Global 3D Non-Rigid Registration of Deformable Objects Using a Single RGB-D Camera

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

***Abstract*— We present a novel global non-rigid registration method for dynamic 3D objects. Our method allows objects to undergo large non-rigid deformations and achieves high-quality results even with substantial pose change or camera motion between views. In addition, our method does not require a template prior and uses less raw data than tracking-based methods since only a sparse set of scans is needed. We simultaneously compute the deformations of all the scans 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 shapes, especially near open boundaries of scans. To cope with 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. The experimental results demonstrate that the proposed method obtains better results than several state-of-the-art methods.**

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

I. Introduction

DYNAMIC 3D reconstruction, which aims to recover dynamic scenes by capturing videos using a single or multiple cameras, becomes increasingly popular in computer graphics and computer vision [1]–[4]. With the availability 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 many applications [5], [6], such as 3D printing, gaming, and movie production, to name a few. However, reconstruction results by KinectFusion [7] for deformable objects have serious drifting artifacts, because a static model is generally assumed. Besides, the point clouds captured by depth cameras are usually polluted by serious 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 research groups have set up multi-camera systems [8], [20]. However, practical applications of such systems are limited, due to high cost, complex maintenance, and lack of portability. To reduce system sizes, some methods use fewer cameras with depth cues [9]–[11]. Using a single camera is cheaper but the problem becomes more ill-posed. A template prior is usually used to reduce the difficulty [3], [12]. However, using a prior template restricts the captured target as they can pre-scan *e.g.* a deformable object, but cannot model deformations beyond the prior template, such as facial expressions and loose clothing. Some methods try to achieve dynamic 3D reconstruction using a single camera without a template prior. Non-rigid structure from motion methods [13]–[15] aim at recovering dynamic 3D shapes from multi-frame 2D images, but they cannot deal with large scale cases. Alternative methods [16], [17] based on tracking and fusion of RGB-D sequences of non-rigidly deforming objects are proposed, but small deformation between two neighboring viewpoints (time instances) is generally assumed. To our knowledge, few work in the literature allows large motions of the subject between different viewpoints using a single camera, which happens commonly 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 without a template prior. The motion of the object between different viewpoints can be very large. Naive solutions of applying pairwise non-rigid registration in succession lead 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. We also 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 demonstrate that the proposed method obtains better results than the state-of-the-art methods.

Manuscript received November 13, 2017; revised September 19, 2018 and January 29, 2019; accepted March 18, 2019. Date of publication April 4, 2019; date of current version August 1, 2019. This work was supported in part by the National Natural Science Foundation of China under Grant 61571322 and Grant 61771339, and in part by the Tianjin Research Program of Application Foundation and Advanced Technology under Grant 18JCYBJC19200. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Gene Cheung. *(Corresponding author: Kun Li.)*

J. Yang and D. Guo are with the School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China.

K. Li and Z. Wu are with the Tianjin Key Laboratory of Cognitive Computing and Application, College of Intelligence and Computing, Tianjin University, Tianjin 300350, China (e-mail: lik@tju.edu.cn).

Y.-K. Lai is with the School of Computer Science and Informatics, Cardiff University, Cardiff CF10 3AT, U.K.

Digital Object Identifier 10.1109/TIP.2019.2909197

The main contributions of this work are summarized as follows:

- 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.

Preliminary results of this work were reported in a conference paper [18]. In addition to more thorough discussions and literature review, the algorithm details are now provided and experimental validation is substantially extended, including quantitative evaluation, evaluation on more datasets, and comparison with more methods. We first summarize previous work in Sec. II. Then, in Sec. III, we present our global registration framework, including several constraint terms and the overall optimization function. The solution to the optimization function is presented in Sec. IV. Finally, we provide experimental results in Sec. V and conclude this paper in Sec. VI.

II. Related Work

In this section, we review recent related work in 3D reconstruction.

A. 3D Reconstruction With Multi-Cameras

Several groups have set up multi-camera systems, in which drifting is not a concern because a relatively complete model is captured at each frame. Starck *et al.* [8] design a system to reconstruct a full human body using 16 cameras, which requires careful positioning of cameras to obtain better raw data. With multiple high-speed cameras, Vlasic *et al.* [19] design a system for high-resolution capture of moving 3D objects at high details using a photometric stereo light stage. Li *et al.* [20] 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. To reduce the number of required cameras, some methods use depth cameras. Tong *et al.* [9] scan a 3D full human body model using three Kinect cameras, but the method assumes that the person keeps still. Ye *et al.* [10] use three hand-held Kinects to reconstruct human skeletal poses, deforming surface geometry and camera poses by deforming template models, which generates relatively fine results. Dou *et al.* [11] scan and track deforming objects using fusion of dynamic input from an eight-Kinect rig, by deforming a human template. Collet *et al.* [21] use over 30 RGB-D cameras and a large studio setting with a green screen and controlled lighting to produce extremely high quality results. Lin *et al.* [22] optimize the placement of multiple Kinect sensors to achieve the desired scanning accuracy, leading to an effective configuration with 16 RGB-D cameras. Dou *et al.* [23] design an approach for live performance capture from eight RGB-D sequences, which is robust to large frame-to-frame motion and topology changes, and generates compelling reconstruction results in real-time.

B. 3D Reconstruction With a Single Camera

Considering the high cost and the difficult maintenance of multi-camera systems, monocular approaches become more and more popular. Some methods focus on scanning persons with a fixed pose. Weiss *et al.* [24] propose to estimate a parametric model of the human shape combining low-resolution image silhouettes with coarse range data. Cui *et al.* [25] capture a full 3D human body model using a single depth camera, which presents fine results, but limits the user to keep a 'T' pose. Li *et al.* [26] adopt a more general non-rigid registration framework which allows a wider range of poses, which demonstrates compelling results but still requires users to keep the *same* pose. Dou *et al.* [27] develop a 3D scanning system which allows a considerable amount of deformations during scanning and shows fine results. However, large deformation between two neighboring viewpoints (time instances) is not allowed.

To reconstruct a dynamic scene using a single camera, non-rigid structure from motion methods are used to recover dynamic 3D shapes from multi-frame 2D images [13]–[15]. However, these methods normally are not able to recover 3D shapes with a large number of vertices. To handle this, some methods capture a static pre-scan as a template prior [11]. Li *et al.* [28] reconstruct the geometry and motion of complex deforming shapes by using a smooth template that provides a crude approximation to the scanned object and serves as a geometric and topological prior. Zollhöfer *et al.* [12] use a single self-contained stereo camera unit to generate spatio-temporally coherent 3D models, which also starts by scanning a smooth template model of the object using KinectFusion and registers the template to the sequences. In particular, it is able to produce compelling reconstruction models for palms and faces. Guo *et al.* [3] reconstruct non-rigid geometry and motions from a single-view depth input captured by a depth sensor, which also uses a template prior and presents fine results. However, scanning a template model in advance is inconvenient and impractical for many applications.

Some methods try to achieve dynamic reconstruction using a single camera without a template prior. Liao *et al.* [29] reconstruct complete 3D deformable models over time by a single depth camera, which is able to reconstruct visually plausible 3D surface deformation results. However, it assumes that the deformation is continuous and predictable in a short temporal interval. Newcombe *et al.* [17] design a dense SLAM system, which is able to reconstruct non-rigidly deforming scenes in real-time. Dou *et al.* [16] propose a 360° performance capture system that can reconstruct arbitrary non-rigid scenes in real-time. However, all the methods have the same limitation that small deformation between the neighboring views (frames) is assumed.

C. Shape Registration for 3D Reconstruction

In 3D reconstruction, registration methods are developed to align scans from multiple views with substantial movement, including rigid and non-rigid registration methods [30]. The former assumes that the object only undergoes rigid body transformation [31]–[33], whereas the latter considers more general deformable models, and is thus more suitable for reconstruction of deformable objects. Typical non-rigid registration methods [34] generalize the iterative closest point (ICP) method from rigid registration, and follow a similar paradigm that alternately optimizes correspondences based on the closest point criterion and local transformations according to the updated correspondences. Recent work introduces various effective data and regularization terms to the non-rigid ICP framework [35]–[39] to improve accuracy and robustness of registration. However, such existing non-rigid registration methods are based on pairwise registration. Applying them to multiple scans in succession leads to error accumulation and the loop closure problem, i.e., when a sequence of scans forms a loop, the last scan fails to align with the first scan due to accumulated drifting.

In this paper, we propose a global non-rigid registration framework based on sparse priors as they are robust to noise and outliers. Multiple scans are aligned simultaneously, which effectively handles error accumulation and avoids the loop closure problem. We further introduce an ARAP constraint to the global non-rigid registration framework to eliminate the shrinking problem, which is more critical for partial scans with limited overlaps, and design a coarse-to-fine multi-resolution scheme to avoid the optimization being trapped into local minima and help to attack large scale problems. Our method only requires sparse views as input and allows large scale deformations of the object during scanning.

III. The Proposed Method

A. Iterative Framework

The aim of global non-rigid registration is to find a set of non-rigid transformations $\mathbf{X}$ that transforms scans for consistent alignment. To this end, an iterative procedure is applied with the following two alternating steps:

Step 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. At the first iteration, we use a technique based on local geometric similarity and diffusion pruning of inconsistent correspondence [40] as it often provides reliable correspondences. Alternative correspondence techniques or manual specification of a few correspondences may instead be used. At other iterations, we update the correspondences by using the closest points between two shapes to find additional correspondences similar to ICP.

Step 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 are to be registered.

B. 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 $\mathcal{U}^{(M+1)} \equiv \mathcal{U}^{(1)}$), let $f_{m\to m+1} : \{1, \cdots, N_m\} \mapsto \{1, \cdots, N_{m+1}\}$ 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 $\mathbf{X}^{(m)} \triangleq \left[\mathbf{X}_1^{(m)}, \cdots, \mathbf{X}_{N_m}^{(m)}\right]^\top$ 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:

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

where $E_{\text{data}}(\mathbf{X})$ is the data term to measure the registration accuracy, $E_{\text{smooth}}(\mathbf{X})$ is the smoothness term to measure the smoothness of local transformations, $E_{\text{rig}}(\mathbf{X})$ is the orthogonality term to measure the rigidness of local transformations and $E_{\text{arap}}(\mathbf{X})$ 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, $f_{m\to m+1}$ (denoted by f_m hereafter for short), between a neighboring pair of overlapping scans $\mathcal{U}^{(m)}$ and $\mathcal{U}^{(m+1)}$. As neighboring surfaces only have partial overlaps, not every point has a corresponding point. Let K_m be the number of corresponding points between $\mathcal{U}^{(m)}$ and $\mathcal{U}^{(m+1)}$, where $K_m \le \min(N_m, N_{m+1})$. For the correspondence mapping f_m, let $f_m(i,1)$ and $f_m(i,2)$ be 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 $\mathcal{U}^{(m+1)}$:

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

where $\|\cdot\|_1$ denotes ℓ_1 norm of a matrix considered as a long vector. The right hand side of the data term (2) can be rewritten as

$$\sum_m \left\| \left(\mathbf{U}_{f_{m,1}}^{(m)} \mathbf{X}^{(m)} - \mathbf{U}_{f_{m,2}}^{(m+1)} \mathbf{X}^{(m+1)} \right) \right\|_1,$$

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

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

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

$$\mathbf{H} = \begin{bmatrix} \mathbf{U}_{f_{1,1}}^{(1)} & -\mathbf{U}_{f_{1,2}}^{(2)} & & & 0 \\ & \mathbf{U}_{f_{2,1}}^{(2)} & -\mathbf{U}_{f_{2,2}}^{(3)} & & \\ & & \ddots & \ddots & \\ & 0 & & \mathbf{U}_{f_{M-1,1}}^{(M-1)} & -\mathbf{U}_{f_{M-1,2}}^{(M)} \\ -\mathbf{U}_{f_{M,2}}^{(1)} & & & & \mathbf{U}_{f_{M,1}}^{(M)} \end{bmatrix}. \tag{4}$$

Smoothness Term: Similar to the pairwise registration, we define the edge set with a neighboring system: For a 3D mesh, the edges are simply defined by the edges of the mesh; for a 3D point set, it can be transformed to a mesh, or the edges can be defined by connecting each point with its K-nearest neighbors (K is typically set to 6). 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 $\mathcal{E}^{(m)} = \left\{e_{ij}^{(m)} \mid \mathbf{u}_j^{(m)} \in \mathcal{N}_i^{(m)}, \mathbf{u}_i^{(m)} \in \mathcal{U}^{(m)}\right\}$. Smoothness is regularized by the ℓ_1 norm of transformation differences on the neighboring system over all the scans $\mathcal{U}^{(m)}$ [39]:

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

which is rewritten into the matrix form:

$$E_{\text{smooth}}(\mathbf{X}) = \sum_m \left\|\mathbf{B}^{(m)}\mathbf{X}^{(m)}\right\|_1. \tag{6}$$

In (6), $\mathbf{B}^{(m)}$ is a sparse matrix, where each row contains only two groups of nonzero entries. For example, assuming the r^{th} row is associated with edge $e_{ij}^{(m)}$, then the entries linked to the reference vertex $\mathbf{u}_i^{(m)}$ are set to $(x_i^{(m)}, y_i^{(m)}, z_i^{(m)}, 1)$, while the ones linked to the neighboring vertex $\mathbf{u}_j^{(m)}$ are set to $(-x_j^{(m)}, -y_j^{(m)}, -z_j^{(m)}, -1)$. Let $\mathbf{B} \triangleq \text{diag}\left(\mathbf{B}^{(1)}, \ldots, \mathbf{B}^{(M)}\right)$, and we have the following form of the smoothness term:

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

Orthogonality Term: In non-rigid registration, each vertex 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 still 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 locally 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 [39]:

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

where $\mathbf{D} = \begin{bmatrix}\mathbf{I}_3 & \mathbf{0}_{3\times 1}\end{bmatrix}$ is a constant 3×4 matrix 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.

As-rigid-as-possible (ARAP) Term: We observe that some vertices of the registered surfaces may have inward shrinkage, 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 $\mathbf{p}_i^{(m)} \triangleq (x_i^{(m)}, y_i^{(m)}, z_i^{(m)})$ is the vertex position of $\mathcal{U}^{(m)}$. We define the ARAP term as follows, similar to [12], [41]:

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

where $w_i^{(m)} = 1$ for vertices with known correspondence and $w_i^{(m)} = 0$ otherwise, and $\mathbf{T}_i{}^{(m)} \in \mathbb{R}^{3\times 3}$ is a rotation matrix. The cotangent weight $w_{ij}^{(m)}$ is is used to reduce mesh discretization bias:

$$w_{ij}^{(m)} = \frac{1}{2}(\cot\alpha_{ij} + \cot\beta_{ij}), \tag{10}$$

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). 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

$$\mathbf{S}_i^{(m)} = \sum_m \sum_{j \in \mathcal{N}(i)} w_{ij}^{(m)} \mathbf{e}_{ij}^{(m)} \mathbf{e}_{ij}^{'(m)\top}. \tag{11}$$

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

$$\mathbf{T}_i^{(m)} = \mathbf{V}_i^m \mathbf{U}_i^{(m)\top}. \tag{12}$$

To minimize E_{arap} *w.r.t.* $\mathbf{X}_i^{(m)}$, similar to $\mathbf{h}_{m,i}$, we denote $\mathbf{h}_{m,j} = \mathbf{e}_{ji}^{'(m)} - \mathbf{e}_{ji}^{(m)}\mathbf{T}_j{}^{(m)}$. Then, we first work out $\frac{\partial E_{\text{arap}}}{\partial \mathbf{p}_i^{'(m)}}$ where

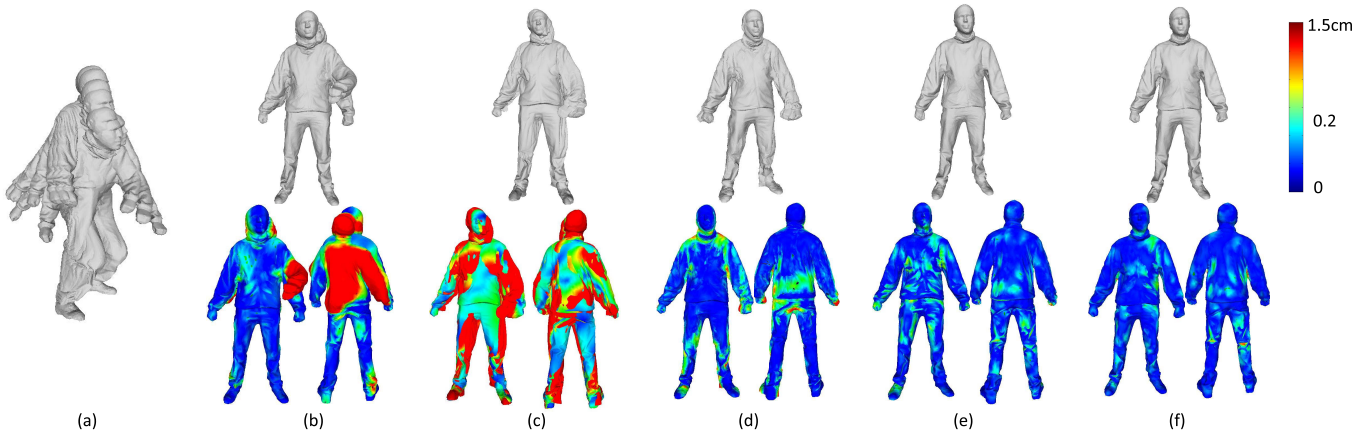


Fig. 1. Comparison results on the *Jumping* dataset: (a) original complete models, (b) the results of [3], (c) the results of PR-GLS [4], (d) the results of [38], (e) the results of [39], and (f) our results.

$\mathbf{p}_i^{'(m)} = \mathbf{u}_i^{(m)}\mathbf{X}_i^{(m)}$ is the transformed vertex position as

$$\begin{aligned}\frac{\partial E_{\text{arap}}}{\partial \mathbf{p}_i^{'(m)}} &= \frac{\partial}{\partial \mathbf{p}_i^{'(m)}}\left(\sum_{\mathbf{j}\in\mathcal{N}(i)} w_{ij}^{(m)}\left\|\mathbf{h}_{m,i}\right\|^2 + \sum_{\mathbf{j}\in\mathcal{N}(i)} w_{ji}^{(m)}\left\|\mathbf{h}_{m,j}\right\|^2\right) \\ &= \sum_{\mathbf{j}\in\mathcal{N}(i)} 2w_{ij}^{(m)}\mathbf{h}_{m,i} + \sum_{\mathbf{j}\in\mathcal{N}(i)} -2w_{ji}^{(m)}\mathbf{h}_{m,j}. \end{aligned} \tag{13}$$

Using the fact that $w_{ij}^{(m)} = w_{ji}^{(m)}$, we obtain

$$\frac{\partial E_{\text{arap}}}{\partial \mathbf{p}_i^{'(m)}} = \sum_{\mathbf{j}\in\mathcal{N}(i)} 4w_{ij}^{(m)}\left(\mathbf{e}_{ij}^{'(m)} - \frac{1}{2}\mathbf{e}_{ij}^{(m)}(\mathbf{T}_i{}^{(m)} + \mathbf{T}_j{}^{(m)})\right). \tag{14}$$

Setting the partial derivatives to zero leads to the following:

$$\sum_{j\in\mathcal{N}(i)} w_{ij}^{(m)}\mathbf{e}_{ij}^{'(m)} = \sum_{j\in\mathcal{N}(i)} \frac{w_{ij}^{(m)}}{2}\mathbf{e}_{ij}^{(m)}\left(\mathbf{T}_i^{(m)} + \mathbf{T}_j^{(m)}\right). \tag{15}$$

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

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

where $\mathbf{L}^{(m)}$ represents the linear combination on the left-hand side of (15), which is the discrete Laplace-Beltrami operator. $\mathbf{b}^{(m)}$ is an n-vector whose i^{th} row contains the right-hand side of (15). The definition of E_{arap} in Eq. (9) depends on both the deformed edges $\mathbf{e}_{ij}^{'(m)}$'s and rotations $\mathbf{T}_i^{(m)}$'s. In our setting, the former are determined by transformations $\mathbf{X}$, which are optimized in the overall energy along with other terms, so we optimize $\mathbf{T}_i^{(m)}$'s dedicated to the energy term such that the formulation (16) only depends on $\mathbf{X}$.

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

$$E_{\text{arap}}(\mathbf{X}) = \left\|\mathbf{L}\mathbf{X} - \mathbf{b}\right\|_F^2. \tag{17}$$

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:

$$\begin{aligned}\min_{\mathbf{X},\mathbf{C},\mathbf{A}} &\left\{\begin{array}{c}\|\mathbf{C}\|_1 + \alpha\|\mathbf{A}\|_1 + \beta\|\mathbf{L}\mathbf{X} - \mathbf{b}\|_F^2 \\ +\lambda\sum_m\sum_i \\ \left\|\mathbf{D}\mathbf{X}_i^{(m)} - \mathbf{R}_i^{(m)}\right\|_F^2\end{array}\right\} \\ \text{s.t. } &\mathbf{C} = \mathbf{H}\mathbf{X}, \mathbf{A} = \mathbf{B}\mathbf{X}, \\ &\mathbf{R}_i^{(m)\top}\mathbf{R}_i^{(m)} = \mathbf{I}_3, , m = 1, \ldots, M, \\ &\det(\mathbf{R}_i^{(m)}) > 0,\end{aligned} \tag{18}$$

where $\mathbf{A}$ and $\mathbf{C}$ are auxiliary variables to facilitate optimization. Then, we solve the constrained minimization (18) using the augmented Lagrangian method (ALM) (see the next subsection for details).

Multi-resolution Approach: Since the transformation $\mathbf{X}_i$ of each vertex i has a rotation $\mathbf{R}_i \in \mathbb{R}^{3\times3}$ and a translation $\mathbf{t}_i \in \mathbb{R}^3$, there are 12 degrees of freedom (DoFs) in total for each $\mathbf{X}_i$. If a scan m has N_m vertices, there are N_m transformations and $12N_m$ DoFs. However, even if each vertex has a positional constraint (with 3 constraints), there are $3N_m$ constraints in total, which are not enough to identify a unique solution. 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 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 s^{th} scale of the shape via standard downsampling [42]. 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. (18) at scale s can be rewritten as:

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

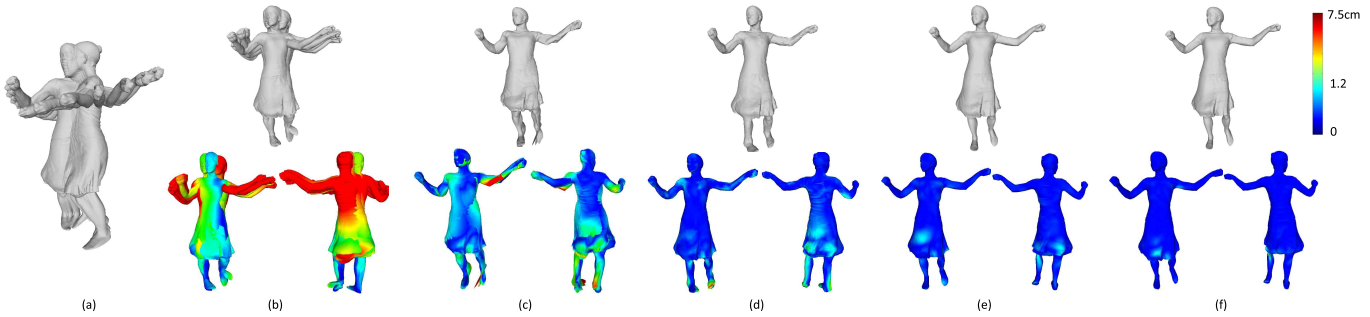


Fig. 2. Comparison results on the *Swing* dataset: (a) original complete models, (b) the results of [3], (c) the results of PR-GLS [4], (d) the results of [38], (e) the results of [39], and (f) our results.

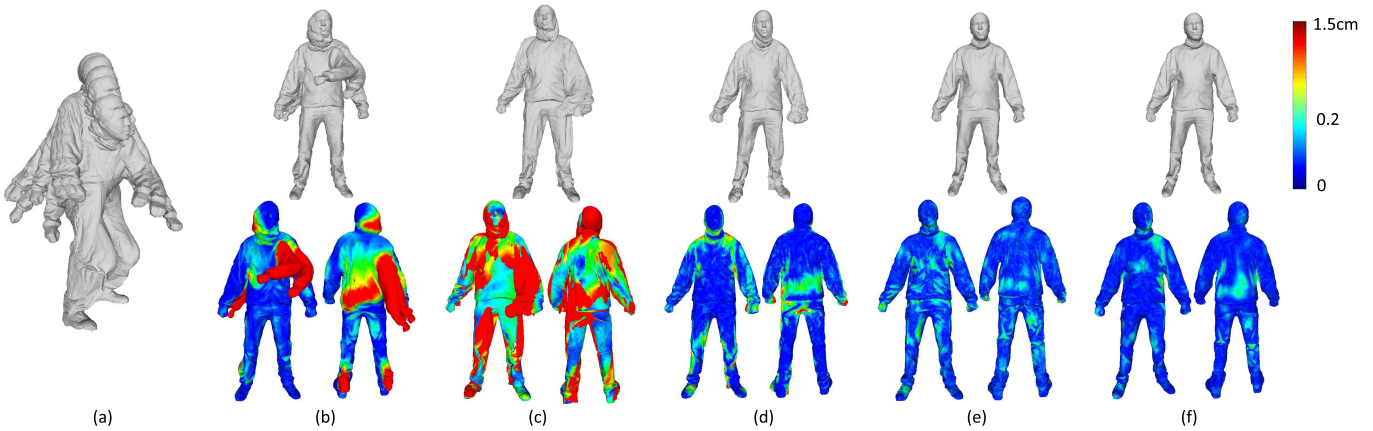


Fig. 3. Comparison results on the *Jumping* dataset with noise ($\sigma = 0.1\bar{l}$): (a) original complete models with noise, (b) the results of [3], (c) the results of PR-GLS [4], (d) the results of [38], (e) the results of [39], and (f) our results.

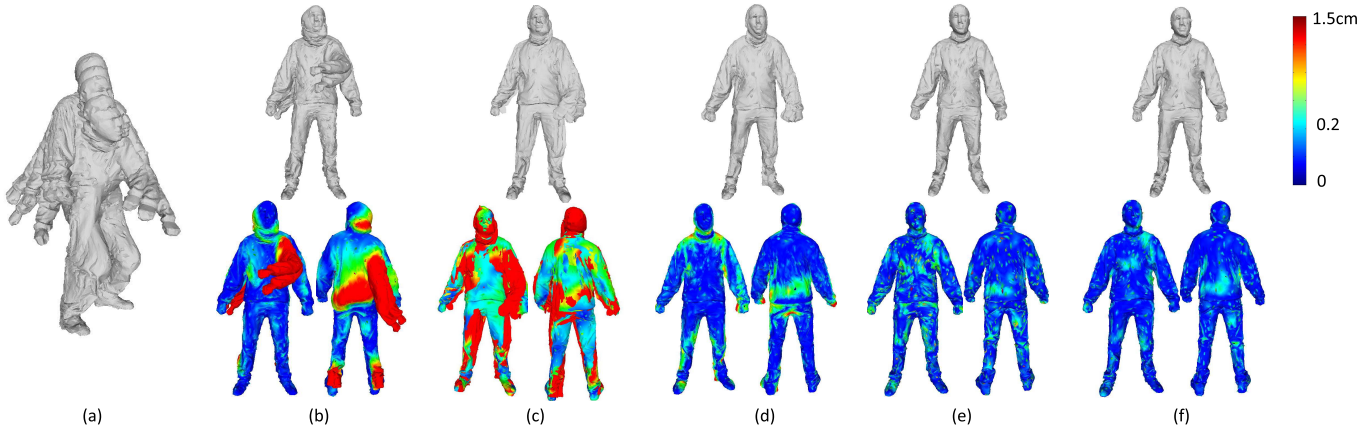


Fig. 4. Comparison results on the *Jumping* dataset with 10% outliers: (a) original complete models with outliers, (b) the results of [3], (c) the results of PR-GLS [4], (d) the results of [38], (e) the results of [39], and (f) our results.

$$\begin{aligned}\mathbf{R}_i^{(m,s)^\top}\mathbf{R}_i^{(m,s)} &= \mathbf{I}_3,\\ \det(\mathbf{R}_i^{(m,s)}) &> 0,\end{aligned}, m = 1,\ldots,M,$$

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

We start our multi-resolution method from the coarsest scale S to solve the optimization problem (19), and use the solution at previous scale as initialization to solve the transformation at next scale until reaching the finest scale. Denote by $\mathbf{u}_i^{(m,s-1)}$ a vertex at the $(s-1)^{\text{th}}$ scale, and by $\Gamma_i^{(m,s)}$ the index set of vertices of $\mathbf{u}_i^{(m,s-1)}$ at the s^{th} scale within the spherical neighborhood. The corresponding deformation $\mathbf{X}_i^{(m,s-1)}$ is estimated by a weighted average of the deformations of vertices within a spherical neighborhood of radius r at the s^{th} scale [28]:

$$\mathbf{X}_i{}^{(m,s-1)} = \sum_{j\in\Gamma_i^{(m,s)}} m_{i,j}\mathbf{X}_j^{(m,s)}. \tag{20}$$

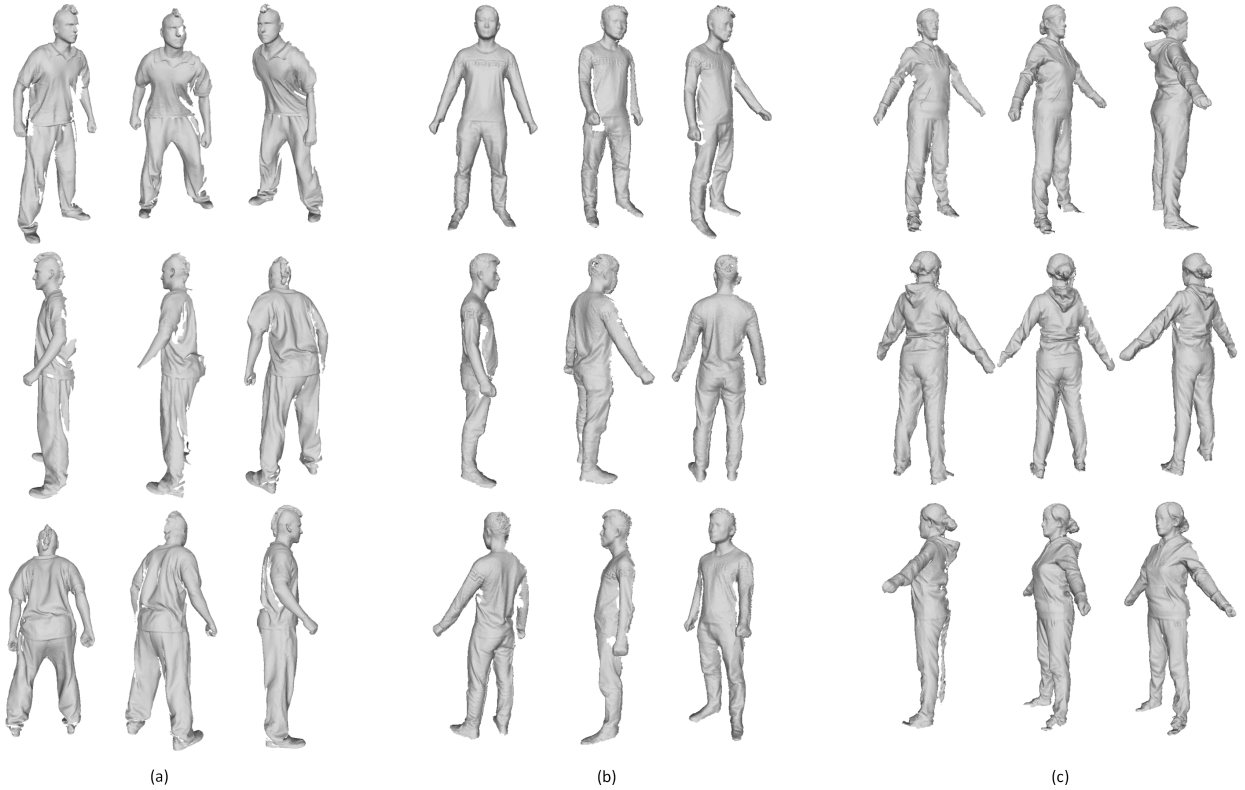


Fig. 5. Sample partial models of test datasets: (a) *Bouncing*, (b) *Waving* and (c) *Flying*.

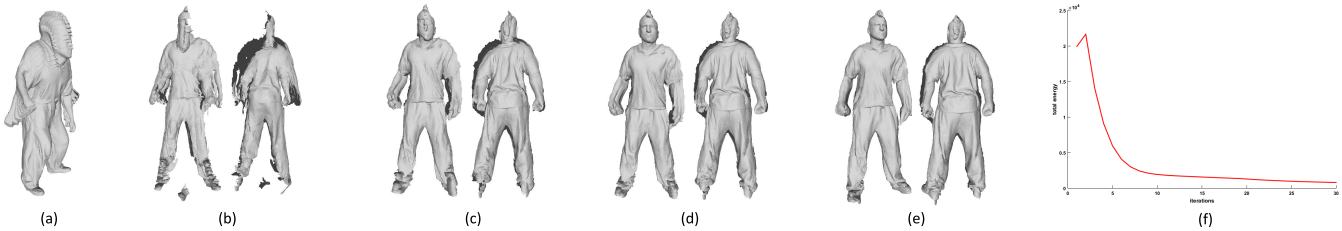


Fig. 6. Iterative results of our method. (a) initial 37 partial scans, (b) registration result after 1 iteration, (c) registration result after 10 iterations, (d) registration result after 20 iterations, (e) registration result after 30 iterations, and (f) total energy vs. the number of iterations.

The weight $m_{i,j}$ showing the contribution of the transformation of $\mathbf{u}_i^{(m,s-1)}$ to that of $\mathbf{u}_j^{(m,s)}$ is defined as

$$\begin{aligned} m_{i,j} &= \max\left(0, \mathbf{d}_{i,j}\right), \\ \mathbf{d}_{i,j} &= 1 - d^2\left(\mathbf{u}_i^{(m,s-1)}, \mathbf{u}_j^{(m,s)}\right) \Big/ r^2, \end{aligned} \tag{21}$$

where r is the effective radius. In our experiment, the radius is set to be twice of the average edge length of the coarser mesh. $d\left(\mathbf{u}_i^{(m,s-1)}, \mathbf{u}_j^{(m,s)}\right)$ represents the Euclidean distance between $\mathbf{u}_i^{(m,s-1)}$ and $\mathbf{u}_j^{(m,s)}$. The weight drops steadily with an increasing distance. Using the matrix notation, we can rewrite Eq. (20) as:

$$\mathbf{X}^{(m,s-1)} = \mathbf{M}^{(m)}\mathbf{X}^{(m,s)}, \tag{22}$$

where $\mathbf{M}^{(m)}$ is constructed by collecting $m_{i,j}$, and therefore $\mathbf{X}^{(m,s-1)}$ is obtained by a linear combination of relevant transformations of $\mathbf{X}^{(m,s)}$. Then, we have:

$$\mathbf{X}^{(s-1)} = \mathbf{M}\mathbf{X}^{(s)}, \tag{23}$$

where $\mathbf{M}$ contains the mapping of transformations from $\mathcal{U}^{(m,s)}$ to $\mathcal{U}^{(m,s-1)}$ for all the scans. Using Eq. (23) to replace $\mathbf{X}$ in Eq. (18), we can get the optimization function Eq. (19).

By using the coarse-to-fine strategy, our method avoids the optimization being trapped into local minima, and handles high resolution shapes effectively. In our implementation, the global registration is performed at two scales in which the coarser-scale shapes have 500∼1000 vertices.

IV. The Proposed ADM-ALM Algorithm

Our numerical algorithm is derived based on the ADM-ALM framework due to the following three appealing merits: 1) convenient handling of equality constraints,

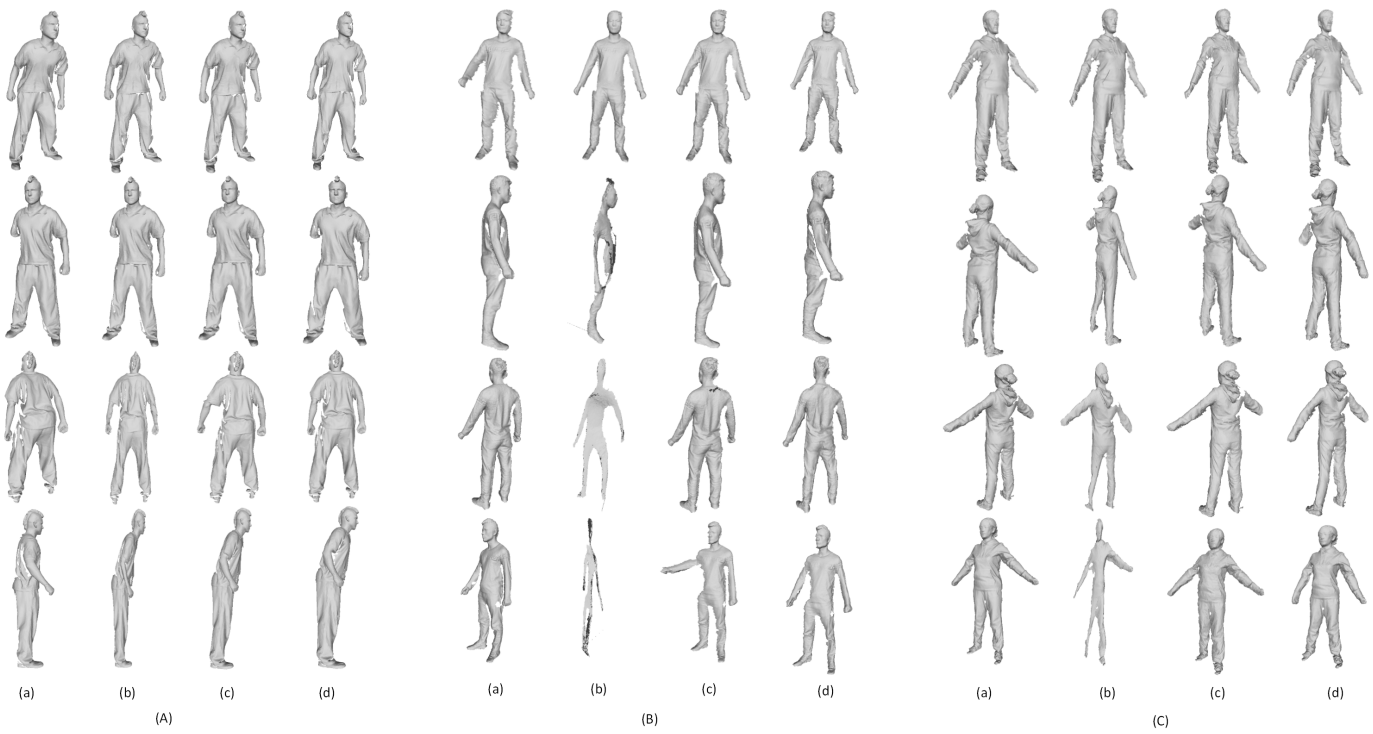


Fig. 7. Comparison of results on datasets: (A) *Bouncing*, (B) *Waving* and (C) *Flying*: (a) original partial scans, (b) the results of [38], (c) the results of [39] and (d) the results of our method.

2) flexible adaptability to large-scale problems with multiple sets of variables, and 3) proven numerical convergence for various minimization models across a wide range of applications.

The ALM method converts the original problem (18) to the iterative minimization of its augmented Lagrangian function:

$$\begin{aligned} L(\mathbf{X},\mathbf{C},\mathbf{A},\mathbf{Y}_1,\mathbf{Y}_2,\mu_1,\mu_2) = {} & \left\|\mathbf{C}\right\|_1 + \alpha\left\|\mathbf{A}\right\|_1 \\ & + \langle \mathbf{Y}_1, \mathbf{C}-\mathbf{H}\mathbf{X}\rangle \\ & + \frac{\mu_1}{2}\left\|\mathbf{C}-\mathbf{H}\mathbf{X}\right\|_F^2 \\ & + \langle \mathbf{Y}_2, \mathbf{A}-\mathbf{B}\mathbf{X}\rangle \\ & + \frac{\mu_2}{2}\left\|\mathbf{A}-\mathbf{B}\mathbf{X}\right\|_F^2 \\ & + \lambda\sum_m\sum_i\left\|\mathbf{D}\mathbf{X}_i^{(m)}-\mathbf{R}_i^{(m)}\right\|_F^2 \\ & + \beta\left\|\mathbf{L}\mathbf{X}-\mathbf{b}\right\|_F^2, \\ \text{s.t.}\quad & \mathbf{R}_i^{(m)\top}\mathbf{R}_i^{(m)} = \mathbf{I}_3, \det(\mathbf{R}_i^{(m)}) > 0, \end{aligned} \tag{24}$$

where (μ_1, μ_2) are positive constants, $(\mathbf{Y}_1, \mathbf{Y}_2)$ are Lagrangian multipliers, and $\langle\cdot,\cdot\rangle$ denotes the inner product of two matrices considered as long vectors. Under the standard ALM framework, $(\mathbf{Y}_1, \mathbf{Y}_2)$ and (μ_1, μ_2) can be efficiently updated. However, each iteration has to solve $\mathbf{X}$, $\mathbf{C}$ and $\mathbf{A}$ simultaneously, which is difficult and computationally demanding. Hence, we resort to the alternate direction method (ADM) [43] to optimize $\mathbf{A}$, $\mathbf{C}$ and $\mathbf{X}$ separately at each iteration:

$$\begin{cases} \mathbf{C}^{(k+1)} = \arg\min_{\mathbf{C}}\ \|\mathbf{C}\|_1 + \left\langle \mathbf{Y}_1^{(k)}, \mathbf{C}-\mathbf{H}\mathbf{X}^{(k)}\right\rangle \\ \qquad\qquad + \frac{\mu_1^{(k)}}{2}\left\|\mathbf{C}-\mathbf{H}\mathbf{X}^{(k)}\right\|_F^2, \\ \mathbf{A}^{(k+1)} = \arg\min_{\mathbf{A}}\ \alpha\|\mathbf{A}\|_1 + \left\langle \mathbf{Y}_2^{(k)}, \mathbf{A}-\mathbf{B}\mathbf{X}^{(k)}\right\rangle \\ \qquad\qquad + \frac{\mu_2^{(k)}}{2}\left\|\mathbf{A}-\mathbf{B}\mathbf{X}^{(k)}\right\|_F^2, \\ \mathbf{X}^{(k+1)} = \arg\min_{\mathbf{X}}\ \left\langle \mathbf{Y}_1^{(k)}, \mathbf{C}-\mathbf{H}\mathbf{X}^{(k)}\right\rangle \\ \qquad\qquad + \frac{\mu_1^{(k)}}{2}\left\|\mathbf{C}-\mathbf{H}\mathbf{X}^{(k)}\right\|_F^2 + \left\langle \mathbf{Y}_2^{(k)}, \mathbf{A}-\mathbf{B}\mathbf{X}^{(k)}\right\rangle \\ \qquad\qquad + \frac{\mu_2^{(k)}}{2}\left\|\mathbf{A}-\mathbf{B}\mathbf{X}^{(k)}\right\|_F^2 + \beta\left\|\mathbf{L}\mathbf{X}^{(k)}-b\right\|_F^2 \\ \qquad\qquad + \lambda\sum_m\sum_i\left\|\mathbf{D}\mathbf{X}_i^{(m)(k)} - \mathbf{R}_i^{(m)(k)}\right\|_F^2, \\ \mathbf{R}_i^{(k+1)} = \arg\min_{\mathbf{R}_i}\ \lambda\left\|\mathbf{D}\mathbf{X}_i^{(k)}-\mathbf{R}_i\right\|_F^2 \\ \qquad\qquad \text{s.t.}\quad \mathbf{R}_i^\top\mathbf{R}_i = \mathbf{I}_3, \det(\mathbf{R}_i) > 0, \\ \mathbf{Y}_1^{(k+1)} = \mathbf{Y}_1^{(k)} + \mu_1^{(k)}\left(\mathbf{C}^{(k+1)} - \mathbf{H}\mathbf{X}^{(k)}\right), \\ \mathbf{Y}_2^{(k+1)} = \mathbf{Y}_2^{(k)} + \mu_2^{(k)}\left(\mathbf{A}^{(k+1)} - \mathbf{B}\mathbf{X}^{(k)}\right), \\ \mu_1^{(k+1)} = \rho_1\mu_1^{(k)}, \quad \rho_1 > 1, \\ \mu_2^{(k+1)} = \rho_2\mu_2^{(k)}, \quad \rho_2 > 1. \end{cases} \tag{25}$$

The $\mathbf{C}$-subproblem has the following closed solution:

$$\mathbf{C}^{(k+1)} = \text{shrink}\left(\mathbf{H}\mathbf{X}^{(k)} - \frac{1}{\mu_1^{(k)}}\mathbf{Y}_1^{(k)}, \frac{1}{\mu_1^{(k)}}\right), \tag{26}$$

where shrink$(\cdot,\cdot)$ is the shrinkage function applied to the matrix element-wisely:

$$\text{shrink}\,(x,\tau) = \text{sign}(x)\max(|x|-\tau,0). \tag{27}$$

The $\mathbf{A}$-subproblem is solved in a similar way:

$$\mathbf{A}^{(k+1)} = \text{shrink}\left(\mathbf{B}\mathbf{X}^{(k)} - \frac{1}{\mu_2^{(k)}}\mathbf{Y}_2^{(k)}, \frac{\alpha}{\mu_2^{(k)}}\right). \tag{28}$$

The $\mathbf{R_i}$-subproblem is solved as follows:

$$\begin{aligned}\mathbf{R}_i^{(k+1)} &= \mathbf{U}\mathbf{I}_3\mathbf{V}^\top,\\ (\mathbf{U},\mathbf{D},\mathbf{V}^\top) &= \text{svd}\left(\mathbf{D}\mathbf{X}_i^{(k+1)}\right).\end{aligned} \tag{29}$$

Being quadratic, the $\mathbf{X}$-subproblem is equivalent to solving the following normal equations according to the first-order optimality condition:

$$\begin{aligned}\mathbf{X}^{(k+1)} &= \mathbf{M}^{-1}\mathbf{g},\\ \mathbf{M} &\triangleq \mu_1^{(k)}\mathbf{H}^\top\mathbf{H} + \mu_2^{(k)}\mathbf{B}^\top\mathbf{B} + \beta\mathbf{L}^\top\mathbf{L} + \lambda\sum_i \mathbf{D}^\top\mathbf{D},\\ \mathbf{g} &\triangleq \mathbf{B}^\top\left(\mathbf{Y}_2^{(k)} + \mu_2^{(k)}\mathbf{A}^{(k+1)}\right)\\ &\quad + \mathbf{H}^\top\left(\mathbf{Y}_1^{(k)} + \mu_1^{(k)}\mathbf{C}^{(k+1)}\right)\\ &\quad + \beta\mathbf{L}^\top\mathbf{b} + \lambda\sum_i \mathbf{D}^\top\mathbf{R}_i^{(k)}.\end{aligned} \tag{30}$$

Note that $\mathbf{M}$ is diagonally-dominant and sparse. We use the sparse solver for linear equations in Matlab to solve the above normal equation. Compared with existing ADM-ALM approaches, our optimization solution has the following major differences: On the one hand, the proposed global registration model incorporates an as-rigid-as-possible (ARAP) term to avoid the collapse of the registered 3D model. Then the $\mathbf{X}$-subproblem involves the Laplace-Beltrami operator in the normal equations. On the other hand, although the compact form of objective function is similar to [38], [39], the algorithm in this work simultaneously optimizes all available partial scans in a global manner. As a result, the construction of the matrix $\mathbf{H}$ in the data term is different from those in [38] and [39]. Also, the proposed model has a large problem scale, and is suitable to be solved by the ADM-ALM framework.

The convergence can be determined by checking if the change of energy is below a threshold. In practice, we find that the setting (5 iterations of the outer loop each involving 25 iterations of the inner loop) is often sufficient to converge to decent results.

V. Experimental Results

In this section, we evaluate the performance of the proposed method on public complete datasets (Section V-A), partial datasets (Section V-B) and real scans (Section V-C).

TABLE I
Quantitative Evaluation for Whole-to-Whole Registration (cm)

Method	$Swing_{clean}$	$Jumping_{clean}$	$Jumping_{noise}$	$Jumping_{outlier}$
L0 [3]	5.2758	1.0162	1.1011	1.1984
PR-GLS [4]	1.1154	1.8122	1.8646	1.9780
L21 [41]	0.5621	0.7878	0.8003	0.8160
L11 [42]	0.2181	0.1392	0.2033	0.2245
Ours	**0.1224**	**0.1281**	**0.1466**	**0.1595**

A. Results on Public Complete Datasets

Firstly, we evaluate the proposed method on the *Jumping* and *Swing* datasets [19], which contain complete models with dramatic deformations. They have known correspondences to allow quantitative evaluation. Fig. 1 and Fig. 2 show the alignment results, compared with four pairwise registration methods [3], [38], [39]. Considering that these methods only register two models, we apply them to 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. The original 10 complete models with very different poses are shown in Fig. 1(a) and Fig. 2(a). The registration errors between the deformed model and the reference model are color-coded on the reference model for visual inspection. Here, the corresponding distance errors are computed using the standard Metro tool [44]. It can be seen that the results of methods [3], [4], [38] have visible misalignment due to error accumulation. Both the results of method [39] and our method are visually well aligned, but our method has smaller error. Table I gives quantitative evaluation with average errors over all the frames. Our method has the smallest errors which demonstrates that our global registration method suppresses error accumulation and produces more accurate registration results.

In order to evaluate the robustness of our method, we also experiment on the dataset with added noise and outliers. For the first test, each vertex is perturbed with Gaussian noise along the normal directions, with the mean set to zero and the deviation σ set as $0.1\overline{l}$ in our experiments, where $\overline{l}$ is the average length of triangle edges on all meshes. Fig. 3 demonstrates the alignment results compared with other three methods. We also evaluate the methods on the dataset with outliers in Fig. 4, in which 10% vertices are perturbed by Gaussian noise. As shown in Fig. 3 and Fig. 4, our method is more robust to noise and outliers than the other three methods. The corresponding distance errors are also shown in Table I.

B. Results on Partial Datasets

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. We select 37 partial models and allow large deformations among the selected models. Sample partial models of *Bouncing* are shown in Fig. 5(a). We use a multi-resolution approach for improved robustness and efficiency.

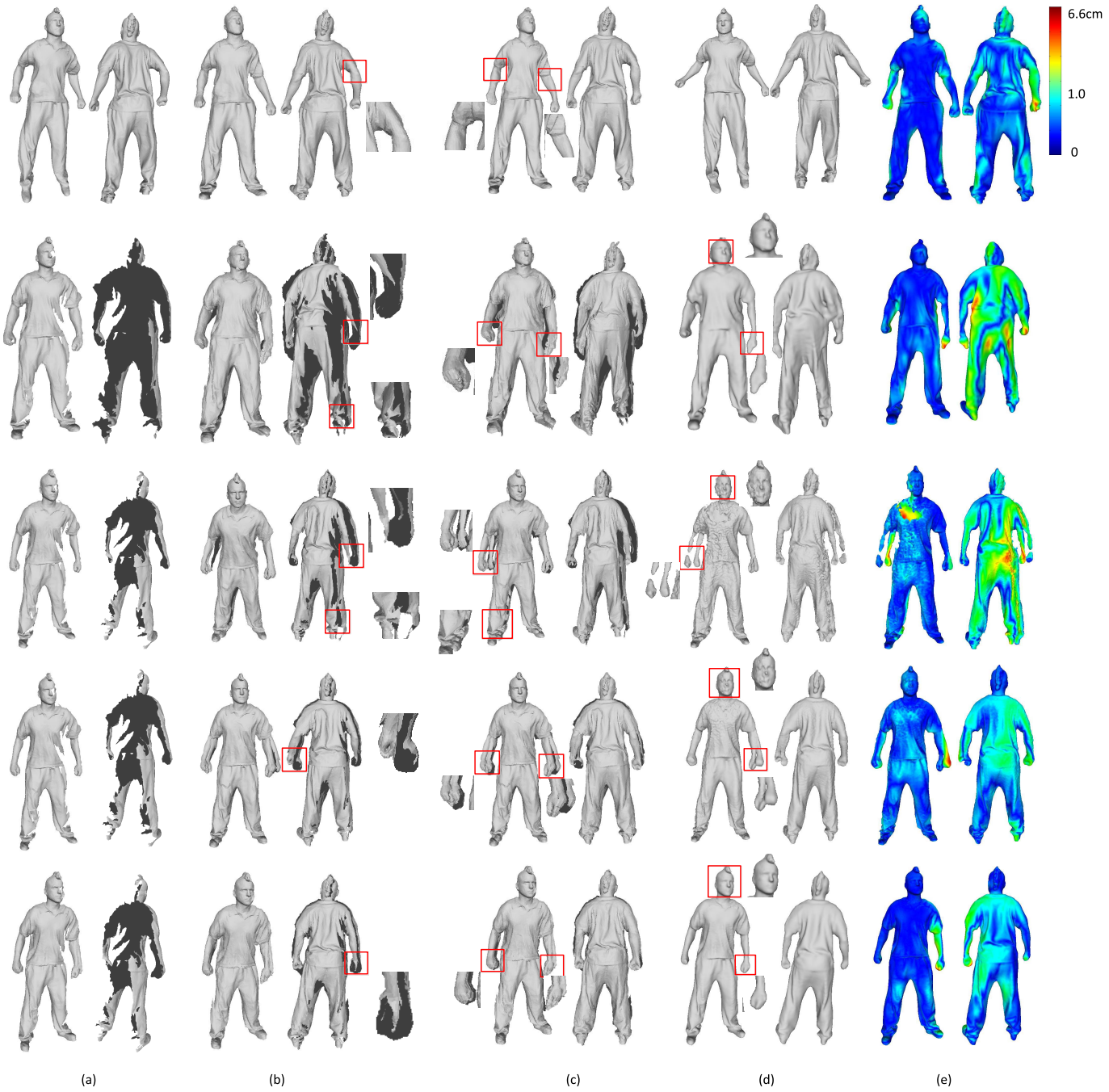


Fig. 8. Comparative results using gradually accumulated scans on the partial dataset *Bouncing*. Top row: L0 [3], second row: PR-GLS [4], third row: L21 [38], fourth row: L11 [39], and bottom row: our method. (a): the results of scans 1-4, (b) the results of scans 1-20, (c) the results of all the scans, (d) Poisson reconstruction results based on (c) (the top row shows the template mesh), (e) corresponding color-coded error distributions.

First, we obtain low-resolution models from the original partial models by downsampling them to $1/10$ of the full resolution. Each original partial model contains about 3,000-5,000 vertices and therefore each low-resolution model has about 300-500 vertices. Then, we find the corresponding points between neighboring scans using the approach [40]. The method is intrinsic and works well even for partial scans with large deformation. The iterative results of our method are shown in Figs. 6(a-e). Fig. 6(f) demonstrates that the total energy reduces steadily over iterations.

We apply our multi-resolution global registration to the set of models from coarse to fine. In Fig. 7(A), we show four partial scans and the deformed scans after registration using our method and alternative methods. It can be seen that the method [38] has a serious shrinkage problem and a similar phenomenon happens to the method [39] although to a lesser extent. Our method produces the best registration result. Fig. 8 and Fig. 9 illustrate the results when scans are accumulated gradually, using the first 4 scans, the first 20 scans, and all the scans for registration, compared with four state-of-the-art

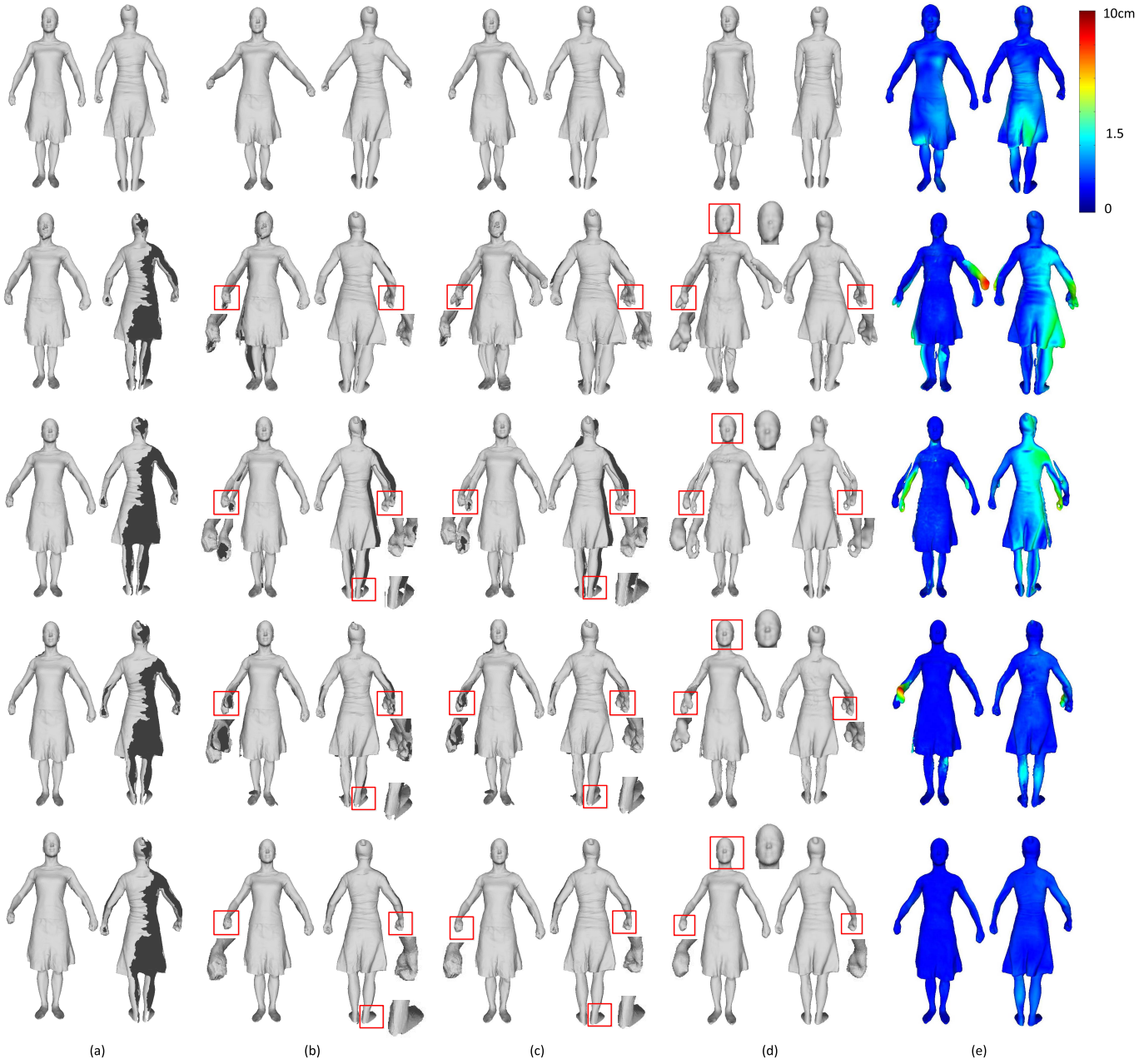


Fig. 9. Comparative results using gradually accumulated scans on the partial dataset *Samba*. Top row: L0 [3], second row: PR-GLS [4], third row: L21 [38], fourth row: L11 [39], and bottom row: our method. (a): the results of scans 1-4, (b) the results of scans 1-20, (c) the results of all the scans, (d) Poisson reconstruction results based on (c) (the top row shows the template mesh), (e) corresponding color-coded error distributions.

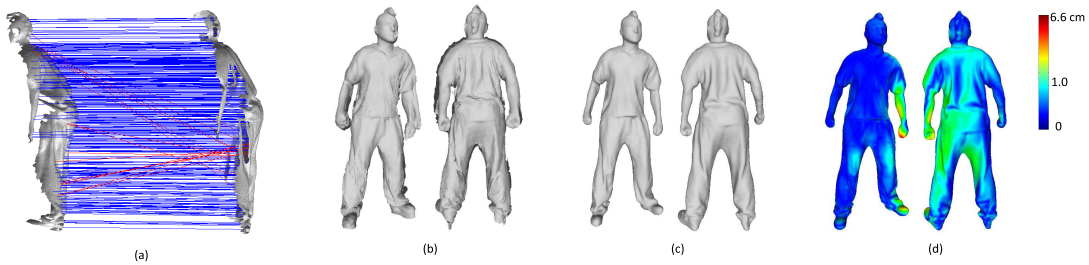


Fig. 10. Comparison results on *Bouncing* dataset with partially incorrect correspondences: (a) part of correspondences for two scans, (b) registration result of all the scans, (c) Poisson reconstruction result, (d) corresponding color-coded error distributions.

methods. We also use standard Poisson reconstruction [45] to obtain watertight meshes. Because the L0 method [3] requires a template mesh for tracking, we choose a complete mesh from the original dataset as the template mesh and register this mesh

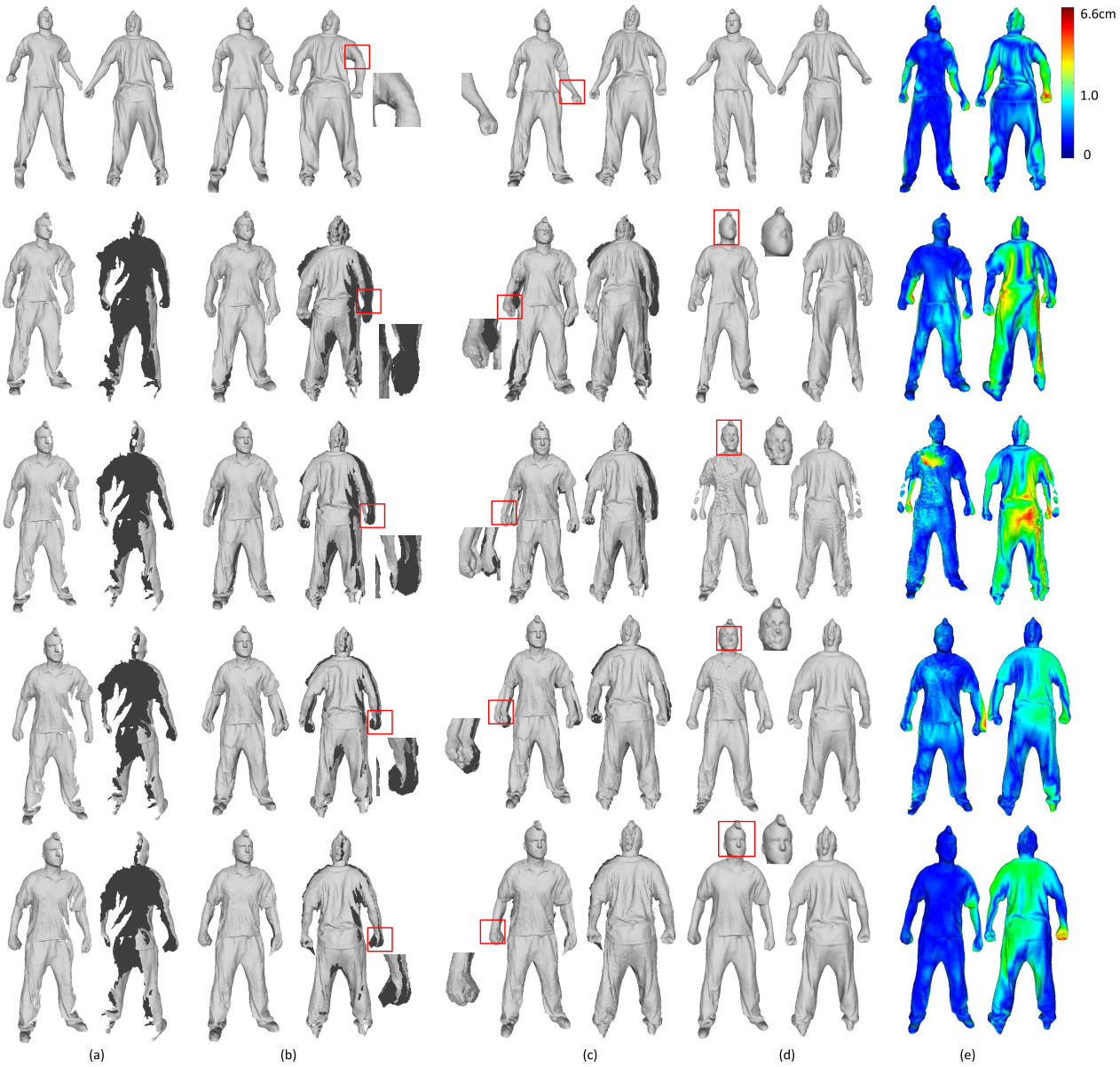


Fig. 11. Comparative results using gradually accumulated scans on the partial dataset *Bouncing* with noise ($\sigma = 0.1\bar{l}$). Top row: L0 [3], second row: PR-GLS [4], third row: L21 [38], fourth row: L11 [39], and bottom row: our method. (a): the results of scans 1-4, (b) the results of scans 1-20, (c) the results of all the scans, (d) Poisson reconstruction results based on (c) (the top row shows the template mesh), (e) corresponding color-coded error distributions.

to the partial meshes. We can see that the L0 method [3] has wrong estimation for the head and dress, and the PR-GLS method [4] has misalignment for the back. The shrinkage problem becomes more and more severe for L21 method [38] in arms and legs. The results of L11 method [39] have clearly visible misalignment in the results of registration and even after Poisson reconstruction, especially in the arms, which are resulted from the accumulation of registration errors. Compared with these methods, the results of our method are smoother and better aligned. By using global registration, our method does not suffer from error accumulation and the use of ARAP constraint avoids shrinking. Our results are better than alternative methods, even with only the first four views. The quantitative evaluation is given in Table II. Our method has the smallest error. The running times of L0 [3], PR-GLS [4], L21 [38], L11 [39], and our method are 20*min*14*s*, 60*min*45*s*, 7*min*56*s*, 19*min*30*s*, 180*min*55*s*, respectively. The experiments were carried out on a desktop PC with an i7 3.4-GHz CPU and 8-GB RAM. Our method is currently implemented using unoptimized MATLAB code.

To evaluate the performance for correspondences with partially incorrect matchings, we obtain two thirds of correspondences using diffusion pruning [40] and the remaining one third using local geometric feature matching based on SHOT signatures [46]. The majority of correspondences from the former are correct while many correspondences from the latter are incorrect due to the ambiguity of local features. One example of correspondences for two partial meshes is shown in Fig. 10(a), and final reconstruction results are shown in Figs. 10(b-d). Wrong correspondences are marked

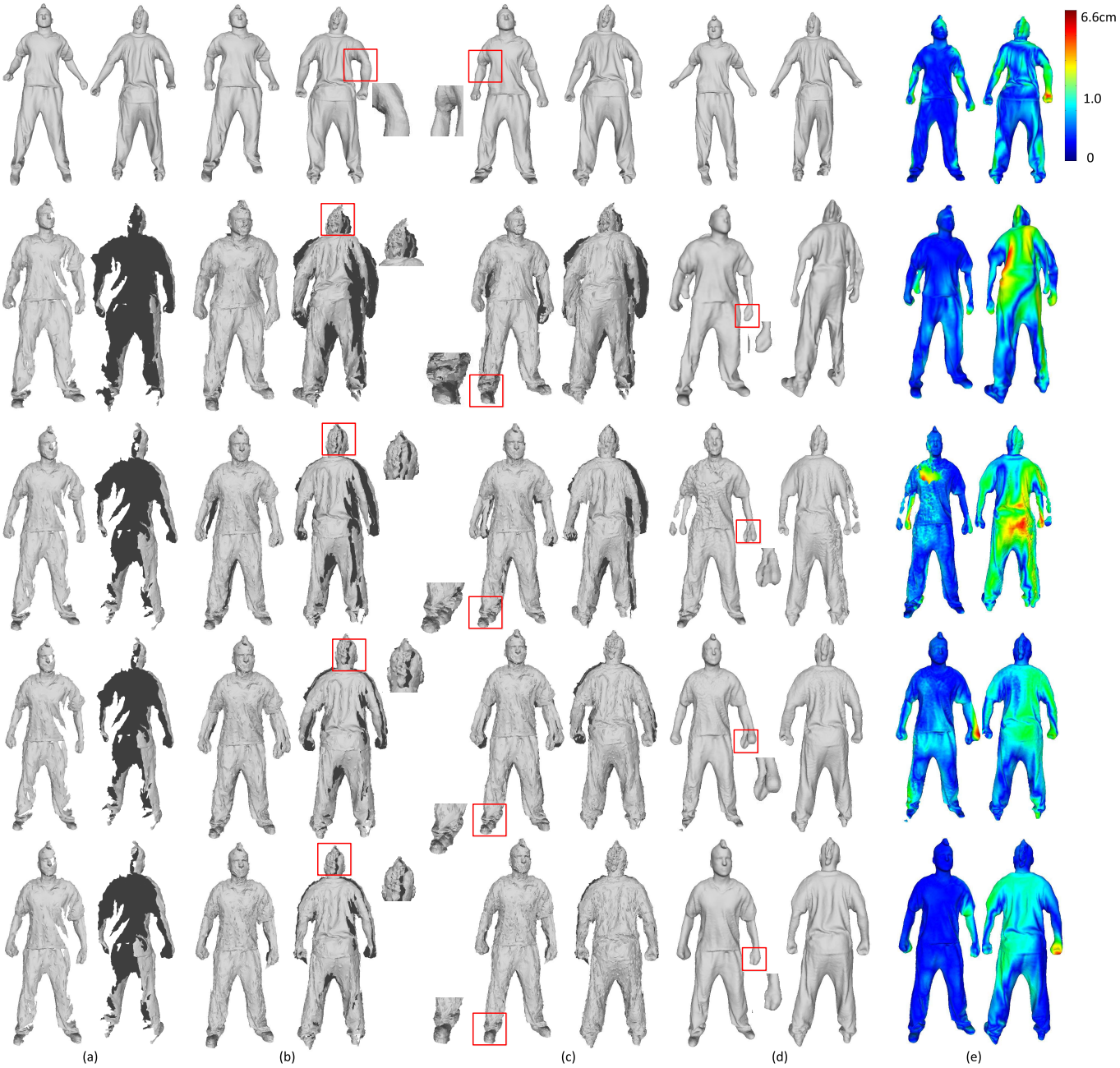


Fig. 12. Comparative results using gradually accumulated scans on the partial dataset *Bouncing* with 10% outliers. Top row: L0 [3], second row: PR-GLS [4], third row: L21 [38], fourth row: L11 [39], and bottom row: our method. (a): the results of scans 1-4, (b) the results of scans 1-20, (c) the results of all the scans, (d) Poisson reconstruction results based on (c) (the top row shows the template mesh), (e) corresponding color-coded error distributions.

as red. It can be seen that our method is robust with respect to incorrect correspondences. Thanks to the regularization terms in our energy function, in particular the transformation smoothness term and the as-rigid-as-possible term, incorrect correspondences which are likely to be substantially different from their neighboring correspondences, are substantially down-weighted, due to large regularization costs if local transformations were to follow them.

To evaluate the robustness of our method, we also experiment on the dataset polluted by dense noise and sparse outliers with the same approach as mentioned in Section V-A. Fig. 11 and Fig. 12 show our results compared with the other methods, and the corresponding distance errors are shown in Table II. The results show that our method is more robust to noise and outliers than the alternative methods.

C. Results on Real Scans

We now test our method on real scans, which are very challenging, because they have much noise and a large number of outliers. Here, we create two datasets scanned using a Kinect: *Waving* and *Flying* datasets. The *Waving* dataset involves large deformations which allows the hands and feet to wave forward and backward. It contains 30 scans with about

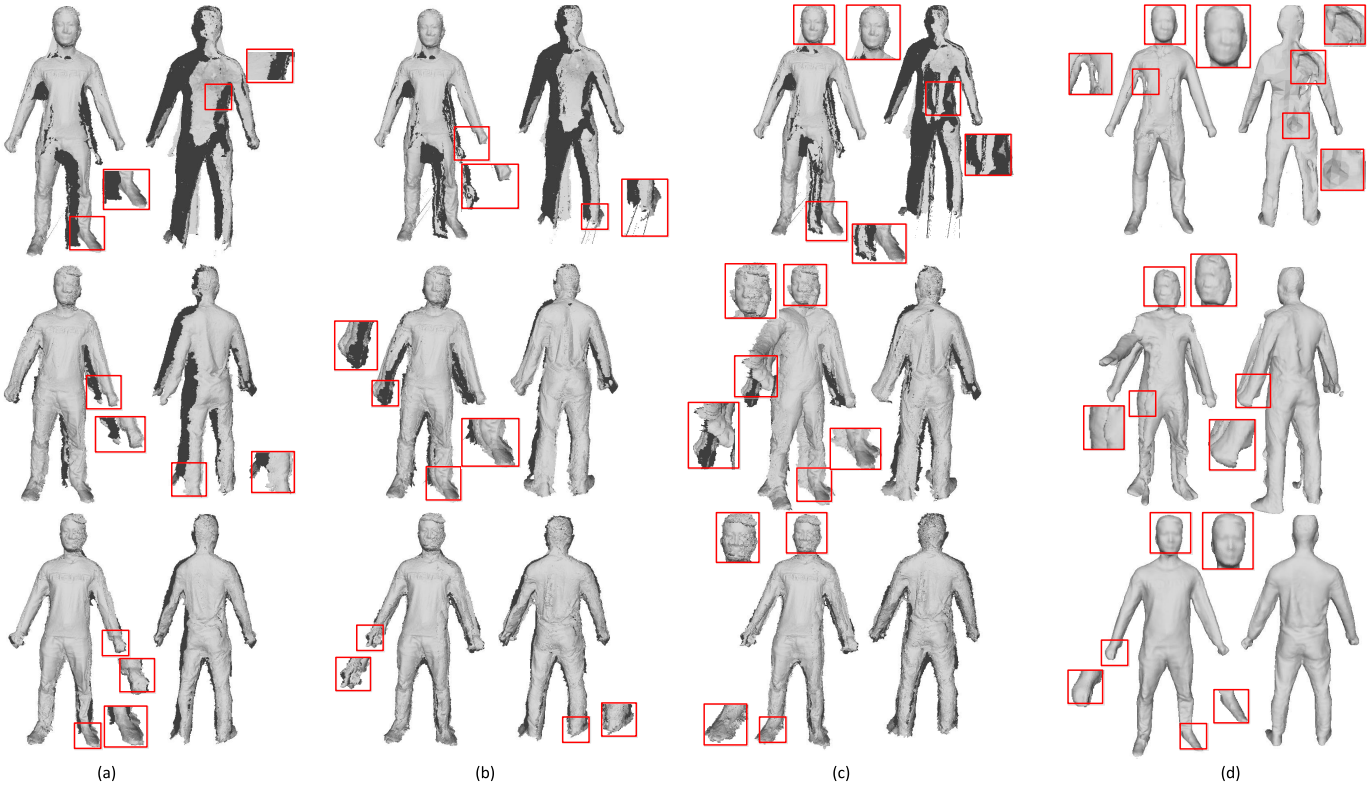


Fig. 13. Comparative results with scans accumulated gradually on the scanned *Waving* dataset. Top row: [38], middle row: [39], bottom row: our method. (a): The results of scans 1-12, (b): the results of scans 1-20, (c) the results of all the scans, (d) the Poisson reconstruction results for (c).

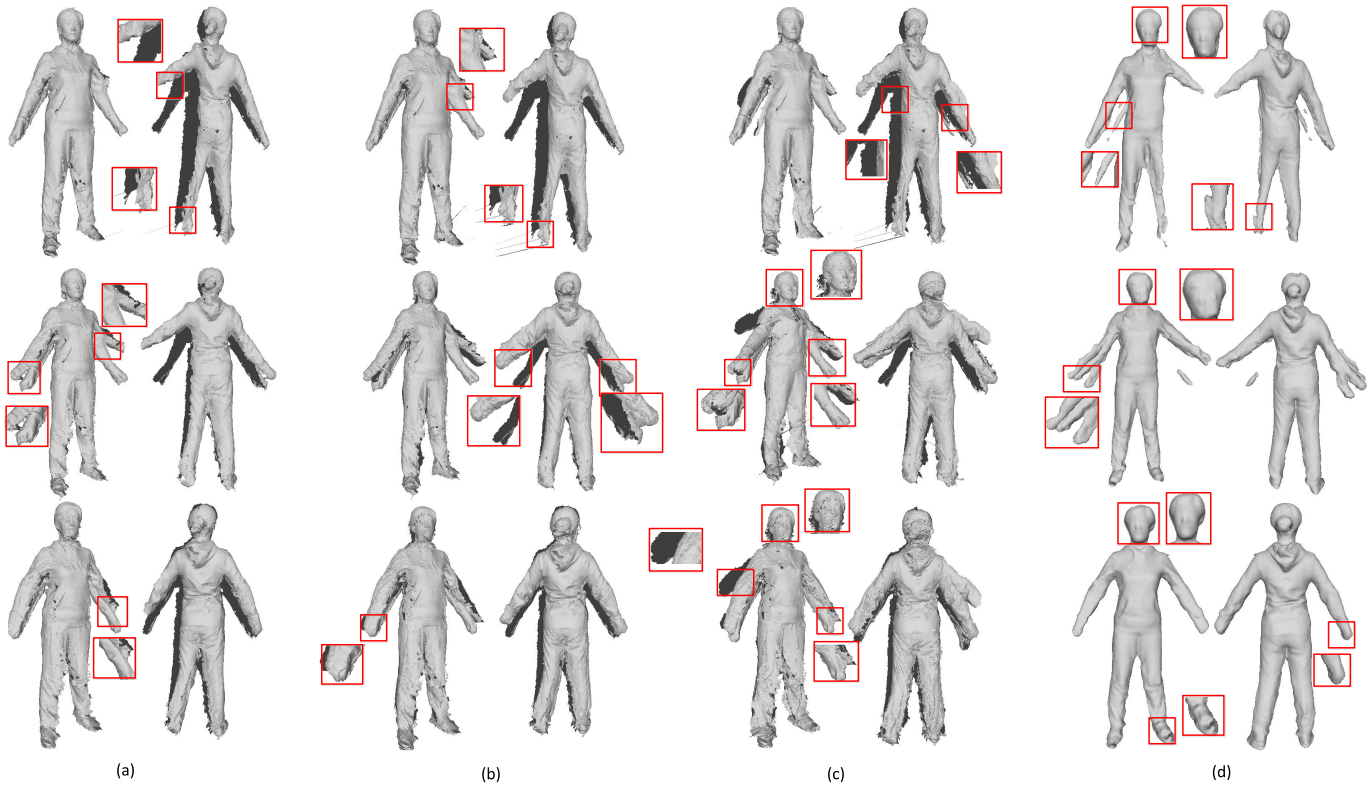


Fig. 14. Comparative results with scans accumulated gradually on the *Flying* dataset. Top row: [38], middle row: [39], bottom row: our method; (a): The results of scans 1-6, (b): the results of scans 1-10, (c) the results of all the scans, (d) the Poisson reconstruction results for (c).

9,000-14,000 vertices for each partial scan. Sample partial models of *Waving* are shown in Fig. 5(b). We obtain the low-resolution models by subsampling with 1/10 of vertices, similar to the clean datasets, and compare our method with

TABLE II
QUANTITATIVE EVALUATION FOR PARTIAL DATASETS (cm)

Method	$Samba_{clean}$	$Bouncing_{clean}$	$Bouncing_{noise}$	$Bouncing_{outlier}$
L0 [3]	0.6318	1.0042	1.0437	1.0145
PR-GLS [4]	1.8008	1.6973	1.7556	2.1227
L21 [41]	1.2959	1.4986	1.5575	1.5417
L11 [42]	0.4910	0.9887	1.0158	1.0225
Ours	**0.2508**	**0.7186**	**0.9849**	**0.9506**

the pairwise registration methods [38], [39] in Fig. 7(B). Similarly, we can see that the method [38] produces highly distorted results and the results of method [39] also contain misalignment. Fig. 13 illustrates the results when scans are accumulated gradually. Since no ground truth data is available, it is not possible to measure the errors quantitatively. However, from visual inspection, it is clear that our global registration method produces superior results. The results of [38] (top row) not only have serious shrinkage but also become more and more flat. With the accumulation of registration errors, the misalignment problem for method [39] also becomes unacceptable, especially in the head and arms. Our method generates significantly better results, including the head and arms.

In order to show the robustness and effectiveness of our method, for the *Flying* dataset, we just use 15 partial scans with dramatic deformations between scans which allow the arms to wave up and down. Sample partial models of *Flying* are shown in Fig. 5(c). As shown in Fig. 7(C), there are serious distortions in the results of method [38], and the transformed scans become more flat. The misalignment of method [39] is also apparent. On the contrary, the results of our method have no such problems. Fig. 14 gives the results when scans are accumulated gradually. The results of method [38] have a serious shrinkage problem and the misalignment problem for method [39] also becomes unacceptable. Our method has better registration results and the reconstructed complete model is accurate and watertight.

VI. CONCLUSION

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 large-scale movement of arms and legs, as well as noise and outliers. In addition, our method allows fewer partial scans to reconstruct a full object.

Our method also has some limitations. First, although our method can handle a wider range of deformations, it becomes more difficult with very few scans, such as the example shown in Fig. 14, since neighboring scans have less overlap. Not all the partial scans are well aligned such as the arms, although the reconstructed complete model removes most artifacts. Second, our current formulation only considers the registration errors of neighboring scans while other scans that have overlap with the current scan will also help for accurate registration. Third, the computation complexity is a little high due to the global formulation. In the future, we will investigate more robust schemes by exploiting potential overlaps between non-adjacent scans and speed up the algorithm using GPU.

REFERENCES

[1] K. Wang, G. Zhang, and S. Xia, "Templateless non-rigid reconstruction and motion tracking with a single RGB-D camera," *IEEE Trans. Image Process.*, vol. 26, no. 12, pp. 5966–5979, Dec. 2017.

[2] W. Huang, X. Cao, K. Lu, Q. Dai, and A. C. Bovik, "Toward naturalistic 2D-to-3D conversion," *IEEE Trans. Image Process.*, vol. 24, no. 2, pp. 724–733, Feb. 2015.

[3] K. Guo, F. Xu, Y. Wang, Y. Liu, and Q. Dai, "Robust non-rigid motion tracking and surface reconstruction using L0 regularization," in *Proc. IEEE Int. Conf. Comput. Vis.*, May 2015, pp. 3083–3091.

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

[5] J. Lei, C. Zhang, Y. Fang, Z. Gu, N. Ling, and C. Hou, "Depth sensation enhancement for multiple virtual view rendering," *IEEE Trans. Multimedia*, vol. 17, no. 4, pp. 457–469, Sep. 2015.

[6] J. Lei, J. Liu, H. Zhang, Z. Gu, N. Ling, and C. Hou, "Motion and structure information based adaptive weighted depth video estimation," *IEEE Trans. Broadcast.*, vol. 61, no. 3, pp. 416–424, Sep. 2015.

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

[8] J. Starck, A. Maki, S. Nobuhara, A. Hilton, and T. Matsuyama, "The multiple-camera 3-D production studio," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 6, pp. 856–869, Jun. 2009.

[9] J. Tong, J. Zhou, L. Liu, Z. Pan, and H. Yan, "Scanning 3D full human bodies using Kinects," *IEEE Trans. Vis. Comput. Graph.*, vol. 18, no. 4, pp. 506–643, Apr. 2012.

[10] G. Ye, Y. Liu, N. Hasler, X. Ji, Q. Dai, and C. Theobalt, "Performance capture of interacting characters with handheld kinects," in *Proc. Eur. Conf. Comput. Vis.*, Oct. 2012, pp. 828–841.

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

[12] M. Zollhözfer *et al.*, "Real-time non-rigid reconstruction using an RGB-D camera," *ACM Trans. Graph.*, vol. 33, no. 4, pp. 1–12, Jul. 2014.

[13] X. Zhou, M. Zhu, S. Leonardos, and K. Daniilidis, "Sparse representation for 3D shape estimation: A convex relaxation approach," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 8, pp. 1648–1661, Aug. 2017.

[14] K. Li, J. Yang, and J. Jiang, "Nonrigid structure from motion via sparse representation," *IEEE Trans. Cybern.*, vol. 45, no. 8, pp. 1401–1413, Aug. 2015.

[15] P. F. U. Gotardo and A. M. Martinez, "Computing smooth time trajectories for camera and deformable shape in structure from motion with occlusion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 10, pp. 2051–2065, Oct. 2011.

[16] M. Dou *et al.*, "Motion2fusion: Real-time volumetric performance capture," *ACM Trans. Graph.*, vol. 36, no. 6, pp. 246:1–246:16, Nov. 2017. doi: 10.1145/3130800.3130801.

[17] R. A. Newcombe, D. Fox, and S. M. Seitz, "DynamicFusion: Reconstruction and tracking of non-rigid scenes in real-time," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 343–352.

[18] D. Guo, K. Li, Y.-K. Lai, and J. Yang, "Global alignment of deformable objects captured by a single RGB-D camera," in *Proc. IEEE Int. Conf. Multimedia Expo*, Jul. 2017, pp. 1554–1559.

[19] D. Vlasic, I. Baran, W. Matusik, and J. Popović, "Articulated mesh animation from multi-view silhouettes," *ACM Trans. Graph.*, vol. 27, no. 3, p. 97, Aug. 2008.

[20] K. Li, Q. Dai, and W. Xu, "Markerless shape and motion capture from multiview video sequences," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 3, pp. 320–334, Mar. 2011.

[21] A. Collet *et al.*, "High-quality streamable free-viewpoint video," *ACM Trans. Graph.*, vol. 34, no. 4, pp. 1–13, Jul. 2015.

[22] S. Lin, Y. Chen, Y.-K. Lai, R. R. Martin, and Z.-Q. Cheng, "Fast capture of textured full-body avatar with RGB-D cameras," *Vis. Comput.*, vol. 32, nos. 6–8, pp. 681–691, Jun. 2016.

[23] M. Dou *et al.*, "Fusion4D: Real-time performance capture of challenging scenes," *ACM Trans. Graph.*, vol. 35, no. 4, pp. 114:1–114:13, Jul. 2016. doi: 10.1145/2897824.2925969.

[24] A. Weiss, D. Hirshberg, and M. J. Black, "Home 3D body scans from noisy image and range data," in *Proc. IEEE Int. Conf. Comput. Vis.*, Nov. 2011, pp. 1951–1958.

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

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

[27] M. Dou, J. Taylor, H. Fuchs, A. Fitzgibbon, and S. Izadi, "3D scanning deformable objects with a single RGBD sensor," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 493–501.

[28] H. Li, B. Adams, L. J. Guibas, and M. Pauly, "Robust single-view geometry and motion reconstruction," *ACM Trans. Graph.*, vol. 28, no. 5, p. 175, Dec. 2009.

[29] M. Liao, Q. Zhang, H. Wang, and R. Yang, "Modeling deformable objects from a single depth camera," in *Proc. IEEE Int. Conf. Comput. Vis.*, Sep. 2009, pp. 167–174.

[30] G. K. Tam *et al.*, "Registration of 3D point clouds and meshes: A survey from rigid to nonrigid," *IEEE Trans. Vis. Comput. Graphics*, vol. 19, no. 7, pp. 1199–1217, Jul. 2013.

[31] H. Lei, G. Jiang, and L. Quan, "Fast descriptors and correspondence propagation for robust global point cloud registration," *IEEE Trans. Image Process.*, vol. 26, no. 8, pp. 3614–3642, Aug. 2017.

[32] S. Bouaziz, A. Tagliasacchi, and M. Pauly, "Sparse iterative closest point," *Comput. Graph. Forum*, vol. 32, no. 5, pp. 113–123, Aug. 2013.

[33] M. Ruhnke, R. Kümmerle, G. Grisetti, and W. Burgard, "Highly accurate 3D surface models by sparse surface adjustment," in *Proc. IEEE Int. Conf. Robot. Autom.*, May 2012, pp. 751–757.

[34] S. Bouaziz and M. Pauly, "Dynamic 2D/3D registration for the Kinect," in *Proc. ACM SIGGRAPH Courses*, May 2013, p. 21:1–21:14.

[35] J. Ma, W. Qiu, J. Zhao, Y. Ma, A. L. Yuille, and Z. Tu, "Robust L2E estimation of transformation for non-rigid registration," *IEEE Trans. Signal Process.*, vol. 63, no. 5, pp. 1115–1129, Mar. 2015.

[36] J. Ma, J. Zhao, J. Tian, A. L. Yuille, and Z. Tu, "Robust point matching via vector field consensus," *IEEE Trans. Image Process.*, vol. 23, no. 4, pp. 1706–1721, Apr. 2014.

[37] J. Ma, H. Zhou, J. Zhao, Y. Gao, J. Jiang, and J. Tian, "Robust feature matching for remote sensing image registration via locally linear transforming," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 12, pp. 6469–6481, Dec. 2015.

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

[39] J. Yang, K. Li, Y.-K. Lai, and D. Guo, "Robust non-rigid registration with reweighted position and transformation sparsity," *IEEE Trans. Vis. Comput. Graphics*, to be published.

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

[41] O. Sorkine and M. Alexa, "As-rigid-as-possible surface modeling," in *Proc. Symp. Geometry Process.*, Jul. 2007, pp. 109–116.

[42] M. Garland and P. S. Heckbert, "Surface simplification using quadric error metrics," in *Proc. ACM SIGGRAPH*, Feb. 1997, pp. 209–216.

[43] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 111–122, Jan. 2011.

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

[45] M. Kazhdan and H. Hoppe, "Screened Poisson surface reconstruction," *ACM Trans. Graph.*, vol. 32, no. 3, p. 29, 2013.

[46] S. Salti, F. Tombari, and L. Di Stefano, "SHOT: Unique signatures of histograms for surface and texture description," *Comput. Vis. Image Understand.*, vol. 125, pp. 251–264, Aug. 2014.

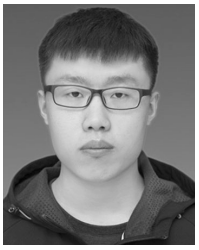
Jingyu Yang (M'10–SM'17) received the B.E. degree from the Beijing University of Posts and Telecommunications, Beijing, China, in 2003 and the Ph.D. degree (Hons.) from Tsinghua University, Beijing, in 2009.

He has been a Faculty Member with Tianjin University, China, since 2009, where he is currently a Professor with the School of Electrical and Information Engineering. He was with Microsoft Research Asia (MSRA) in 2011 and the Signal Processing Laboratory, EPFL, Lausanne, Switzerland, in 2012 and from 2014 to 2015. He has authored or coauthored more than 80 high-quality research papers. His research interests include image/video processing, 3D imaging, and computer vision. His coauthored paper received the Best 10% Paper Award in the IEEE VCIP 2016 and the Platinum Best Paper Award in the IEEE ICME 2017. He served as a Special Session Chair for the VCIP 2016 and an Area Chair for the ICIP 2017.

Daoliang Guo received the B.E. degree from the School of Electronic Information Engineering, Anhui University, in 2015 and the M.E. degree from the School of Electrical and Information Engineering, Tianjin University, in 2018. His research interests include 3D registration, reconstruction, and animation.

Kun Li received the B.E. degree from the Beijing University of Posts and Telecommunications, Beijing, China, in 2006 and the master's and Ph.D. degrees from Tsinghua University, Beijing, in 2011.

She visited the École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, in 2012 and from 2014 to 2015. She is currently an Associate Professor with the School of Computer Science and Technology, Tianjin University, Tianjin, China. Her research interests include dynamic scene 3D reconstruction and image/video processing. She was selected into the Peiyang Scholar Program of Tianjin University in 2016. She was a recipient of the Platinum Best Paper Award in IEEE ICME 2017.



Zhenchao Wu received the B.E. degree from the School of Computer Science, Xi'an Polytechnic University, Xi'an, China, in 2017. He is currently pursuing the M.E. degree with the College of Intelligence and Computing, Tianjin University, Tianjin, China. His interests are mainly in 3D registration and reconstruction.

Yu-Kun Lai received the bachelor's and Ph.D. degrees in computer science from Tsinghua University in 2003 and 2008, respectively. He is currently a Reader of visual computing with the School of Computer Science & Informatics, Cardiff University. His research interests include computer graphics, geometry processing, image processing, and computer vision. He is on the Editorial Board of *The Visual Computer*.