Global Correspondence Optimization for Non-Rigid Registration of Depth Scans

Hao Li   Robert W. Sumner   Mark Pauly

Applied Geometry Group
ETH Zurich

Abstract
We present a registration algorithm for pairs of deforming and partial range scans that addresses the challenges of non-rigid registration within a single non-linear optimization. Our algorithm simultaneously solves for correspondences between points on source and target scans, confidence weights that measure the reliability of each correspondence and identify non-overlapping areas, and a warping field that brings the source scan into alignment with the target geometry. The optimization maximizes the region of overlap and the spatial coherence of the deformation while minimizing registration error. All optimization parameters are chosen automatically; hand-tuning is not necessary. Our method is not restricted to part-in-whole matching, but addresses the general problem of partial matching, and requires no explicit prior correspondences or feature points. We evaluate the performance and robustness of our method using scan data acquired by a structured light scanner and compare our method with existing non-rigid registration algorithms.

Categories and Subject Descriptors (according to ACM CCS): I.3.5 [Computer Graphics]: Computational Geometry and Object Modeling

1. Introduction

Surface registration is a fundamental problem in geometric modeling and 3-D shape acquisition. Most scanning systems provide partial surface data that must be aligned and merged to obtain a complete digital representation of the acquired object. Computing such an alignment is a challenging task, since only subsets of the input scans overlap, and these regions of correspondence are not known a priori. Correspondence errors lead to significant degradation in the final quality of the scanned object. Consequentially, the registration problem is one of the most difficult and crucial steps in the shape acquisition pipeline.

A substantial amount of research has been devoted to the registration of rigid objects, where two scans can be aligned with a global Euclidean transformation [RL01]. Lifting this rigidity assumption makes the registration problem significantly more challenging. Objects that undergo deformations while being scanned cannot be registered by a single global rigid motion. Registration of non-rigid shapes thus requires estimating both correspondence and a suitable warping function that matches the deformation of the acquired object. False correspondences are debilitating, since they can lead to strong distortions of the shape that are not consistent with the object's deformation.

We present a registration algorithm for partial scans of deforming shapes that addresses the challenges of non-rigid registration within a single non-linear optimization (Figure 1). Our algorithm simultaneously solves for correspondences between points on source and target scans, confi-
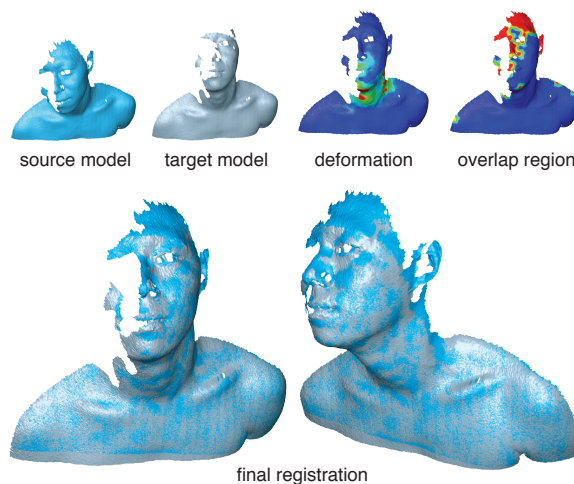
Figure 1: *Non-rigid registration. A global, non-linear optimization simultaneously solves for the parameters of the deformation, a dense set of point correspondences, and the region of overlap between the two surfaces.*

dence weights that measure the reliability of each correspondence and identifies non-overlapping areas, and a warping field that brings the source scan into alignment with the target geometry. The optimization maximizes the region of overlap and the spatial coherence of the deformation while minimizing registration error. Poor local minima are avoided

with an iterative execution schedule that detects sub-optimal convergence and restarts the optimization with improved initial conditions so that a better result is obtained. Our system employs a reduced model of deformation that separates the geometric complexity of the scans from the complexity of the optimization, thereby enhancing performance. A non-linear deformation energy avoids unnatural shearing artifacts by keeping the deformation locally as-rigid-as-possible.

Our registration system enhances the shape acquisition process by contributing a new tool to align partial scans of deforming shapes. We have developed our system in the context of range scans from a single depth scanner, as we feel this is an appropriate acquisition setting for deforming objects. However, we make few additional assumptions. Our system is robust to considerable deformations and does not require high-speed acquisition. Our method is not restricted to part-in-whole matching, but addresses the general problem of partial matching where the overlap region is a subset of both shapes. Finally, our algorithm requires no explicit prior correspondences or feature points, which makes it more robust to settings where markers are hard to place and track on the scanned subject or when feature extraction methods yield unreliable key points.

2. Related Work

Shape registration is pervasive in computer graphics, impacting 3-D scanning, motion capture, shape morphing, texture transfer, and statistical shape analysis. As such, a tremendous amount of attention has been given to this problem. In order to make this discussion more tractable and germane, we focus only on previous work that addresses the problem of non-rigid registration for 3-D scans.

Non-Rigid Alignment of Rigid Objects. Device non-linearities and calibration errors can cause low frequency warps on the scans of rigid objects. To correct such distortions, Ikemoto and coworkers [IGL03] introduce a non-rigid registration technique that decomposes the input scans using a coarse-to-fine hierarchy of locally rigid pieces that are allowed to translate and rotate with respect to one another. The advantage of this method is that no specific characterization of the warp is required since a continuous deformation is approximated from the convergence of the piecewise rigid model. However, the running time of this technique is quadratic in the number of patches. Brown and Rusinkiewicz [BR04] address this scalability issue using thin-plate splines to represent smooth warps and a hierarchical ICP method to find good feature correspondences between subdivided patches. An extension of this method for the simultaneous alignment of a large number of scans with locally weighted ICP matching has been recently presented by the same authors [BR07].

Template-Based Registration for Large Deformations. The registration of scans with large-scale deformations, such as those of an articulated body, requires a more general and flexible deformation model. Moreover, local shape matching techniques might fail if the shape is distorted beyond a certain limit. Previous solutions to this problem commonly involve the use of template models that are warped toward the input scans [BV99, ACP03, PMG*05, ASK*05, ARV07]. The template model provides a strong geometric prior and thus leads to high-quality reconstructions with automated hole-filling and noise removal. Correspondence estimation is often facilitated by the use of tracked marker points or hand-selected feature correspondences. Park and Hodgins [PH06] propose using a large set of markers to accurately capture the dynamic motion of human bodies. An algorithm that does not require hand selected markers has been developed by Anguelov and colleagues [ASP*04] where a joint probabilistic model over all point-to-point correspondences is optimized between two shapes. While the method is fully automatic and is able to recover significant movements of articulated parts and non-rigid deformations, it requires that one of the input shapes is a subset of the other. A different approach that also requires a template model has been proposed by Bronstein and colleagues [BBK06]. They address the partial matching problem with a multi-dimensional scaling algorithm that aims at minimizing the distortion of the mapping between two surfaces.

Registration of Real-time Range Scans. Recent advances in real-time structured light scanners [WLG07] enable the continuous capture of deforming objects and produce dense sequences of range scans. In the most general setting, no prior template shape is given and no markers or explicit feature point correspondences are available. In this case, a true partial matching problem must be solved, as opposed to a part-in-whole matching as for template-based techniques. Existing work on the pairwise alignment of dense scans typically assumes that pairwise scans undergo small deformations while having a significant amount of overlap. Two recent papers present innovative approaches for this scenario.

Mitra and colleagues [MFO*07] present a registration method for dense time-series of point clouds that does not explicitly compute correspondence. Instead, they aggregate all scans into a 4-D space-time surface and estimate inter-frame motion from kinematic properties of this surface. This technique requires the deformation of adjacent frames to be sufficiently small as it is designed primarily for articulated motions. Wand and coworkers [WJH*07] introduce a statistical framework that computes a globally optimal shape and deformation of the complete model over every frame. The method relies on an initial pairwise registration of all adjacent scans using a non-rigid ICP variant based on a deformation model proposed in [HTB03]. Pairwise correspondences are then iteratively improved during the optimization assuming the input scans deform smoothly over time.

Non-Rigid ICP. We speak of non-rigid ICP algorithms (N-ICP) whenever correspondences are solely based on spatial proximity heuristics and the transformation is represented by a general deformation. For point cloud representations, deformation can be achieved using a skeletal link structure which connects neighboring points as demonstrated in [HTB03]. For manifold surfaces, scan alignment algorithms usually use a deformation model based on smooth local affine transforms. Because the deformation model allows
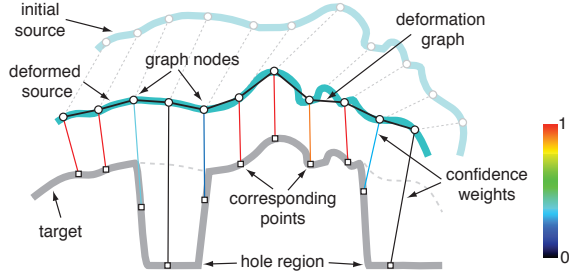
Figure 2: *Our non-rigid registration algorithm simultaneously solves for the affine transformations of each node of the deformation graph, the corresponding points on the target shape, and the confidence weights. The latter evolve during the optimization to identify the region of overlap.*

Figure 3: *Data representations. From left to right: input depth map, reconstructed triangle mesh, continuous weighted least squares (WLS) approximation.*

many degrees of freedom, it is usually used in conjunction with sparse hand selected feature correspondences and a procedure that iteratively reduces the stiffness [ACP03, SP04]. While a template model is still required, the optimal step non-rigid ICP (N-ICP 1) proposed by Amberg and colleagues [ARV07] demonstrates several successfully aligned examples without the use of hand selected correspondences. Techniques for improving the convergence of rigid ICP algorithms [PHYH06] extend to the non-rigid setting proposed by Pauly and coworkers [PMG*05]. This non-rigid ICP variant (N-ICP 2) is mainly different in that a combination of point-to-plane [CM92] and point-to-point metrics is used in order to avoid penalizing correspondences due to large featureless regions.

Global Correspondence Optimization. In a typical capture session, neither the amount of deformation nor the percentage of overlap between adjacent frames of range scans can be predicted. Misalignments between adjacent pairs of scans can hinder the performance of subsequent algorithms that focus on global reconstruction. In contrast to rigid ICP, we show in this paper that using proximity heuristics to determine correspondences is less reliable when large deformations are present. Instead of heuristically determining approximated corresponding points, our algorithm simultaneously solves for correspondences, confidence weights, and deformation field within a single optimization.

3. Registration Overview

The core task of our system is the registration of two partial scans of a deforming object: a source scan is registered to a target scan captured at a different point in time (c.f. Figure 2). Each scan represents only a portion of the entire object. Since new parts of the object may have come into view and other parts may have become occluded in between the two captures, the region of overlap is a subset of both scans. Furthermore, the object may have undergone both rigid and non-rigid deformation, such as global Euclidean motion, pose changes, or changes in facial expression.

We employ a non-linear deformation system that favors
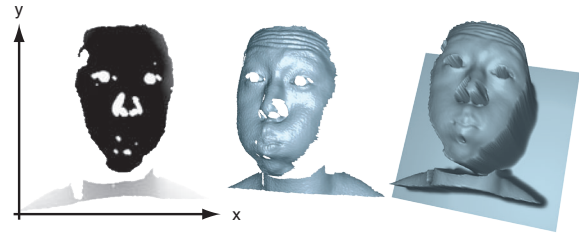
natural deformations by maximizing both rigidity and consistency. The deformation is controlled by correspondences expressed at points distributed evenly over the source so that each point has a corresponding position on the target shape. The optimization solves simultaneously for both the deformation parameters as well as the correspondence positions. Since the deformation algorithm is designed to favor the most natural deformations, the optimizer will update the target correspondence positions so as to achieve a natural deformation. Inconsistent correspondences are penalized and the features of the source and target are naturally aligned with one another, since such an alignment leads to a lower deformation energy state. Since some source points have no corresponding position on the target due to partial overlap, we augment each correspondence with a weight that is also solved for by the optimizer. We design an energy functional so that this weight is naturally brought to zero when an appropriate correspondence cannot be found. In doing so, the zero-weighted correspondences indicate non-overlapping regions and do not influence the deformation.

4. Input Data

We develop our work in the context of depth maps, in which a 2-D image in the *xy*-plane stores a depth value along the *z*-direction. Our data is acquired using a range scanner based on structured light [WLG07] that operates at approximately 15 frames per second with a capture volume of $40 \times 30 \times 60 \, \text{cm}^3$. Please note that, although the scanner has a high frame rate, our registration algorithm is robust under much lower frame rates. We demonstrate examples where the registered scans are spaced as many as 20 frames apart, or 1.33 seconds apart in time.

A different 3-D triangle mesh is extracted for both the source and target scans from their respective depth maps by triangulating the pixel grid and assigning the *z*-direction of each vertex to be the corresponding depth value. As a post-process, triangles with an edge length greater than a threshold value of 0.5 cm are deleted, since these occur at occlusion boundaries where there is a large depth disparity. Additionally, we delete all connected components with fewer than 200 triangles (see Figure 3).

5. Deformation Model

We require a deformation model that favors high-quality, natural shape deformations. In order to be general, we avoid domain-specific solutions, such as a kinematic skeleton that would have to be customized a priori to each acquired shape. The deformation system should also be robust to noisy scans composed of multiple connected components. Finally, efficiency is imperative, since the deformation parameters will be coupled with many other unknowns into a large optimization problem.

Given these constraints, we build upon the embedded deformation approach of Sumner and colleagues [SSP07]. The source scan is augmented with a reduced deformable model in the form of a deformation graph. Graph nodes, defined by their positions $\mathbf{x}_1 \ldots \mathbf{x}_n$, are chosen by uniformly sampling the range image. One affine transformation is associated with each node and induces a deformation on the nearby space. The influence of nearby nodes is blended by the embedded deformation algorithm in order to deform the scan vertices or the graph nodes themselves. Undirected edges connect nodes of overlapping influence to indicate local dependencies. The affine transformation for node i is specified by a 3×3 matrix $\mathbf{A}_i$ and a 3×1 translation vector $\mathbf{b}_i$. In this way, the collection of all per-node affine transformations expresses a non-rigid deformation of the scan and the graph.

Although the embedded deformation method can, in principle, represent both rigid and non-rigid deformations, the performance of our registration system is enhanced by modeling these two quantities separately. Thus, we augment the embedded deformation framework with a global rigid transformation defined by a rotation matrix $\mathbf{R}$ (parameterized in axis-angle form) and a translation vector $\mathbf{t}$. The rotation is relative to the center-of-mass $\mathbf{g}$ of the scan. The source graph nodes and mesh vertices are deformed by first applying the local non-rigid embedded deformation routine and then the rigid transformation so that a vertex $\mathbf{v}_j$ is transformed to $\tilde{\mathbf{v}}_j$ according to:

$$\tilde{\mathbf{v}}_j = \Phi_{\text{global}} \circ \Phi_{\text{local}} \left(\mathbf{v}_j \right), \qquad (1)$$

where

$$\Phi_{\text{global}} \left(\mathbf{v}_j \right) = \mathbf{R}(\mathbf{v}_j - \mathbf{g}) + \mathbf{g} + \mathbf{t} \qquad (2)$$

and

$$\Phi_{\text{local}} \left(\mathbf{v}_j \right) = \sum_{i=1}^{n} w_i(\mathbf{v}_j) \left[\mathbf{A}_i(\mathbf{v}_j - \mathbf{x}_i) + \mathbf{x}_i + \mathbf{b}_i \right]. \qquad (3)$$

The weights $w_i(\mathbf{v}_j), i \in 1 \ldots n$, are nonzero for the k-nearest nodes ($k = 4$ in our examples) and defined by

$$w_i(\mathbf{v}_j) = \frac{1 - \|\mathbf{v}_j - \mathbf{x}_i\|/d_{\max}}{\sum_{p=1}^{k} 1 - \|\mathbf{v}_j - \mathbf{x}_p\|/d_{\max}}, \qquad (4)$$

where $d_{\max}$ is the distance to the $k+1$-nearest node.

This deformation model forms the first building block in our overall optimization strategy. Both the global rigid transformation and the per-node affine transformations are treated as unknowns in the optimization. Two energy functionals

control the deformation. The E_{rigid} term penalizes the deviation of each transformation from a pure rigid motion. Consequently, local features deform as rigidly as possible avoiding shearing or stretching artifacts. E_{rigid} is defined by

$$E_{\text{rigid}} = \sum_i \text{Rot}(\mathbf{A}_i), \qquad (5)$$

where

$$\begin{aligned}
\text{Rot}(A) &= (\mathbf{a}_1^{\text{t}} \mathbf{a}_2)^2 + (\mathbf{a}_1^{\text{t}} \mathbf{a}_3)^2 + (\mathbf{a}_2^{\text{t}} \mathbf{a}_3)^2 + \\
&\quad (1 - \mathbf{a}_1^{\text{t}} \mathbf{a}_1)^2 + (1 - \mathbf{a}_2^{\text{t}} \mathbf{a}_2)^2 + (1 - \mathbf{a}_3^{\text{t}} \mathbf{a}_3)^2
\end{aligned}$$

and $\mathbf{a}_1$, $\mathbf{a}_2$, and $\mathbf{a}_3$ are the column vectors of $\mathbf{A}_i$. A second energy term, E_{smooth}, serves as a regularizer for the deformation by indicating that the affine transformations of adjacent graph nodes should agree with one another:

$$E_{\text{smooth}} = \sum_i \sum_{j \in \mathcal{N}(i)} \left\| \mathbf{A}_i(\mathbf{x}_j - \mathbf{x}_i) + \mathbf{x}_i + \mathbf{b}_i - (\mathbf{x}_j + \mathbf{b}_j) \right\|_2^2,$$

where $\mathcal{N}(i)$ consists of all nodes that share an edge with node i. Finally, a fitting term, defined in Section 6, provides a desired position for each graph node and is the impetus that induces deformation.

6. Correspondence

For each graph node, we associate one correspondence value that indicates the corresponding position on the target shape. This position is initialized via a closest point computation and subsequently updated by the optimizer. Since our range scans are created from captured depth maps, the depth map itself provides a natural parameterization for each scan. Thus, we represent the correspondence position for node i by its (u_i, v_i) values in the parameter domain of the target depth map. The function $\mathbf{c}(\mathbf{u}_i)$ maps from the parameter domain back to the 3-D position:

$$\mathbf{c}(\mathbf{u}_i) = \begin{bmatrix} \mathbf{u}_i \\ c(\mathbf{u}_i) \end{bmatrix}, \qquad (6)$$

where $\mathbf{u}_i = [u_i, v_i]^{\text{t}}$ and $c(\mathbf{u}_i)$ is a scalar function that gives the z-value of the mapped point.

In practice, we obtain better performance by allowing (u_i, v_i) to be transformed by the deformation model's rigid transformation and define the transformed coordinates $\tilde{\mathbf{u}}_i$ according to:

$$\tilde{\mathbf{u}}_i = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} (\mathbf{R}(\mathbf{c}(\mathbf{u}_i) - \mathbf{g}) + \mathbf{g} + \mathbf{t}) + \mathbf{d}_i. \qquad (7)$$

We introduce the energy term E_{fit}, which strives to move each source graph node to its corresponding position on the target shape:

$$E_{\text{fit}} = \sum_{i=1}^{n} \|\tilde{\mathbf{x}}_i - \mathbf{c}(\tilde{\mathbf{u}}_i)\|_2^2, \qquad (8)$$

where $\tilde{\mathbf{x}}_i$ is the deformed position of node i. The (u_i, v_i) parameters for each graph node become unknowns of the optimization, which allows the corresponding points to move along the surface of the target scan. This parameterization

of the target scan is a key ingredient of our method, since it automatically constrains the corresponding points to lie on the target scan and avoids the need for re-projection during the optimization as in the case of non-rigid ICP approaches.

Subsequent numeric computations (Section 8) require computing partial derivatives with respect to the target scan's parameter domain. For efficiency and numeric robustness, we precompute the required derivitives by first building a continuous approximation of the target shape. Since our shape is defined on a function graph rather than a manifold, we favor a weighted least squares (WLS) approximation using a 2-D quadratic polynomial basis $[1, u, v, uv, u^2, v^2]$ and a Wendland function of degree 5 as a weighting function (see [Wen05] for details). Partial derivatives are precomputed for each pixel in the depth map and bilinearly interpolated at runtime.

7. Partial Overlap

One principle challenge of our registration framework is the ambiguity introduced by scans that only overlap one another partially. For some portions of the source mesh, no corresponding point exists on the target and this region of overlap is not known a priori. Instead, our system computes it automatically. We accomplish this task by making a modification both to the data representation and the correspondence energy functional.

Each range image contains portions where object measurements were obtained and "empty" regions where no object was detected. We preprocess the target range image by filling each empty pixel with a large value l so that the empty regions are replaced by deep holes after the mesh is reconstructed (see Figure 3). We set l to be twice the maximum depth value measured by the scanner so that hole regions lead to a large penalty in the fitting energy E_{fit}.

One consequence of this change is that the hole regions will be treated as outliers in the WLS reconstruction due to the disparity between the object depths and the hole depth, leading to artifacts in the reconstruction. Thus, we perform a feathering operation in which a morphological erosion detects the border of the object region, and a smoothing filter is applied to the hole and border regions to ensure a smooth blend between the two [HSZ87]. The closest point variables (Section 6) do not distinguish between object geometry and hole geometry and are free to move between the two via the (u, v) mapping. Second, we associate a weight parameter ω_i with each correspondence and modify E_{fit} as follows

$$E_{\text{fit}}^* = \sum_{i=1}^n \omega_i^2 \|\tilde{\mathbf{x}}_i - \mathbf{c}(\tilde{\mathbf{u}}_i)\|_2^2. \tag{9}$$

Finally, we introduce the confidence energy term

$$E_{\text{conf}} = \sum_{i=1}^n (1 - \omega_i^2)^2. \tag{10}$$

Values of ω_i close to one indicate a reliable correspondence, while values close to zero indicate that no appropriate correspondence is found. The E_{conf} energy aims at maximizing the number of reliable correspondences by pushing the

weights towards one, and thus maximizes the region of overlap. Now we consider what the optimization may do when a portion of the source mesh has no match on the target. First of all, we force each graph node to have some correspondence on the target regardless of whether there actually is an overlap between that portion of the source and the target. Without the modifications presented in this section, partial overlap would result in many bad correspondences and cause significant artifacts in the computed deformation, since the bad correspondences would pull the target shape in incompatible directions. Such unnatural deformations are high energy states, since the deformation model favors smooth deformations that maximize rigidity.

However, using our ω formulation, the source regions that are not present in the target can freely match to the hole regions. There is a high cost in terms of E_{fit}^* to such matches as the hole region is far away. Again, deforming the source to the position of correspondence in the hole also yields a high energy, since such large deformations are penalized. Thus, the minimum energy configuration naturally occurs when the ω_i parameter is reduced to zero by the optimizer. While this incurs some cost from E_{conf}, the cost is less than the alternatives. As a consequence, the optimizer naturally detects non-overlapping regions via the ω_i parameters (Figure 2).

8. Optimization

We sum the individual energy terms from the previous sections to form the full objective function of our optimization:

$$E = \alpha_{\text{rigid}}E_{\text{rigid}} + \alpha_{\text{smooth}}E_{\text{smooth}} + \alpha_{\text{fit}}E_{\text{fit}}^* + \alpha_{\text{conf}}E_{\text{conf}}. \tag{11}$$

The unknowns comprise the global rigid transformation, the affine transformations of the deformation graph, the (u, v) parameter domain coordinates for each graph node correspondence, and the confidence weights ω_i for each node. The number of optimization variables is thus $15n + 6$ with n the number of deformation graph nodes.

We solve this nonlinear least-squares problem using the Levenberg-Marquardt algorithm [MNT04]. Since the system matrix is sparse, we solve the normal equations in each iteration using a direct solver that employs sparse Cholesky factorization [SG04]. A simple heuristic is employed to automatically adapt the optimization weights. Initially, $\alpha_{\text{rigid}} = 1000$, $\alpha_{\text{smooth}} = 100$, and $\alpha_{\text{conf}} = 100$. Each value is halved whenever $|F_k - F_{k-1}| < 10^{-5}(1 + F_k)$, with $F_k = E(\theta_k)$, until $\alpha_{\text{rigid}} < 1$, $\alpha_{\text{smooth}} < 0.1$, and $\alpha_{\text{conf}} < 1$. The weight α_{fit} is held constant at 0.1 during the optimization. The adaptation of weights initially favors global rigid alignment and subsequently lowers the stiffness of the object to allow increasing deformation as the optimization progresses. This automatic procedure is used for all shown examples; hand-tuning is not necessary.

Iterative Improvement. We detect convergence when $|F_k - F_{k-1}| < 10^{-6}(1 + F_k)$. As with any non-linear optimization, our system converges to a local optimum that may not represent the best possible global solution. We employ an iterative improvement algorithm to find a better local minimum

by teleporting the solution to a different position in the energy landscape and restarting the optimizer from this new position. When the system converges, all correspondences are recalculated via a closest-point computation. Next, poor correspondences are detected using three criteria. A correspondence is poor if it is in a hole region, if the distance from the source graph node to the corresponding point on the target is greater than 2 cm, or if normals are inconsistent. A surface normal is maintained for each graph node and transformed along with the graph deformation. If the dot product of this transformed normal with the surface normal at the corresponding position on the target shape is less than 0.6, then the normals are considered inconsistent. The ω value for each poor correspondence is set to zero, and the ω value for all others is set to one. These heuristics are executed every time the optimization converges, and the optimizer is restarted with the new correspondence and ω values. The entire process (including the iterative improvement) converges when $|F_k - F_{k-1}| < 10^{-8}(1 + F_k)$.

Final Warp. The continuous WLS approximation of the target scan is essential for the optimization, since it allows a unified treatment of valid depth samples and holes, and gives access to derivative information of the corresponding points $\mathbf{c}(u_i, v_i)$ with respect to the unknowns (u_i, v_i). As a result, the above registration procedure computes a warp between the source scan and the WLS approximation of the target. Since the WLS approximation smooths some features, our algorithm performs one final step to find a deformed source scan that matches the target scan more accurately. The correspondence positions are projected from the WLS approximation onto the target scan. Any projected correspondence that is farther than 2 cm from its source graph node is discarded. The remaining correspondences are used to solve one final deformation problem in which they are fixed as constants and are not controlled by the optimization. The per-node ω values are also removed from the optimization, so that only the deformation model parameters are solved for. This essentially warps the source scan directly to the target scan using the valid correspondences found during the optimization.

9. Results

To illustrate the performance of our proposed technique, we conduct a series of experiments on synthetic data and real scans. The results are then compared with those of two recent non-rigid ICP variants. Our proposed global correspondence optimization algorithm and N-ICP 1 implementation were both performed on a 3.0 GHz Quad-Core Intel Xeon machine with 8 GB RAM. The longest computation was for the torso example from Figure 6 and required 219 iterations until convergence which took 2 min 19 s in total.

Convergence and Robustness. Our first test case (Figure 4) evaluates the performance of our method on synthetic data with given ground truth correspondence. The source depth map is created by sampling a digital model of an elephant. We simulate a pose change by applying a warp to this model in order to obtain the deformed target shape. In addition, parts of both surfaces have been removed so that only sub-

sets of both models are in correspondence. The optimization correctly classifies this region of overlap by solving for appropriate correspondences and confidence weights as indicated by the color-coded energy terms. Large fitting errors E_{fit} are balanced by low confidence weights and hence a high value in E_{conf}. Since the objective function includes the augmented fitting term E_{fit}^*, where the squared distances to the corresponding points are scaled by the confidence weights, an optimal trade-off between alignment and deformation is found. The graph illustrates how the distance of the corresponding points with respect to the ground truth data evolves as the optimization progresses.

In Figures 5 to 7 we apply our method to real range scans. To simulate the effects of fast motion, we skipped several frames for the target shape. Figure 5 shows a non-rigid registration of a human face. Our algorithm accurately captures the deformation on the cheeks, a mostly feature-less region, and the global rotation of the head. At the same time the relevant features of the face, such as nose, mouth, eyes, and ears are correctly aligned. Figures 6 and 7 show registration of scans of articulated objects, where most of the deformation is concentrated on a small region of the model. These examples are challenging for a marker-less algorithm, since the surface parts that contain most of the important features, the face or the fist, are substantially different in both scans and overlap only partially. Most of the correspondences are located on the torso and the arm, so that these regions dominate the energy terms of the objective function. Still, our algorithm is capable of aligning the facial features and accurately determines the regions of overlap. In this respect, the coupled optimization of correspondence points and deformation can be seen as a form of point-to-plane metric used for rigid ICP (c.f. [CM92]) where sliding along the surface is allowed and featureless regions would not penalize the optimization. The main difference is that in our optimization, the correspondence points with high confidence weights remain on the target surface, reducing the effect of approximation errors using a point-to-plane metric. In Figure 7, the algorithm correctly captures the bending of the arm but produces a slight misalignment in the fist. As the visualizations of the E_{conf} energy shows, few reliable correspondences have been found in this region. This is mostly due to the inferior quality of the input data that leads to a poor WLS approximation of the depth map.

Comparison to Related Work. The performance of different registration methods is shown in Figure 9. For each algorithm, we examine four test cases on the torso in order to visualize how different the poses of the input data can be, so that the registration is still able to converge successfully. We increase the number of frames between the source and target input scans for each test as shown in the first four columns. The last column shows the first occurrences of notable misalignments for each approach. The initial alignment of our input data as shown in the first row is followed by a sequence of rigid alignments performed using rigid ICP with geometric stable sampling [GIRL03], normal compatibility pruning, and the point-to-plane metric. Subtle misalignments are already visible between frame 0 and 2.

N-ICP 1: The third row of Figure 9 shows the best results we could achieve using our implementation of the optimal step non-rigid ICP method (N-ICP 1) from [ARV07]. Instead of using the deformation model from [ACP03] as in the original work, we employed the the model from [SP04] which is known to produce comparable results The input meshes were decimated to 5% of the original size using the algorithm from [GH97] as the registration algorithm relies on dense correspondences which cannot be handled efficiently by the deformation model for our high resolution input scans.

Although a correct alignment could be found for the easiest test case, the registration fails for a registration between frame 0 and 6. The primary issue here is that the employed point-to-point metric in N-ICP 1 penalizes sliding of correspondences during the deformation process. A noticeable shrinking can also be observed which is due to the employed deformation model which only enforces smoothness over the local affine transforms.

N-ICP 2: The non-rigid ICP method described in [PHYH06, PMG*05] uses a combination of point-to-plane and point-to-point metric is used. The deformation model is also based on smooth affine transformations [ACP03]. The artifacts on the boundaries are due to the fact that correspondences to boundary edges were not pruned.

Besides the slight misalignment of the arm region for the pair of frames 0 and 6, the method breaks down more severely on the face region for the frames 0 and 12. The warped source shape is geometrically closer to the target scan than our approach. However the correspondences are semantically wrong as illustrated on the texture visualization in last row of Figure 9. We can depict semantic regions such as the mouth, the nose, the eyes, and the ears on the deformed source mesh. Ideally, the texture of each semantic region should correspond to its geometry. In addition to the slight geometric distortions on the nose, we observe that N-ICP 2 fails in matching the mouth region. The texture of the upper lip is matched to the geometry of the lower lip. Our method does not show any visible misalignments.

All methods break down for the registration between frame 0 and 20. As the amount of overlaps is too small for this pair of input data, our method cannot find reliable correspondences and a faithful alignment is not returned.

Table Cloth: We perform a comparison for the registration between two frames of a draping cloth falling onto an invisible sphere as shown in Figure 8. This example was created using a cloth simulation [GHF*07] and then depth sampled. The cloth input is particularly challenging as the deformation is more complex than for articulated objects and no stationary rigidness is present. Correspondence search is therefore more difficult and a correct registration would basically rely on an appropriate regularization during the deformation process. Unlike other non-rigid ICP methods, the alignment of our registration algorithm is able to recover the amount of isometry between the source and target shapes. The superior quality of our method is amplified when comparing the distortion of the checkers around the folded regions.

Limitations and Future Work. Our optimization relies on the fact that the target surface is parameterized. Although most acquisition systems provide a parameterization implicitly (e.g. depth maps or 2D sweep patterns), we intend to generalize our method to manifold surfaces using on-the-fly computed local parameterizations on the target surface for each correspondence point.

Combining the correspondence and deformation estimation into a single, non-linear optimization is essential for the effectiveness of our method. However, this global scheme leads to a comparatively high computational cost for real-time acquisitions. One important means to improve performance is the reduced deformable model that decouples the computational complexity from the size of the input scans. Currently we use a uniformly sampled graph that does not adapt to the geometry or the deformation of the processed data. Consequently, the smallest feature that we want to capture determines the resolution of the graph and thus leads to highly over-sampled graphs in mostly rigid regions. On the other hand, if the graph is too coarse, small-scale deformations cannot be captured accurately. This effect is noticeable for the bending arm example (see Figure 7), where the fingers are not appropriately matched due to inadequate resolution of the deformation graph.

Future work will address these issues and investigate the use of time- and space-adaptive deformation graphs. Apart from improving the quality of the results, such an adaptive scheme should lead to substantial performance gains by exploiting the spatial and temporal coherence of typical scan sequences. We also plan to investigate multi-view scanning systems and the integration of our approach with global methods for aligning entire time-sequences of scans.

10. Conclusion

We have introduced a registration algorithm for partial range scans of deforming shapes. Our method is designed for acquisition settings where the region of overlap is not known a priori, no explicit correspondences are provided, and minimal assumptions on the type of deformation are made. The key insight of this paper is that a simultaneous optimization of all of these properties yields a robust method that achieves semantically correct and accurate registrations even in the presence of considerable deformation and partial overlap. Ultimately, the problem of partial registration of deforming scans is ill-posed and no algorithm will be applicable for all acquisitions scenarios. We believe, however, that our method pushes the limit of what can be achieved with minimal prior information and provides an important building block of a comprehensive system for non-rigid shape acquisition.
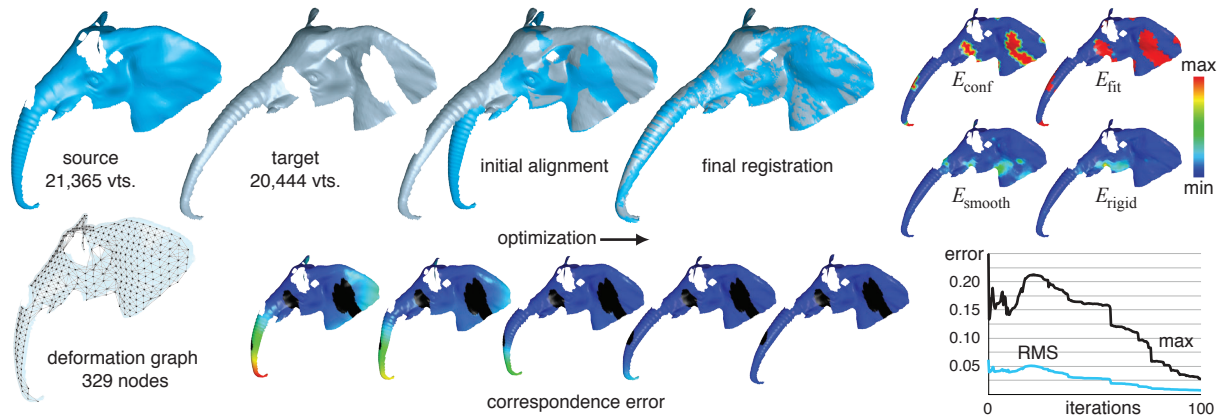
Figure 4: *Evaluation with synthetic data. The color-coded images on the top right show the different energy terms of the objective function. The bottom row visualizes how the distance to the ground truth correspondences evolves during the optimization, the graph shows the maximum and RMS error relative to the bounding box diagonal.*

References

[ACP03] ALLEN B., CURLESS B., POPOVIĆ Z.: The space of human body shapes: reconstruction and parameterization from range scans. *ACM Trans. Graph. 22*, 3 (2003).

[ARV07] AMBERG B., ROMDHANI S., VETTER T.: Optimal step nonrigid icp algorithms for surface registration. In *Proceedings of IEEE CVPR* (June 2007).

[ASK*05] ANGUELOV D., SRINIVASAN P., KOLLER D., THRUN S., RODGERS J., DAVIS J.: Scape: shape completion and animation of people. *ACM Trans. Graph. 24*, 3 (2005).

[ASP*04] ANGUELOV D., SRINIVASAN P., PANG H.-C., KOLLER D., THRUN S., DAVIS J.: The correlated correspondence algorithm for unsupervised registration of nonrigid surfaces. In *Advances in Neural Inf. Proc. Systems 17*. 2004.

[BBK06] BRONSTEIN A. M., BRONSTEIN M. M., KIMMEL R.: Generalized multidimensional scaling: a framework for isometry-invariant partial surface matching. *Proc. National Academy of Sciences (PNAS) 103* (2006).

[BR04] BROWN B., RUSINKIEWICZ S.: Non-rigid range-scan alignment using thin-plate splines. In *Symp. on 3D Data Processing, Visualization, and Transmission* (2004).

[BR07] BROWN B. J., RUSINKIEWICZ S.: Global non-rigid alignment of 3-d scans. *ACM Trans. Graph. 26*, 3 (2007), 21.

[BV99] BLANZ V., VETTER T.: A morphable model for the synthesis of 3D faces. In *Proceedings of SIGGRAPH* (1999).

[CM92] CHEN Y., MEDIONI G.: Object modelling by registration of multiple range images. *Image Vision Comput. 10*, 3 (1992), 145–155.

[GH97] GARLAND M., HECKBERT P. S.: Surface simplification using quadric error metrics. *Computer Graphics 31*, Annual Conference Series (1997), 209–216.

[GHF*07] GOLDENTHAL R., HARMON D., FATTAL R., BERCOVIER M., GRINSPUN E.: Efficient simulation of inextensible cloth. *ACM Trans. Graph. 26*, 3 (2007).

[GIRL03] GELFAND N., IKEMOTO L., RUSINKIEWICZ S., LEVOY M.: Geometrically stable sampling for the icp algorithm. In *Proc. Int. Conf. on 3D Digital Imaging and Modeling* (2003).

[HSZ87] HARALICK R. M., STERNBERG S. R., ZHUANG X.: Image analysis using mathematical morphology. *IEEE PAMI 9*, 4 (1987), 532–550.

[HTB03] HÄHNEL D., THRUN S., BURGARD W.: An extension of the ICP algorithm for modeling nonrigid objects with mobile robots. In *Proc. of IJCAI* (2003).

[IGL03] IKEMOTO L., GELFAND N., LEVOY M.: A hierarchical method for aligning warped meshes. In *Proc. of 4th Int. Conference on 3D Digital Imaging and Modeling* (2003).

[MFO*07] MITRA N. J., FLORY S., OVSJANIKOV M., GELFAND N., GUIBAS L., POTTMANN H.: Dynamic geometry registration. In *Symposium on Geometry Processing* (2007).

[MNT04] MADSEN K., NIELSEN H., TINGLEFF O.: *Methods for Non-Linear Least Squares Problems*. Tech. rep., Technical University of Denmark, 2004.

[PH06] PARK S. I., HODGINS J. K.: Capturing and animating skin deformation in human motion. *ACM Trans. Graph. 25*, 3 (2006).

[PHYH06] POTTMANN H., HUANG Q.-X., YANG Y.-L., HU S.-M.: Geometry and convergence analysis of algorithms for registration of 3d shapes. *Int. J. Comput. Vision 67*, 3 (2006).

[PMG*05] PAULY M., MITRA N. J., GIESEN J., GROSS M., GUIBAS L. J.: Example-based 3d scan completion. In *Symposium on Geometry Processing* (2005).

[RL01] RUSINKIEWICZ S., LEVOY M.: Efficient variants of the ICP algorithm. In *Proceedings of the Third Intl. Conf. on 3D Digital Imaging and Modeling* (2001).

[SG04] SCHENK O., GÄRTNER K.: Solving unsymmetric sparse systems of linear equations with pardiso. *Future Gener. Comput. Syst. 20*, 3 (2004), 475–487.

[SP04] SUMNER R. W., POPOVIĆ J.: Deformation transfer for triangle meshes. *ACM Trans. Graph. 26*, 3 (2004).

[SSP07] SUMNER R. W., SCHMID J., PAULY M.: Embedded deformation for shape manipulation. *ACM Trans. Graph. 26*, 3 (2007), 80.

[Wen05] WENDLAND H.: *Scattered Data Approximation*. Cambridge University Press, 2005.

[WJH*07] WAND M., JENKE P., HUANG Q., BOKELOH M., GUIBAS L., SCHILLING A.: Reconstruction of deforming geometry from time-varying point clouds. In *Symposium on Geometry processing* (2007), pp. 49–58.

[WLG07] WEISE T., LEIBE B., GOOL L. V.: Fast 3d scanning with automatic motion compensation. In *Proc. CVPR* (2007).
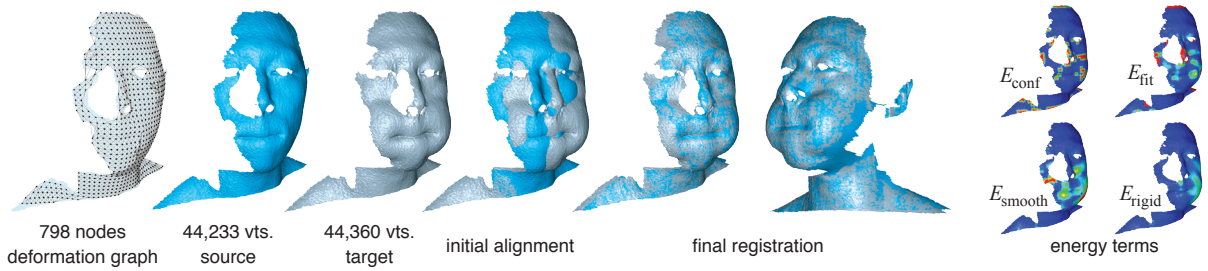
Figure 5: *Registration of a facial expression. The deformation energies E_{smooth} and E_{rigid} illustrate that most of the deformation is concentrated on the cheeks. This example also contains a substantial global rotation that is accurately solved for by the optimization.*
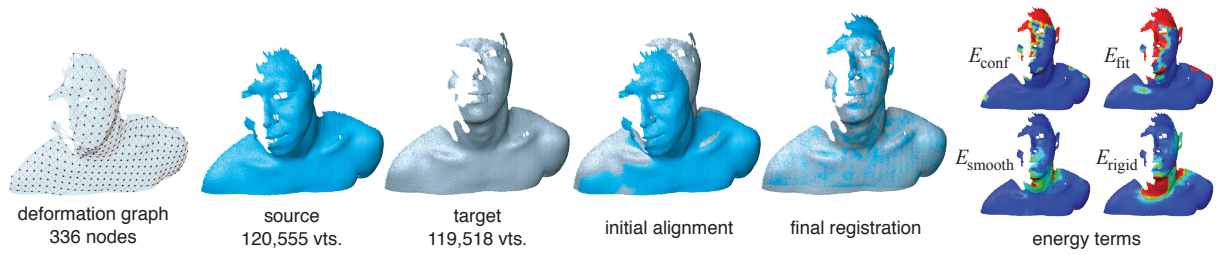
Figure 6: *Registration of a human torso. As indicated by E_{conf}, large regions in the face are only present in one of both models.*
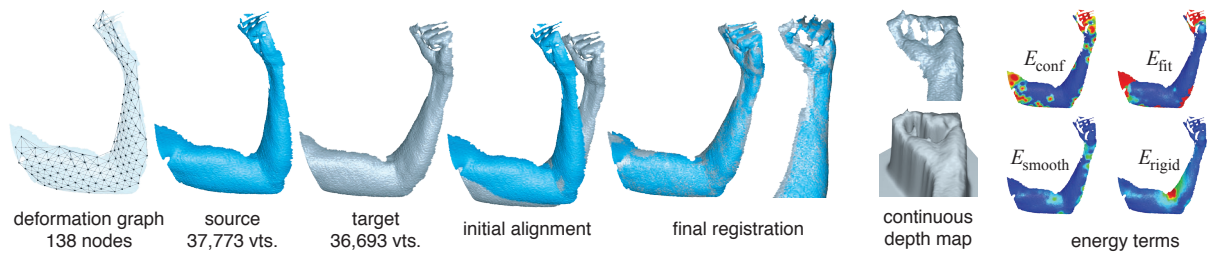

Figure 7: *Registration of a bending arm. The poor quality of the continuous weighted least squared approximation of the depth map of the fist explains why few reliable correspondences are found in that region.*
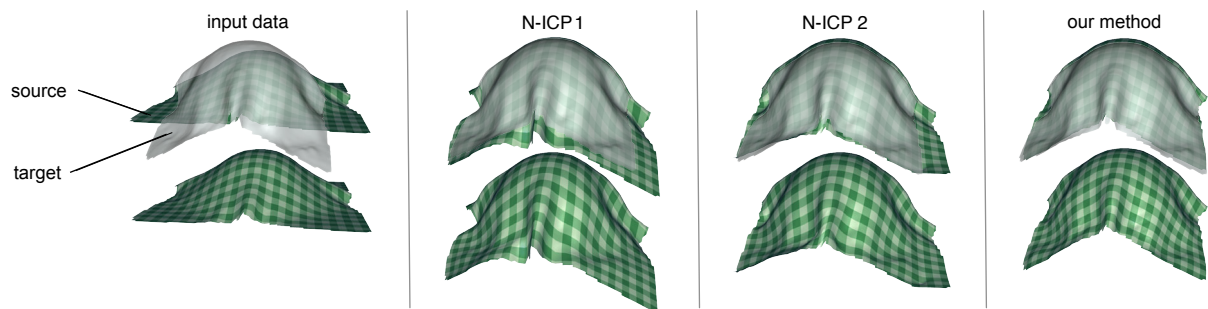
Figure 8: *Performance comparison of the registration between two depth scans of a draping table cloth. While non-rigid ICP methods fail to preserve isometry, our method converges to the correct result.*
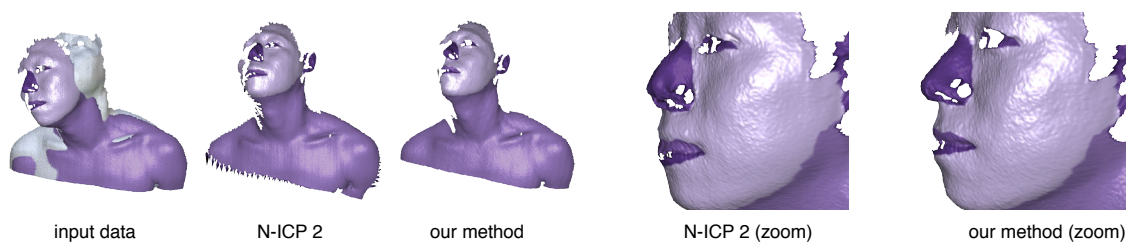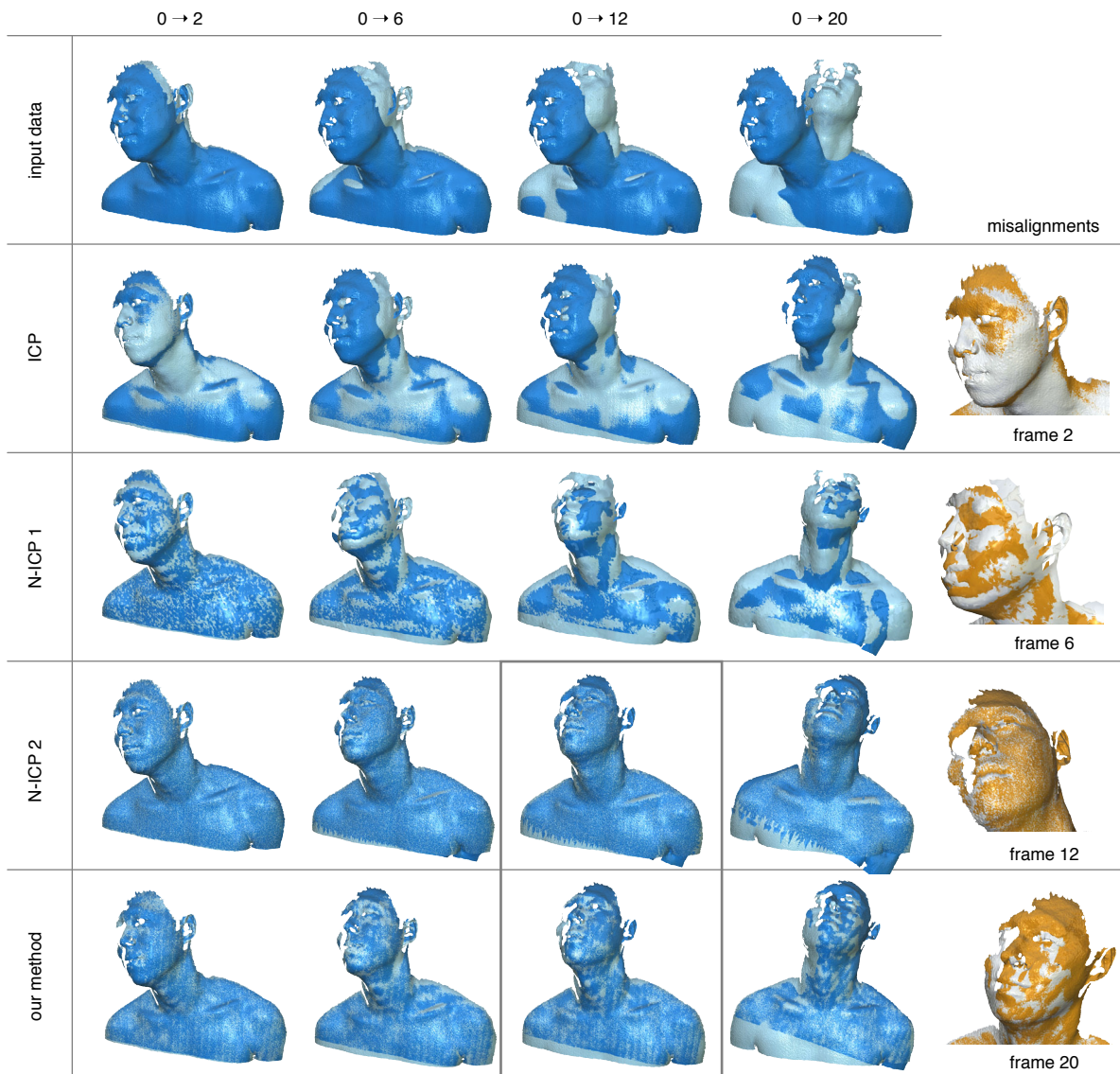
Figure 9: *We compare the performance of our method with two recent non-rigid ICP methods (blue) and highlight misalignments (yellow). While for the frames* 0 *and* 12 *the results of N-ICP 2 are geometrically closer to the target, the texture visualization below (purple) shows that our method yields a semantically more correct alignment.*