite to the mesh edge (i, j) (for a boundary edge, only one such angle exists). $\mathbf{T}_i{}^{(m)} \in \mathbb{R}^{3\times 3}$ is a rotation matrix. Given the positions of deformed vertices, $\mathbf{T}_i{}^{(m)}$ can be explicitly obtained using the singular value decomposition (SVD) of $\mathbf{S}_i^{(m)}$, where $\mathbf{S}_i^{(m)}$ is defined as

$$\mathbf{S}_i^{(m)} = \sum_m \sum_{j\in\mathcal{N}(i)} w_{ij}^{(m)} \mathbf{e}_{ij}^{(m)^T} \mathbf{e}_{ij}^{'(m)}. \tag{12}$$

Using SVD, we can obtain $\mathbf{S}_i^{(m)} = \mathbf{V}_i^m \mathbf{\Sigma}_i^{(m)} \mathbf{U}_i^{(m)^T}$, and $\mathbf{T}_i^{(m)}$ is solved as:

$$\mathbf{T}_i^{(m)} = \mathbf{V}_i^m \mathbf{U}_i^{(m)^T}. \tag{13}$$

To minimize E_{arap} *w.r.t.* $\mathbf{X}_i^{(m)}$, we first work out $\frac{\partial E_{arap}}{\partial \mathbf{p}_i^{'(m)}}$ where $\mathbf{p}_i^{'(m)} = \mathbf{u}_i^{(m)} \mathbf{X}_i^{(m)}$ is the transformed vertex position:

$$\sum_{\mathbf{j}\in\mathcal{N}(i)} w_{ij}^{(m)} (\mathbf{p}_i^{'(m)} - \mathbf{p}_j^{'(m)}) = \sum_{\mathbf{j}\in\mathcal{N}(i)} \frac{w_{ij}^{(m)}}{2} (\mathbf{p}_i^{(m)} - \mathbf{p}_j^{(m)})(\mathbf{T}_i^{(m)} + \mathbf{T}_j^{(m)}). \tag{14}$$

Using matrix-vector notation, E_{arap} can be rewritten as

$$E_{arap}\left(\mathbf{X}\right) = \sum_m \left\|\mathbf{L}^{(m)}\mathbf{X}^{(m)} - \mathbf{b}^{(m)}\right\|_F^2, \tag{15}$$

where $\mathbf{L}^{(m)}$ represents the linear combination on the left-hand side, which is the discrete Laplace-Beltrami operator. $\mathbf{b}^{(m)}$ is an n-vector whose ith row contains the right-hand side expression. In our 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.

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

$$E_{arap}\left(\mathbf{X}\right) = \left\|\mathbf{L}\mathbf{X} - \mathbf{b}\right\|_F^2. \tag{16}$$

Boundary Conditions: For the optimization to have a unique solution, some boundary conditions need to be set. One way is to set a scan *e.g.* $\mathcal{U}^{(1)}$ to be fixed, *i.e.* with $\mathbf{X}_i^{(1)}$ to be identity transformation for each vertex of the scan.

With all these terms, we have the following minimization problem:

$$\begin{aligned} \min_{\mathbf{X},\mathbf{C},\mathbf{A}} & \left\|\mathbf{C}\right\|_1 + \alpha\left\|\mathbf{A}\right\|_1 + \lambda\sum_m\sum_i \left\|\mathbf{D}_i\mathbf{X}_i^{(m)} - \mathbf{R}_i^{(m)}\right\|_F^2 \\ & + \beta\left\|\mathbf{L}\mathbf{X} - \mathbf{b}\right\|_F^2, \\ \text{s.t.} \quad & \mathbf{C} = \mathbf{H}\mathbf{X}, \mathbf{A} = \mathbf{B}\mathbf{X}, \mathbf{R}_i^{(m)^T}\mathbf{R}_i^{(m)} = \mathbf{I}_3, \det(\mathbf{R}_i^{(m)}) > 0, \end{aligned} \tag{17}$$

where $\mathbf{A}$ and $\mathbf{C}$ are auxiliary variables to facilitate optimization. Then, we solve the constrained minimization (17) using the augmented Lagrangian method (ALM). The ALM method converts the original problem (17) to iterative minimization of its augmented Lagrangian function:

$$\begin{aligned} L(\mathbf{X}, \mathbf{C}, \mathbf{A}, \mathbf{Y}_1, \mathbf{Y}_2, \mu_1, \mu_2) &= \left\|\mathbf{C}\right\|_1 + \alpha\left\|\mathbf{A}\right\|_1 \\ & + \langle\mathbf{Y}_1, \mathbf{C} - \mathbf{H}\mathbf{X}\rangle + \frac{\mu_1}{2}\left\|\mathbf{C} - \mathbf{H}\mathbf{X}\right\|_F^2 \\ & + \langle\mathbf{Y}_2, \mathbf{A} - \mathbf{B}\mathbf{X}\rangle + \frac{\mu_2}{2}\left\|\mathbf{A} - \mathbf{B}\mathbf{X}\right\|_F^2 \\ & + \lambda\sum_m\sum_i\left\|\mathbf{D}_i\mathbf{X}_i^{(m)} - \mathbf{R}_i^{(m)}\right\|_F^2 + \beta\left\|\mathbf{L}\mathbf{X} - \mathbf{b}\right\|_F^2, \\ \text{s.t.} \quad & \mathbf{R}_i^{(m)^T}\mathbf{R}_i^{(m)} = \mathbf{I}_3, \det(\mathbf{R}_i^{(m)}) > 0, \end{aligned} \tag{18}$$

we resort to the alternate direction method (ADM) [17] to optimize $\mathbf{A}$, $\mathbf{C}$ and $\mathbf{X}$ separately at each iteration.

Multi-Resolution Approach: Considering the transformation $\mathbf{X}_i$ of each point i has a rotation $\mathbf{R}_i \in \mathbb{R}^{3\times 3}$ and a translation $\mathbf{t}_i \in \mathbb{R}^3$, hence, there are 12 degrees of freedom (DoFs) in total for each $\mathbf{X}_i$. However, if a scan m has N_m vertices, there are N_m transformations and $12N_m$ DoFs, which is not enough to identify a unique solution with N_m constraints. One way of addressing this is to rely on regularization, but the high complexity remains. We further use a coarse-to-fine approach, which can not only provide a promising solution, but also deal with the large scale problems efficiently.

Suppose that we decompose the shapes up to S scales. For any shape $\mathcal{U}^{(m)}$, denote by $\mathcal{U}^{(m)(s)}$ the s^{th} scale of the shape via standard downsampling [18]. We obtain S multi-resolution shapes, $\mathcal{U}^{(m)(S)}, \mathcal{U}^{(m)(S-1)}, \cdots, \mathcal{U}^{(m)(0)}$, where $\mathcal{U}^{(m)(S)}$ is the shape at the coarsest resolution and $\mathcal{U}^{(m)(0)} \equiv \mathcal{U}^{(m)}$ is at the full resolution. The optimization Eq. (17) at scale s can be rewritten as:

$$\begin{aligned} \min_{\mathbf{X},\mathbf{C},\mathbf{A}} & \left\|\mathbf{C}\right\|_1 + \alpha\left\|\mathbf{A}\right\|_1 + \lambda\sum_m\sum_i\left\|\mathbf{D}_i\mathbf{X}_i^{(m)(s)} - \mathbf{R}_i^{(m)(s)}\right\|_F^2 \\ & + \beta\left\|\mathbf{L}\mathbf{M}\mathbf{X}^{(s)} - \mathbf{b}\right\|_F^2, \\ \text{s.t.} \quad & \mathbf{C} = \mathbf{H}\mathbf{M}\mathbf{X}^{(s)}, \mathbf{A} = \mathbf{B}\mathbf{M}\mathbf{X}^{(s)}, \\ & \mathbf{R}_i^{(m)(s)^T}\mathbf{R}_i^{(m)(s)} = \mathbf{I}_3, \det(\mathbf{R}_i^{(m)(s)}) > 0, \end{aligned} \tag{19}$$

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 $\mathcal{U}^{(m)(s)}$.

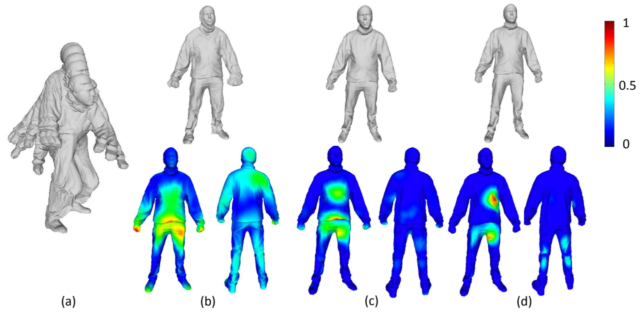

Fig. 1. Results on *Jumping* dataset: (a) original models, (b) results of [8], (c) results of [9] and (d) our results.

4. EXPERIMENTAL RESULTS

We evaluate the performance of our method on public datasets (Section 4.1), and real scans (Section 4.2).

4.1. Results on Public Datasets

Firstly, we evaluate the proposed method on the *Jumping* dataset [19], which are complete models and contain dramatic deformations. They have known correspondences to allow quantitative evaluation. Fig. 1 shows the alignment results, compared with the pairwise registration methods [8, 9]. Considering that these methods only register two models, we register all the models in sequence with the previous registration result used as the next target model. Besides, we select the first partial model as the reference pose model for all the methods. We select 10 complete models shown in Fig. 1(a), which have different motions. The results show the registered models and the average fitting errors between the deformed model and the reference model. The fitting errors are color-coded on the reference model for visual inspection. It can be seen that the method [8] has large average fitting errors, while the sum of average fitting errors over all the vertices of the method [9] and our method are 1.0757 and 0.4425, respectively. Our global registration method suppresses error accumulation and produces more accurate registration results.

We also evaluate our method on a clean partial dataset extracted from the *Bouncing* dataset [19]. Since the original models are complete, we extract the visible part of each complete model with a virtual camera rotating around the model. Here, we select 35 partial models and allow large deformations for the selected models. We use a multi-resolution approach. First, we get low-resolution models form the original partial models by downsampling them to $1/10$ of the full resolution. There are about 3,000-5,000 vertices for the original partial models and 300-500 vertices for low-resolution models. Then, we find the corresponding points between neighboring scans through the approach in [20]. Finally, we solve the global registration problem from coarse to fine (Eq. (19)).

Fig. 2 illustrates the results when scans are accumulated gradually, using the first 4 scans, the first 20 scans, and all the scans for registration. We use standard Poisson reconstruction [21] to obtain watertight meshes. It can be seen that the shrinkage problem becomes more and more severe for the right hand and right leg using the method [8]. Due to the accumulation of registration errors, the results of method [9] have clearly visible misalignment even after Poisson reconstruction, especially for the arms. The results of our method shown in Fig. 2 are smoother and better aligned, such as the left arm and the head. By using global registration, our method does not suffer from error accumulation, and the use of ARAP constraint avoids shrinking.

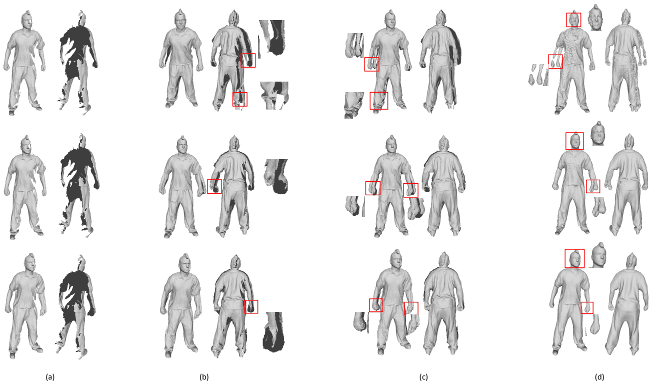

Fig. 2. Comparative results on gradually accumulated partial scans using the method [8] (top row), the method [9] (middle row), and our method (bottom row): (a) results of scans 1-4, (b) results of scans 1-20, (c) results of all the scans and (d) Poisson reconstruction results based on (c).

4.2. Results on Real Scans

We test our method on real scans, which are very challenging, because they have much noise and a large number of outliers. There is a dataset scanned using a Kinect v2.0: *Waving*. Fig. 3 illustrates the results for *Waving* dataset when scans are accumulated gradually. The results of [8] (top row) not only have serious shrinkage but also become more and more flat. With the the accumulation of registration error, the misalignment problem for the method [9] also becomes unacceptable, especially in the head and arms. Our method generates significantly better results, such as the head and arms.

5. CONCLUSIONS

This paper proposes a novel global sparse non-rigid alignment method which registers a sequence of scans with dramatic deformations simultaneously to reconstruct a complete object with a single RGB-D camera. We formulate the energy function with dual sparsity on both data term and smooth term, along with the local rigidity constraint and the ARAP (as-rigid-as-possible) constraint. It is solved by the alternating direction method under the augmented Lagrangian multiplier (ADM-ALM) framework which has exact solutions and guaranteed convergence. Experimental results on public datasets and real scanned datasets show that our method is effective and robust for challenging deformations, such as the large-scale movement of arms and legs. In addition, our method

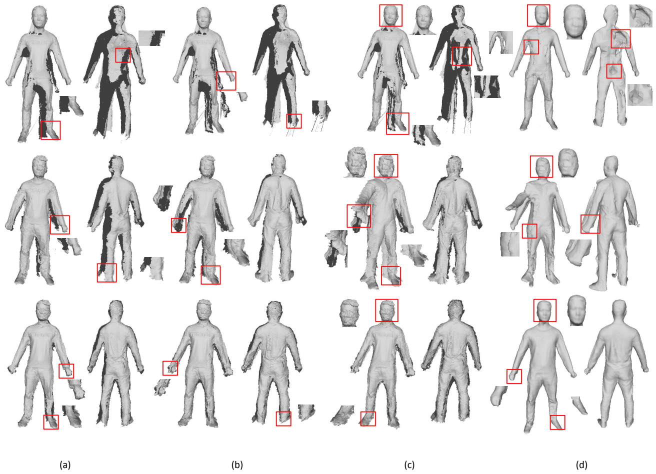

Fig. 3. Comparative results with scans accumulated gradually on the scanned *Waving* dataset using the method [8] (top row), the method [9] (middle row), and our method (bottom row): (a) results of scans 1-12 , (b) results of scans 1-20, (c) results of all the scans and (d) Poisson reconstruction results.

allows fewer partial scans to reconstruct a full object.

Our method has some limitations. Firstly, although our method can handle a wide range of deformations, it becomes more difficult with fewer scans, since neighboring scans have less overlap. Our current global registration implementation only considers neighboring scans. The results may not be ideal if some scans do not have sufficient overlap with adjacent scans. In the future, we will investigate more robust schemes by exploiting potential overlaps between non-adjacent scans.

6. REFERENCES

[1] Kaiwen Guo, Feng Xu, Yangang Wang, Yebin Liu, and Qionghai Dai, “Robust non-rigid motion tracking and surface reconstruction using L0 regularization,” in *Proc. ICCV*, 2015, pp. 3083–3091.

[2] 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.

[3] E. De Aguiar *et al.*, “Performance capture from sparse multi-view video,” *ACM TOG*, vol. 27, no. 3, pp. 15–19, 2008.

[4] R. A. Newcombe *et al.*, “KinectFusion: Real-time dense surface mapping and tracking,” in *IEEE ISMAR*, 2011, pp. 127–136.

[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. 643–50, 2012.

[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] M. Dou, J. Taylor, H. Fuchs, A. Fitzgibbon, and S. Izadi, “3D scanning deformable objects with a single RGBD sensor,” in *Proc. CVPR*, 2015, pp. 493–501.

[8] J. Yang, K. Li, K. Li, and Y.-K. Lai, “Sparse non-rigid registration of 3D shapes,” *Comp. Graph. Forum*, vol. 34, no. 5, pp. 89–99, 2015.

[9] J. Yang, K. Li, Y.-K. Lai, and D. Guo, “Robust non-rigid registration with reweighted dual sparsities,” *https://arxiv.org/abs/1703.04861*, 2016.

[10] J. Starck, A. Maki, S. Nobuhara, A. Hilton, and T. Matsuyama, “The multiple-camera 3-d production studio,” *IEEE TCSVT*, vol. 19, no. 6, pp. 856–869, 2009.

[11] M. Dou, H. Fuchs, and J. M. Frahm, “Scanning and tracking dynamic objects with commodity depth cameras,” in *IEEE ISMAR*, 2013, pp. 99–106.

[12] G. Ye, Y. Liu, N. Hasler, X. Ji, Q. Dai, and C. Theobalt, “Performance capture of interacting characters with handheld kinects,” in *Proc. ECCV*, 2012, pp. 828–841.

[13] M. Zollhöfer *et al.*, “Real-time non-rigid reconstruction using an RGB-D camera,” *ACM TOG*, vol. 33, no. 4, pp. 1–12, 2014.

[14] Y. Cui, W. Chang, T. Nöll, and D. Stricker, “KinectAvatar: Fully automatic body capture using a single kinect,” in *Proc. ACCV Workshops*, 2012, pp. 133–147.

[15] R. A. Newcombe, D. Fox, and S. M. Seitz, “DynamicFusion: Reconstruction and tracking of non-rigid scenes in real-time,” in *Proc. CVPR*, 2015, pp. 343–352.

[16] O. Sorkine and M. Alexa, “As-rigid-as-possible surface modeling,” *SGP*, pp. 109–116, 2007.

[17] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein, “Distributed optimization and statistical learning via the alternating direction method of multipliers,” *Foundations and Trends® in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[18] M. Garland and P. S. Heckbert, “Surface simplification using quadric error metrics,” in *ACM SIGGRAPH*, 1997, pp. 209–216.

[19] 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.

[20] 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.

[21] M. Kazhdan and H. Hoppe, “Screened poisson surface reconstruction,” *ACM TOG*, vol. 32, no. 3, pp. 29, 2013.