Robust Single-View Geometry and Motion Reconstruction

Hao Li* ETH Zurich
Bart Adams† KU Leuven
Leonidas J. Guibas‡ Stanford University
Mark Pauly§ ETH Zurich

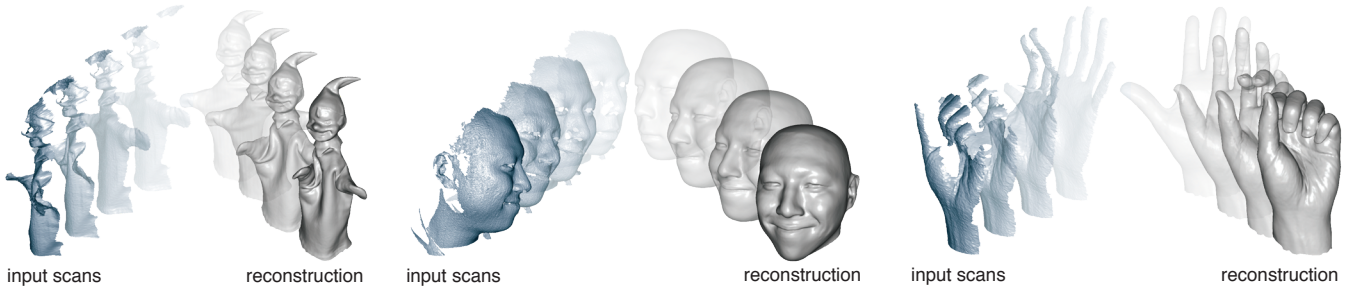

Figure 1: *Reconstruction of complex deforming objects from high-resolution depth scans. Our method accurately captures the global topology and shape motion, as well as dynamic, small-scale details, such as wrinkles and folds.*

Abstract

We present a framework and algorithms for robust geometry and motion reconstruction of complex deforming shapes. Our method makes use of a smooth template that provides a crude approximation of the scanned object and serves as a geometric and topological prior for reconstruction. Large-scale motion of the acquired object is recovered using a novel space-time adaptive, non-rigid registration method. Fine-scale details such as wrinkles and folds are synthesized with an efficient linear mesh deformation algorithm. Subsequent spatial and temporal filtering of detail coefficients allows transfer of persistent geometric detail to regions not observed by the scanner. We show how this two-scale process allows faithful recovery of small-scale shape and motion features leading to a high-quality reconstruction. We illustrate the robustness and generality of our algorithm on a variety of examples composed of different materials and exhibiting a large range of dynamic deformations.

Keywords: animation reconstruction, non-rigid registration, partial scans, 3D scanning, geometry synthesis, template tracking

1 Introduction

Accurate digitization of complex real-world objects is one of the central problems in visual computing. Commercial solutions for rigid objects are widely available and can be considered a mature technology. However, many of the assumptions of rigid scanning methods are no longer valid in a dynamic setting where the acquired shape is in motion and deforms. High temporal and spatial resolution is essential to faithfully recover the small-scale geometric detail that is often created as a result of the dynamic motion of the scanned model. Recent advances in 3D scanning technology facilitate the acquisition of dynamic objects, but pose substantial challenges for reconstruction algorithms. We consider the problem of marker-less, high-resolution geometry and motion reconstruction from single-view scans of a deforming shape. The main advantage of single-view 3D scanners is the simplicity of the acquisition setup, requiring no calibration or synchronization of multiple sensing units. However, single-view reconstruction of dynamic shapes is particularly challenging, since every scan covers a small section of the object's surface. Large and complex shape deformations constantly create or destroy geometric detail, such as wrinkles or folds in cloth, that needs to be distinguished from acquisition noise.

We address these challenges by introducing a novel template-based dynamic registration algorithm that offers significant improvements in terms of accuracy and robustness over previous methods. A key feature of our approach is the separation of large-scale motion from small-scale shape dynamics. We introduce a time- and space-adaptive deformation model that robustly captures the large-scale deformation of the object with minimal assumptions about the dynamics of the motion and without requiring an underlying physical model or kinematic skeleton. Our method dynamically adds degrees of freedom to the deformation model where needed, effectively extracting a generalized skeleton for the acquired shape. Small-scale dynamics are handled by a novel detail-synthesis method that computes a displacement field to adjust the deformed template to match the high-resolution input scans. The combination of these tools allows the efficient processing of extended scan sequences and yields a complete high-resolution geometry representation of the scanned object with full correspondences over all time instances.

We make a clear distinction between *static* and *dynamic* detail. Static detail includes all small-scale geometric features that are *persistent* in the shape and are not affected by the motion of the object. In the example shown in Figure 2, the mouth, eyes, and nose of the hand-puppet are static detail, since the entire face region is rigid. Dynamic detail consists of features that are *transient*. Deformation of the object can cause dynamic detail to appear and disappear, such as the folds in the body of the puppet. Our non-rigid registration method makes use of a template model to reconstruct the overall motion of the shape and provide a geometric prior for shape completion and topology control. In contrast to recent methods in performance capture [de Aguiar et al. 2008; Vlasic et al. 2008], we deliberately remove fine-scale detail from the template to avoid confusing static detail with dynamic detail. High-resolution templates from rigid scans typically have all detail "baked in", even transient features that are then erroneously transferred to all reconstructed surfaces (see also Figure 10). Our detail synthesis method automatically extracts detail from the high-resolution 3D input scans, propagates detail into occluded regions, and separates salient features from high-frequency noise.

*Applied Geometry Group, E-mail:hao@inf.ethz.ch
†Computer Graphics Group, E-mail:bart.adams@cs.kuleuven.be
‡Geometric Computing Group, E-mail:guibas@cs.stanford.edu
§Applied Geometry Group, E-mail:pauly@inf.ethz.ch

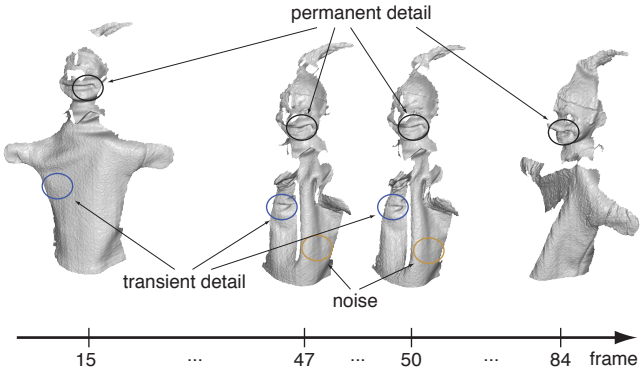

Figure 2: *Deforming shapes typically contain both permanent detail, such as the face region of the puppet, and transient detail, such as the dynamic folds in the cloth. Transient detail still persists over a number of adjacent frames and can thus be distinguished from temporally incoherent noise.*

Contributions. The methods we propose are general in that they are not specifically designed for a certain acquisition setup or particular motion models. Our tool requires no user interaction beyond aligning the template with the first scan and specifying a few global parameters. The main technical contributions of this paper are

- an efficient non-rigid registration method based on a non-linear deformation model that automatically adapts to the motion of the scanned object,
- a detail synthesis method that employs a spatio-temporal analysis of detail vectors to propagate detail into occluded regions and remove high-frequency acquisition noise,
- the integration of these methods into a complete 3D geometry and motion reconstruction framework.

The reconstructed surface meshes come with temporally consistent correspondences, which enables further applications such as mesh editing, texturing, or signal processing to be applied to the animation sequence. We demonstrate the versatility of our approach by showing high-resolution reconstructions of highly deformable shapes such as cloth, as well as the more coherent motion of articulated shapes. In addition, our purely data-driven algorithm is able to accurately reproduce subtle secondary motions such as hand tremor, or the behavior of complex materials such as the crumpling of a paper bag.

2 Related Work

Non-rigid registration methods were initially developed to align 3D scans of rigid objects that are distorted due to device non-linearities and calibration inaccuracies [Ikemoto et al. 2003; Brown and Rusinkiewicz 2004; Brown and Rusinkiewicz 2007]. These methods achieve highly accurate alignments for subtle warps, but are not suitable for large-scale deformations such as a bending arm.

Template-Based Methods. More general deformation models have been proposed to capture dynamic shapes [Allen et al. 2003; Sumner et al. 2007; Botsch and Sorkine 2008]. Various methods make use of a template model to simplify correspondence estimation and provide a prior for geometry and topology reconstruction, often relying on a small set of manually specified correspondences [Blanz and Vetter 1999; Allen et al. 2003; Pauly et al. 2005; Amberg et al. 2007]. Several unsupervised methods were proposed that require no manual intervention [Anguelov et al. 2004; Bronstein et al. 2006], but typically lead to higher computational complexity that makes these methods less suitable for long scan sequences.

Park and Hodgins [2006; 2008] developed a system that uses a very dense and large set of markers to capture and synthesize dynamic motions such as muscle bulging and flesh jiggling. While high resolution motions can be captured accurately, marker-based motion capture systems typically have a time-consuming calibration process and high hardware cost, and require actors to wear unnatural skin-tight clothing with optical beacons.

Marker-less methods are widely used in the acquisition and modeling of facial animations. In [Zhang et al. 2004], the deformation of an accurate face template is driven by time-coherent optical flow features and geometric closest point constraints. Since many features in a human face are persistent, their system can robustly handle long sequences of facial animations. More recently, several papers avoid the use of markers to reproduce complex animations of human performances and cloth deformations from multi-view video [Bradley et al. 2008; de Aguiar et al. 2008; Vlasic et al. 2008]. The latter two methods initialize the recording process with a high resolution full-body laser scan of the subject in a static pose. A low-resolution template model is created to robustly recover complex motions by combining various tracking and silhouette fitting techniques. Details of the high resolution models are then transferred back to the animated template. While large-scale deformations such as flowing garments are nicely captured, fine-scale geometric details such as folds that are not persistent in the surface are captured in the high-resolution model, remaining permanently throughout the reconstructed animation and possibly yielding unnatural deformations. An extension of this approach has been presented in [Ahmed et al. 2008] that follows a similar rationale to our method. A low-resolution template is tracked and subsequently enriched with local detail extracted from the acquired data. However, the specifics of this system differ substantially from our solution. The input stems from a multi-view acquisition system using eight video cameras, the template tracking is based on a shape-skeleton and silhouette matching, and the detail synthesis is performed based on surface normals reconstructed using shape from shading.

Registration Without A Template. Since creating an accurate and sufficiently detailed template of a deforming object can be difficult, various methods have been proposed that do not rely on a complete model. The algorithm presented by Mitra and colleagues [2007] aggregates all scans into a 4D space-time surface and estimates inter-frame motion from kinematic properties of the deforming surface. Süssmuth and coworkers [2008] introduced a space-time approach that first computes an implicit 4D surface representation. A template is extracted from the initial frame and warped to the subsequent frames by maximizing local rigidity. These methods require adjacent frames to be sufficiently dense in space and time and are mainly designed for articulated motions. Sharf and colleagues [2008] introduced a volumetric space-time reconstruction technique that represents shape motion as an incompressible flow of material through time. This strong regularization makes the method particularly suitable for very noisy input data. Wand and coworkers [2007] introduced a statistical framework that performs pairwise alignment and merging over all adjacent scans within a global non-linear optimization process, leading to high computational cost. While significant performance improvements were achieved in a follow-up work using a volumetric meshless deformation model [Wand et al. 2009], this approach is still substantially slower than our method. In addition, the lack of a template can lead to topological ambiguities and misalignments for unseen parts (see also Figure 14).

Several researchers have proposed *pairwise* non-rigid registration algorithms that are specifically designed for large deformations. Li and coworkers [2008] developed a registration framework that simultaneously solves for point correspondences, surface deformation, and region of overlap within a single global optimization. Our registration method uses similar components, but avoids the coupled non-linear optimization of correspondences and deforma-

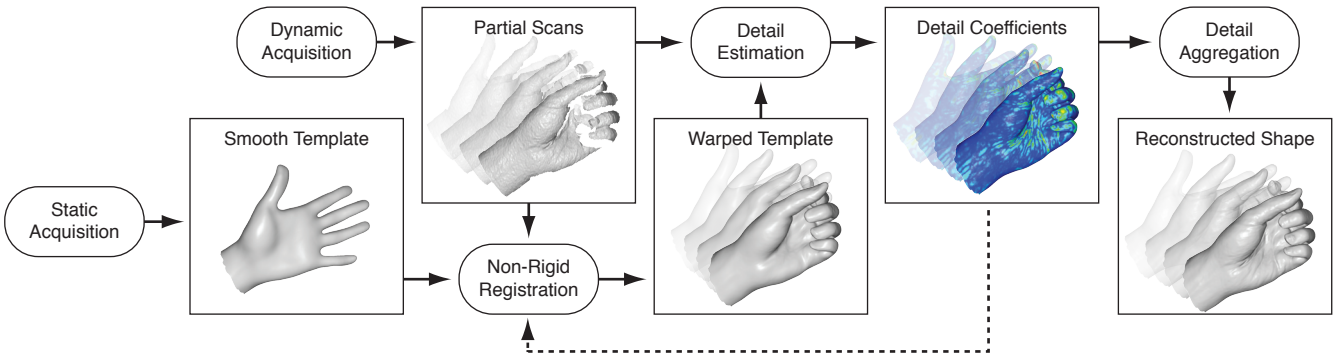

Figure 3: *Processing pipeline. A smooth template mesh is registered to each of the input scans using a non-linear, adaptive deformation model. Small-scale detail coefficients are estimated and integrated into the template. The final reconstruction is obtained through detail aggregation and filtering to propagate detail into occluded regions and separate salient features from noise.*

tion to obtain an efficient alignment method for extended scan sequences. Chang and Zwicker [2008] solve a discrete labeling problem to detect the set of optimal correspondences and apply graph cuts to optimize for a consistent deformation from source to target. They extend their scheme in [Chang and Zwicker 2009] using a reduced space deformation model represented by a volumetric grid that encloses the underlying scan. Although significant motion and occlusions can be handled, their deformation field representation breaks down for topologically difficult scenarios such as shapes with nearby or touching surfaces. Huang and colleagues [2008] suggested a registration technique that finds an alignment by diffusing consistent closest point correspondences over the target shape while preserving isometries as much as possible. Their implementation has proved to be efficient for large isometric deformations, yet the correspondence search is sensitive to topological changes and holes that commonly occur in partial acquisition systems.

3 Overview

Our dynamic acquisition system shown in Figure 4 provides dense depth maps with a spatial resolution of 0.5 mm at 25 frames per second (see [Weise et al. 2007] for a description of a similar acquisition setup). This allows us to capture fine-scale geometric detail of deforming objects at high temporal resolution. However, input scans are typically highly incomplete and contain considerable amounts of measurement noise. We found that a template model is essential as a geometric and topological prior for the robust reconstruction of shapes that undergo complex deformations, in particular for single-view acquisition, where large parts of the object are occluded.

Figure 3 gives an overview of our processing pipeline. Static acquisition is used to reconstruct the initial template. We remove all high-frequency detail from the template using low-pass filtering to avoid transferring potentially transient features to future scans. This significantly simplifies template construction since we do not require high geometric precision. To initialize the computations, we manually specify a rigid alignment of the template to the first frame of the scan sequence and apply one step of the pairwise non-rigid registration method described in Section 4.

We propose a two-scale approach to reconstruct a complete and consistent surface for each frame. Template registration uses a non-linear reduced deformable model to recover the large-scale motion and align the template to each of the input scans (Section 4). The template-to-scan registration makes use of detail coefficients estimated in the previous frame to enable feature locking and improve the alignment accuracy. The final reconstruction is then obtained using a separate detail synthesis pass that runs once forward and once backward in time to aggregate and propagate detail into occluded regions (Section 5).

4 Template Registration

The registration stage captures the large-scale motion of the subject by fitting a coarse template shape to every frame of the scan sequence. Scans do not have to be a subset of the geometry described by the template, as in most previous methods (e.g. [Allen et al. 2003]). Our method robustly handles part-in-part registration, as opposed to the simpler part-in-whole matching (see e.g. Figure 12). We assume minimal prior knowledge about the acquired motion and thus employ a general deformation model to capture a sufficiently large range of shape deformations. We extend the embedded deformation framework proposed in [Sumner et al. 2007] to automatically adapt to the motion of the captured data. This allows recovering unknown complex material behavior and improves the robustness and efficiency of the registration.

4.1 Surface Deformation Model

Embedded deformation computes a warping field using a deformation graph to discretize the underlying space. Each node $\mathbf{x}_i$ of the graph induces a deformation within a local influence region of radius r_i. We represent such a local deformation as an affine transformation specified by a 3×3 matrix $\mathbf{A}_i$ and a 3×1 translation vector $\mathbf{b}_i$. Graph nodes are connected by an edge whenever two nodes influence the same vertex of the mesh. A vertex $\mathbf{v}_j$ of the embedded shape is mapped to the position

$$\mathbf{v}'_j = \sum_{\mathbf{x}_i} \overline{w}(\mathbf{v}_j, \mathbf{x}_i, r_i) \left[\mathbf{A}_i(\mathbf{v}_j - \mathbf{x}_i) + \mathbf{x}_i + \mathbf{b}_i\right], \qquad (1)$$

where $\overline{w}(\mathbf{v}_j, \mathbf{x}_i, r_i)$ are the normalized weights $w(\mathbf{v}_j, \mathbf{x}_i, r_i) = \max(0, (1 - d^2(\mathbf{v}_j, \mathbf{x}_i)/r_i^2)^3)$ with $d(\mathbf{v}_j, \mathbf{x}_i)$ the distance between $\mathbf{v}_j$ and $\mathbf{x}_i$. We exploit the topological prior of the template and replace Euclidean distances in the original formulation by

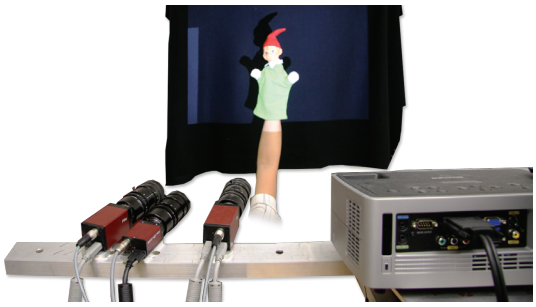

Figure 4: *Our real-time structured light scanner based on active stereo delivers high resolution input scans from a single view.*

geodesic distances measured on the template mesh. This improvement avoids distortion artifacts that often occur when geodesically distant parts of the object come into close contact (Figure 8). We use a variant of the fast marching method to efficiently compute approximate geodesic distances [Kimmel and Sethian 1998].

During non-rigid registration we solve for the unknown transformations $(\mathbf{A}_i, \mathbf{b}_i)$. A feature preserving deformation field is obtained by maximizing local rigidity using the energy

$$E_{\text{rigid}} = \sum_{\mathbf{x}_i} \Big((\mathbf{a}_1^{\mathsf{T}}\mathbf{a}_2)^2 + (\mathbf{a}_1^{\mathsf{T}}\mathbf{a}_3)^2 + (\mathbf{a}_2^{\mathsf{T}}\mathbf{a}_3)^2 + (1-\mathbf{a}_1^{\mathsf{T}}\mathbf{a}_1)^2 + (1-\mathbf{a}_2^{\mathsf{T}}\mathbf{a}_2)^2 + (1-\mathbf{a}_3^{\mathsf{T}}\mathbf{a}_3)^2 \Big) \tag{2}$$

that measures the deviation of the column vectors $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$ of $\mathbf{A}_i$ from orthogonality and unit length. An additional regularization term ensures smoothness of the deformation. We extend the original formulation of [Sumner et al. 2007] using the geodesic distance weights to handle non-uniformly sampled graph nodes:

$$E_{\text{smooth}} = \sum_{\mathbf{x}_i} \sum_{\mathbf{x}_j} \overline{w}(\mathbf{x}_i, \mathbf{x}_j, r_i + r_j) \, \| \mathbf{A}_i(\mathbf{x}_j - \mathbf{x}_i) + \mathbf{x}_i + \mathbf{b}_i - (\mathbf{x}_j + \mathbf{b}_j) \|_2^2 . \tag{3}$$

Minimizing these combined energies with the fitting term defined below yields affine transformations for each node, which in turn define a smooth deformation field for the template mesh. We solve this non-linear problem using a standard Gauss-Newton algorithm as described in [Sumner et al. 2007; Li et al. 2008].

4.2 Robust Pairwise Registration

We iteratively compute closest point correspondences in the spirit of non-rigid ICP methods, followed by a pruning and deformation step. To avoid local minima in the non-linear optimization, we use the simple yet effective technique proposed by [Li et al. 2008] that progressively relaxes the regularization energies of the deformation model. Similar strategies were also applied in [Allen et al. 2003; Amberg et al. 2007]. In this way, the template can be accurately aligned to scans that undergo considerable deformations without the use of sparse, high-dimensional features.

Since our input data is sufficiently coherent in time, we repeatedly use closest point correspondences between the template and each input scan to determine the optimal deformation. In order to obtain an accurate fit, we augment the smooth template with detail information extracted from the previous frame. Template vertices $\mathbf{v}_i^j$ of frame j are displaced in the direction of the corresponding surface normal $\mathbf{n}_i^j$ yielding $\tilde{\mathbf{v}}_i^j = \mathbf{v}_i^j + d_i^{j-1}\mathbf{n}_i^j$, where d_i^{j-1} is the detail coefficient of frame $j-1$ (see Section 5). The correspondence energy combines the point-to-point and the point-to-plane metric to avoid incorrect correspondences in large featureless regions:

$$E_{\text{fit}} = \sum_{(\mathbf{v}_i^j, \mathbf{c}_i^j) \in \mathcal{C}} \alpha_{\text{point}} \left\| \tilde{\mathbf{v}}_i^j - \mathbf{c}_i^j \right\|_2^2 + \alpha_{\text{plane}} |\mathbf{n}_{\mathbf{c}_i^j}^{\mathsf{T}} (\tilde{\mathbf{v}}_i^j - \mathbf{c}_i^j)|^2, \tag{4}$$

where $\mathbf{c}_i^j$ denotes the closest point on the input scan from $\tilde{\mathbf{v}}_i^j$ with corresponding surface normal $\mathbf{n}_{\mathbf{c}_i^j}$. We use $\alpha_{\text{point}} = 0.1$ and $\alpha_{\text{plane}} = 1$ in all our experiments. Correspondences are discarded if they are too far apart, have incompatible normal orientations, lie on the boundary of the partial input scans, or stem from back-facing or self-occluded vertices of the template.

Iterative Optimization. For each template-to-scan alignment, we initialize the registration with high stiffness weights $\alpha_{\text{smooth}} = 10$ and $\alpha_{\text{rigid}} = 100$. We then alternate in each iteration between correspondence computation and template deformation by minimizing

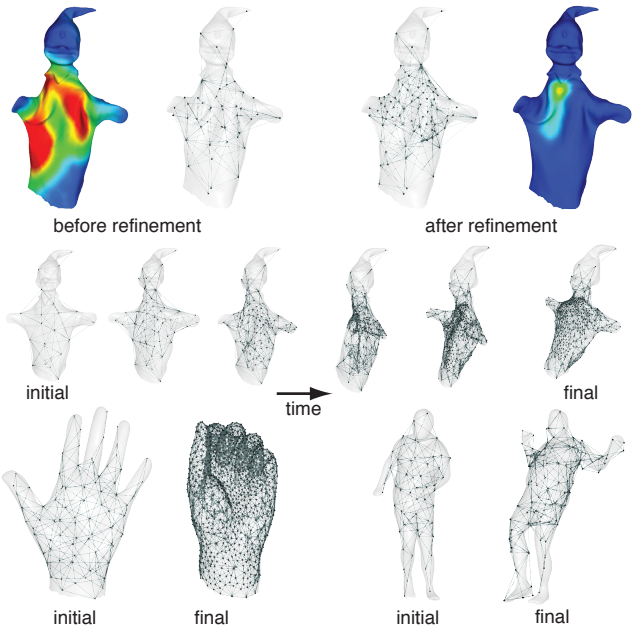

Figure 5: *The deformation graph is dynamically refined during non-rigid registration to adapt to the deformation of the scanned object. Color-coded images indicate the regularization energy that determines where new nodes are added to the graph. The bottom row shows the initial and final deformation graphs for the hand and the sumo reconstruction.*

$E_{\text{tot}} = E_{\text{fit}} + \alpha_{\text{smooth}} E_{\text{smooth}} + \alpha_{\text{rigid}} E_{\text{rigid}}$. If the relative total energy did not change significantly between iterations j and $j+1$ (i.e., $|E_{\text{tot}}^{j+1} - E_{\text{tot}}^{j}| / E_{\text{tot}}^{j} < \sigma$), we additionally relax the regularization weights to $\alpha_{\text{smooth}} \leftarrow \frac{1}{2}\alpha_{\text{smooth}}$ and $\alpha_{\text{rigid}} \leftarrow \frac{1}{2}\alpha_{\text{rigid}}$. This relaxation strategy effectively improves the robustness by avoiding suboptimal local minima and allows handling pairs of scans that undergo significant deformations. In all our experiments we use $\sigma = 0.005$. The iterative optimization is repeated until $\alpha_{\text{rigid}} < 0.1$ or until a maximum number of iterations $N_{\max} = 100$ is reached.

Note that detail information of the previous frame is only used to improve the accuracy of the registration by enabling geometric feature locking. The resulting continuous space deformation is applied to the template vertices without added detail. As discussed in Section 5 the final detail coefficients are obtained through a separate detail synthesis pass.

4.3 Dynamic Graph Refinement

We replace the static, uniform sampling of the deformation graph in [Sumner et al. 2007] and [Li et al. 2008] with a spatially and temporally adaptive node distribution. While the idea of adaptive mesh deformation has been explored in previous work, for instance in the context of multi-resolution shape modeling from images [Zhang and Seitz 2000], we propose to adapt the degrees of freedom of the deformation model instead of the geometry itself in order to improve registration robustness and efficiency.

A hierarchical graph representation is pre-computed from a dense uniform sampling of graph nodes by successively merging nodes in a bottom-up fashion. The initial uniform node sampling corresponds to the highest resolution level $l = L_{\max}$ of the deformation graph that we restrict to roughly one tenth of the number of mesh vertices. We thus avoid over-fitting in regions of small-scale deformations, which are instead captured by our detail synthesis method (Section 5). We uniformly sub-sample the nodes of each level by repeatedly increasing their average sampling distance $r_{l-1} = 4\, r_l$

until l reaches L_{min}. Each of the remaining nodes $\mathbf{x}_i^l$ from level $l \in L_{\text{min}} \ldots L_{\text{max}}$ form a cluster $\mathcal{C}_i^l$ which contains every node from the level below $\mathbf{x}_i^{l+1}$ that is not closer to any other cluster from l. The resulting cluster hierarchy is then used for adaptive refinement. We choose $L_{\text{min}} = L_{\text{max}}/2$ for all our experiments.

Refinement Criterion. Registration starts with a coarse uniform graph at level L_{min} and dynamically adapts the graph resolution by inserting nodes in regions with high regularization residual (E_{smooth}), which indicates a strong discrepancy of neighboring node transformations (see Figure 5). In all our examples we set the threshold for refinement to 10% of the highest regularization value. One step of refinement substitutes every $\mathbf{x}_i^l$ that exhibits high regularization with all nodes contained in $\mathcal{C}_i^l$. To avoid unnecessary refinements for every new upcoming target frame, adaptive refinement is only performed if the global regularization term is still above a certain threshold, i.e. $E_{\text{smooth}} > 0.01$, for the maximum number of iteration $N_{\text{max}} = 100$ of pairwise registration.

The dynamic refinement effectively learns an adaptive deformation model that is consistent with the motion of the scanned object. Additional nodes will be inserted automatically in regions of high deformation, while large rigid parts can be accurately deformed by a single graph node. In addition to being less susceptible to local minima, this leads to significant performance improvements (up to a factor of four in our examples) as compared to a uniform sampling with a high level of node redundancy. As illustrated in Figure 5, our adaptive model is suitable for a wide variety of dynamic objects, from articulated shapes to complex cloth folding.

4.4 Multi-Frame Stabilization

The warped template $\mathcal{T}^{j-1}$ obtained after alignment to scan $j-1$ is the zero-energy state when aligning to scan j for each frame of the entire template warping process. For surface regions that are visible in the scan, dynamic details, such as cracks and fissures in paper-like materials can be accurately captured, since the method prevents the template from deforming back to its initial undeformed state. However, unobserved template parts are inherently prone to accumulation of misalignments, especially for lengthier scan sequences as illustrated in Figure 6. In contrast to our formulation, classical template fitting methods [Zhang et al. 2004; de Aguiar et al. 2008; Vlasic et al. 2008] warp the same initial template to each recorded frame and thus, use a deformation model that behaves globally elastic in time. For complex articulated subjects, such as human bodies, missing data in occluded regions would pull the template back to its original shape, which can be very different to the one of the current frame. Therefore, multi-view acquisition systems are usually used in combination with sparse and robust feature tracking [de Aguiar et al. 2008] and sometimes enhanced with manual intervention [Vlasic et al. 2008] to ensure reliable tracking.

In our dense acquisition setting, the surface coverage of the template by the input scans is spatially and temporally coherent over time. Thus, for non-occluded regions, the template shape from a closer time instance represents in general a more likely shape prior than the initial template $\mathcal{T}_{\text{init}}$. On the other hand, we make the assumption that no better knowledge exists than $\mathcal{T}_{\text{init}}$ for template regions that are never observed or not seen for an extended period.

To address this issue we introduce a time-dependent combination of plastic and elastic deformation to accurately track exposed surface regions and reduce the accumulation of errors in less recently observed parts of the scanned object. After the pairwise registration of $\mathcal{T}^{j-1}$ to scan j as presented in Section 4.2, we obtain the plastically deformed template $\mathcal{T}^j$. A weight c_i^j for visibility confidence can then be defined for each vertex $\mathbf{v}_i^j \in \mathcal{T}^j$ as $c_i^j = \max\{0, (P + j_i^{\text{last}} - j)/P\}$ with j_i^{last} the last frame where $\mathbf{v}_i$ has been observed, and P a constant (we chose $P = 30$ in all our examples) that defines a temporal confidence range of visibility. All template vertices with $c_i^j = 1$ are visible in the current frame, while $c_i^j = 0$ represent those that are no longer considered confident. For the same frame, an elastically deformed template $\tilde{\mathcal{T}}^j$ with vertices $\tilde{\mathbf{v}}_i^j$ is created by warping $\mathcal{T}_{\text{init}}$ to the current frame j using the linearized thin-plate energy as described in [Botsch and Sorkine 2008]. Hard positional constraints are defined for all vertices with confidence $c_i^j = 1$. The resulting template $\bar{\mathcal{T}}^j$ with vertices $\bar{\mathbf{v}}_i^j$ is obtained by linearly blending $\mathcal{T}^j$ and $\tilde{\mathcal{T}}^j$ with the confidence weights for visibility yielding the vertices $\bar{\mathbf{v}}_i^j = c_i^j \mathbf{v}_i^j + (1 - c_i^j)\tilde{\mathbf{v}}_i^j$.

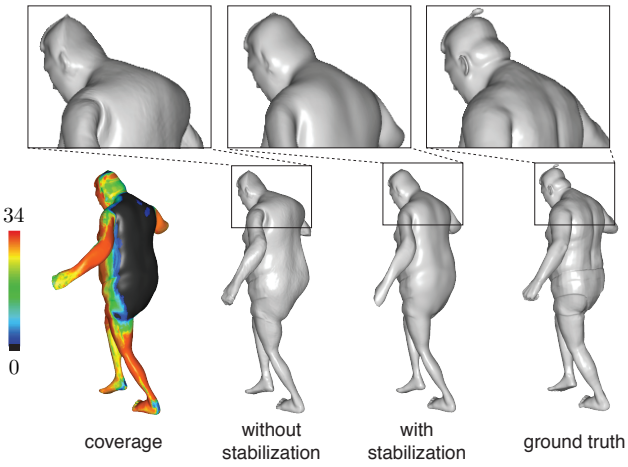

Figure 6: *A hybrid plastic and elastic deformation model is used to stabilize the registration for multiple input frames as repeated pairwise alignment is susceptible to error accumulation. The accumulation of misalignments is shown on frame 30 of the sumo sequence.*

5 Detail Synthesis

Non-rigid registration aligns the template sequentially with all input scans. The resulting deformation fields induced by the graph capture the large-scale deformation but might miss small deformations that give rise to dynamic detail such as wrinkles and folds. To recover fine-scale detail at the spatial resolution of the scanner, we perform a separate detail synthesis stage that is composed of two steps: First, a per-vertex optimization from local correspondences is applied to estimate detail coefficients for each vertex of the template. These preliminary detail coefficients are the ones used for template alignment as detailed in Section 4. After the template has been registered to the entire scan sequence, we perform an additional pass that exploits the temporal coherence of the scan sequence to improve the reconstruction quality by propagating detail into occluded regions.

Linear Mesh Deformation. Since the deformed template is already well-aligned with the input scan, we employ an efficient linear mesh deformation algorithm similar to [Zhang et al. 2004] to estimate detail coefficients. For each vertex $\mathbf{v}_i$ in the template mesh, we trace an undirected ray in normal direction $\mathbf{n}_i$ and find the closest intersection point on the input scan. In case an intersection point $\mathbf{c}_i$ is found, a point-to-point correspondence constraint is created, if both points have the same normal orientation and are sufficiently close. Since the template has no high-frequency detail, its normal vector field is smooth, leading to spatially coherent correspondences. We compute the detail coefficients d_i by minimizing the energy resulting from the extracted correspondences subject to a regularization constraint

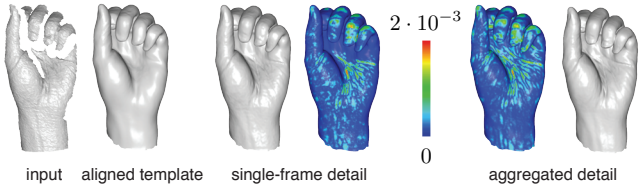

Figure 7: *Detail synthesis. Reconstructing detail from the current frame leads to lack of detail in occluded regions. Aggregating detail over temporally adjacent frames propagates detail into hole regions and reduces noise. The color-coded images show the magnitude of the detail coefficients relative to the bounding box diagonal.*

$$E_{\text{detail}} = \sum_{i \in \mathcal{V}} \|\mathbf{v}_i + d_i \mathbf{n}_i - \mathbf{c}_i\|_2^2 + \beta \sum_{(i,j) \in \mathcal{E}} |d_i - d_j|^2, \qquad (5)$$

where $\mathcal{V}$ and $\mathcal{E}$ are index sets of mesh vertices and edges, respectively. The parameter β balances detail synthesis with smoothness and is set to $\beta = 0.5$ in all our experiments. The resulting system of equations is linear and sparse and can thus be solved efficiently.

Aggregation. The linear mesh deformation method described above estimates detail coefficients independently for each frame in those regions of the object that are observed by a particular scan. To transfer detail to occluded regions we perform a separate processing pass that aggregates detail coefficients using a so-called *exponentially weighted moving average*. We use the formulation of Roberts [1959] and define this moving average as

$$\overline{d}_i^j = (1 - \gamma)\overline{d}_i^{j-1} + \gamma d_i^j \qquad (6)$$

with γ set to 0.5 in all our examples. The influence of past detail coefficients decays quickly in this formulation, which is important, since transient or dynamic detail such as wrinkles and folds might not persist during deformation. Note that details in the template only disappear when they vanish in the input scans of succeeding frames. For instance, the details of a rigid object will persist and not fade toward zero coefficients since only observed coefficients are combined during detail synthesis. When processing scan j, we first update the vertices $\mathbf{v}_i^j \leftarrow \mathbf{v}_i^j + \overline{d}_i^{j-1}\mathbf{n}_i^j$ and perform the linear mesh deformation described in the previous section. This yields the new detail coefficients d_i^j that are then used to update the moving average $\overline{d}_i^j$, which will in turn be employed to process the subsequent scans. The entire detail aggregation process is performed by running sequentially once forward and once backward through the scans while performing the linear mesh deformation and updating the moving averages. Going back and forth allows us to back-propagate persistent details seen at future instances to earlier scans (see Figure 7). As a final step, we apply a band-limiting bilateral filter [Aurich and Weule 1995] that operates in the time domain and detail range to further reduce temporal noise.

6 Results

We show a variety of acquired geometry and motion sequences processed with our system that exhibit substantially different dynamic behavior. Accurate reconstruction of these objects is challenging due to the high noise level in the scans, missing data caused by occlusions or specularity, unknown correspondences, and the large and complex motion and deformations of the acquired objects. The statistics for the results are shown in Table 1. All templates were constructed by performing an online rigid registration technique similar to [Rusinkiewicz et al. 2002] on our acquired data, followed by a surface reconstruction technique based on algebraic point set surfaces described in [Guennebaud and Gross 2007]. Given the roughly aligned template mesh, our system runs completely automatically without any user intervention. Only few parameters (such as the weighting coefficients of the different energy terms) have to be chosen manually. For all examples, we use the same initial parameter settings. During optimization we automatically adapt the parameters using the approach detailed in Section 4.

Figure 9 shows the warped template and final reconstruction of the puppet. This example is particularly difficult due to the close proximity of multiple surface sheets when closing the puppet's hands. The reconstruction of a hand in Figure 8 demonstrates that our detail synthesis method is capable of capturing the intricate folds and wrinkles of human skin, even though the scans contain a large amount of measurement noise. Figure 12 illustrates how detail is propagated correctly into occluded regions, which leads to a plausible high-resolution reconstruction even for parts of the model that have not been observed in a particular scan. Figure 13 shows the reconstruction of a crumpling paper bag. Despite substantial holes

	Puppet	Head	Hand	Paper Bag	Sumo
# Scans	100	200	35	85	34
Min # Points per Scan	23k	53k	19k	82k	85k
Max # Points per Scan	37k	68k	25k	123k	86k
Input Data Size (Mb)	430	1,690	120	145	430
# Template Vertices	48k	64k	46k	64k	107k
Begin # Graph Nodes	20	152	77	37	52
End # Graph Nodes	100	458	1238	86	110
Output Data Size (Mb)	530	2,030	180	960	540
Registration Time	39	247	15	65	26
Detail Synthesis Time	26	92	8	36	23
Total Time	65	339	23	101	49

Table 1: *Statistics for the results shown in this paper. All computations were performed on a 3.0 GHz Dual Quad-Core Intel Xeon machine with 8 GB RAM. Timings are measured in minutes and include I/O operations.*

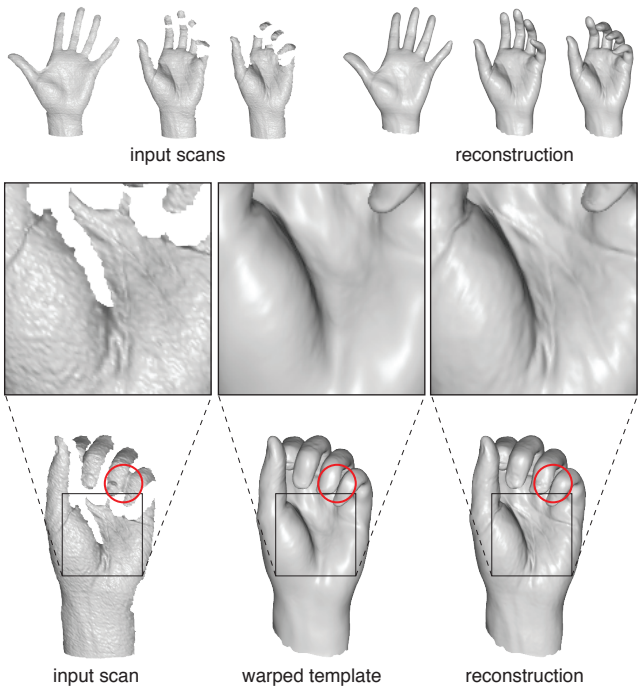

Figure 8: *The zooms illustrate how high-frequency detail such as the skin folds is faithfully recovered and transferred to occluded regions. Even though the scan is connected at the fingertips, shape topology is correctly recovered (red circle).*

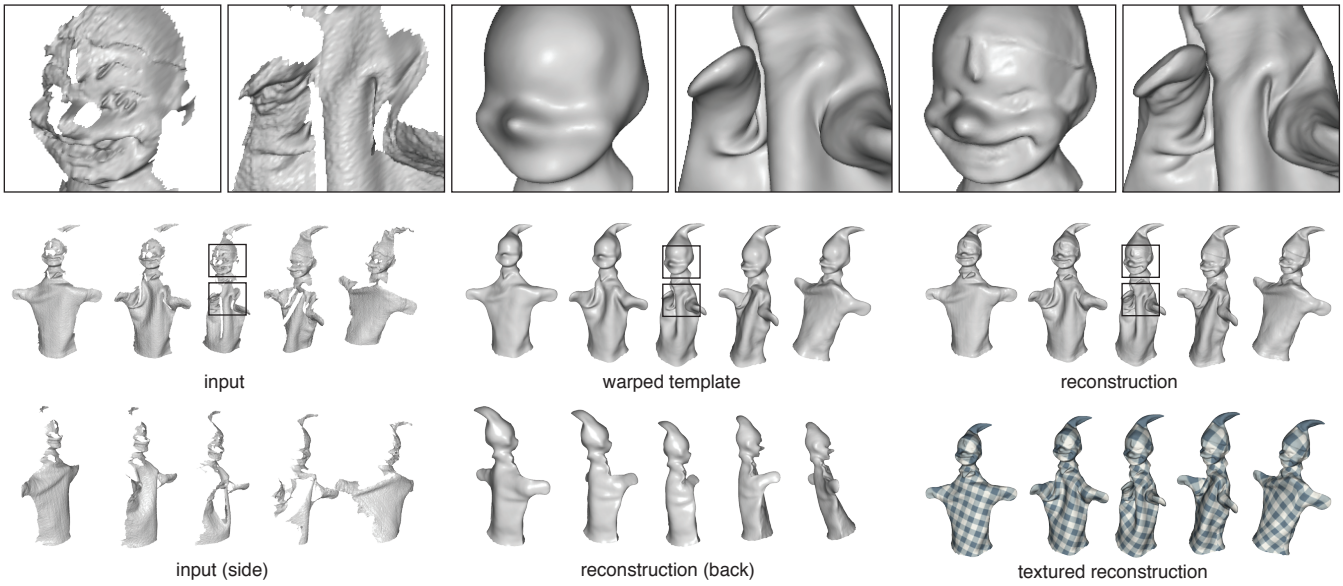

Figure 9: *The global motion of the puppet's shape as well as fine-scale static and dynamic detail are captured accurately using the template registration and detail synthesis algorithm. The intricate folds of the cloth are handled robustly in the registration.*

caused by oversaturation in the reflections, the dynamics of the material as well as sharp geometric creases are faithfully captured.

7 Evaluation

Figure 10 illustrates the difference between tracking a high-resolution template versus our two-scale approach that separates global shape motion and dynamic detail reconstruction. For comparison we use the first frame of our two-scale reconstruction as the high-resolution template, which is then aligned with the input scan sequence using the registration method of Section 4. As can be seen in the zoom, dynamic detail created by the motion, in particular in the cloth, is not captured accurately. In contrast, our detail synthesis approach avoids the artifacts created by "baked-in" geometric detail and leads to a high-quality reconstruction of both static and dynamic detail. While a fairly large range of template smoothness can be tolerated, an overly coarse template can deteriorate the reconstruction as shown in Figure 11.

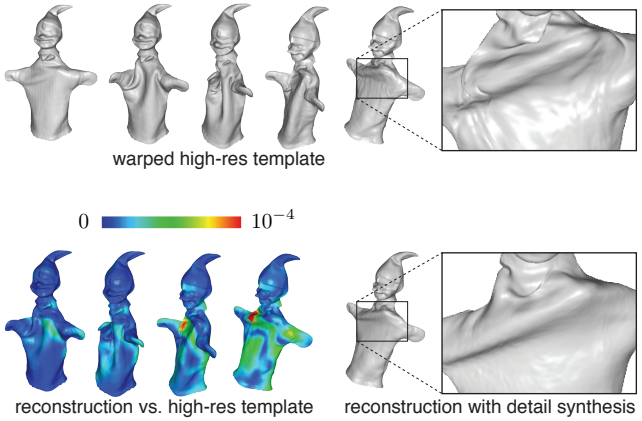

Figure 10: *Warping a high-resolution template without detail synthesis leads to inferior results as compared to our two-scale reconstruction approach (cf. Figure 9). The color coding shows the distance between both results relative to the bounding box diagonal.*

The necessity of using a template for robust reconstruction of complex deforming shape is illustrated in Figure 14. The method of [Wand et al. 2009] that avoids the use of a template cannot track the motion of the fingers accurately. In particular, the correspondence estimation fails when previously unseen parts of the shape, such as the back of the fingers, come into view. Figure 15 shows a comparison of our method to the dynamic registration approach of [Süssmuth et al. 2008] using the same template in both reconstructions.

We evaluate the robustness of the template tracking and detail synthesis method using the ground truth comparison shown in Figure 16. The scanning process has been simulated by creating a set of artificial depth maps from a fixed viewpoint. The ground-truth animation of the 3D model was obtained from dense motion capture data provided by [Park and Hodgins 2006]. In order to test the stability of the template tracking, we sampled the entire sequence at successively lower temporal resolution. The non-rigid

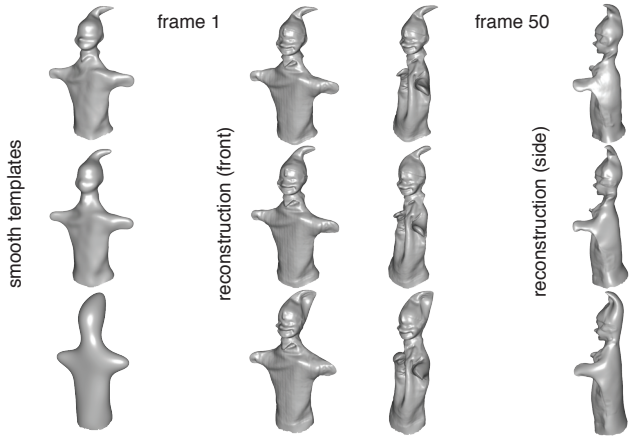

Figure 11: *Evaluation of the reconstruction (frame 1 and 50) for three different initial templates. The upper row shows the original template. The coarser template in the second row is produced by surface reconstruction from points that are uniformly subsampled at half of the density of the original template. The last row illustrates the reconstruction using an even coarser template. This is obtained from only 25% of the initial point density.*

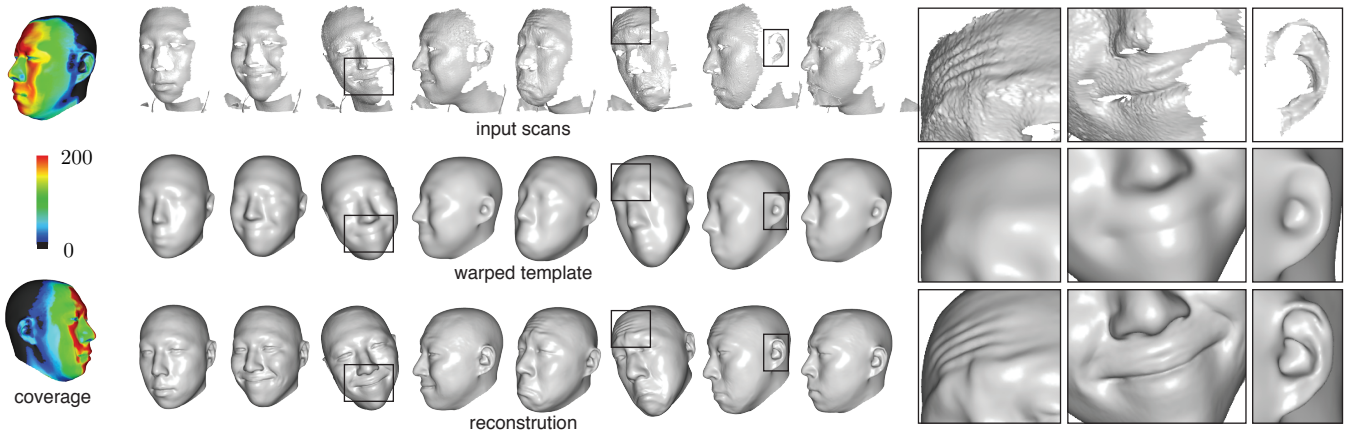

Figure 12: *Our method faithfully recovers both the large-scale motion of the turning head, as well as the dynamic features created by the expression, such as wrinkles on the forehead or around the mouth. Intricate geometric details such as the ears are accurately captured, even though they are only observed in few frames. The color-coded images show the number of frames a certain region has been observed.*

registration robustly aligns the template with the scans for a temporally sub-sampled sequence consisting of only 34 frames. The large inter-frame motion, especially of the arms and legs, is tracked correctly, even though our correspondence computations do not make use of feature points, markers, or user assistance. Template tracking breaks down at 17 frames, where the fast motion of the arms cannot be recovered anymore (see Figure 17 (a)). Detail synthesis for the 34-frame sequence reliably recovers most of the fine-scale geometry correctly. Artifacts appear in the fingers and toes due to the coarse approximation of the template. In addition, drawbacks of the single-view acquisition become apparent in regions that are not observed by the scanner, such as the back of the sumo. Quantitatively, we measured the maximum of the average distance over all frames as 0.0012, the maximum of the maximum distance over all frames as 0.0283 as a fraction of the bounding box diagonal.

Limitations. We make few assumptions on the geometry and motion of the scanned objects. The correspondence estimation based on closest points, however, requires a sufficiently high acquisition frame-rate as otherwise, misalignments can occur, as shown in Figure 17 (a). Similarly, for parts of the shape that are out of view for an extended period of time, registration can fail if these regions have undergone deformations while not being observed by the scanner. In such a case, our system would require user interaction to re-initialize the registration. This is an inherent limitation of single-view systems where more than half of the object surface is occluded at any time instance. However, even some multi-view systems (e.g. [Vlasic et al. 2008]) permit user assistance to adjust incorrect optimizations. Similar manual assistance might be required for longer sequences, where the scanner infrequently produces inferior data in certain frames. These frames need to be removed

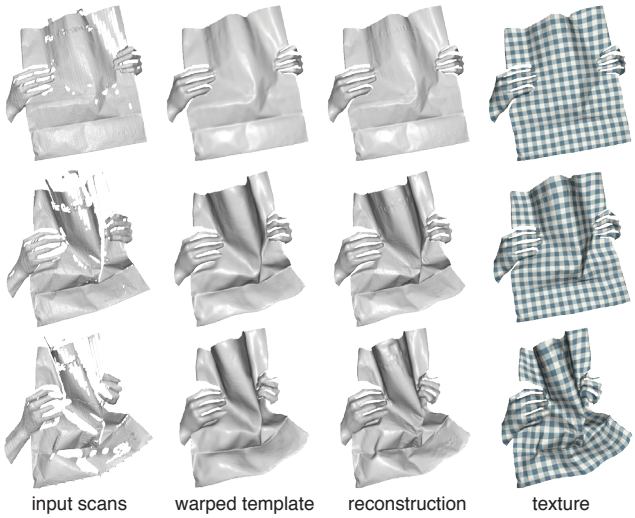

Figure 13: *Sharp creases and intricate folds created by the complex, non-smooth deformation of a crumpling paper bag are captured accurately.*

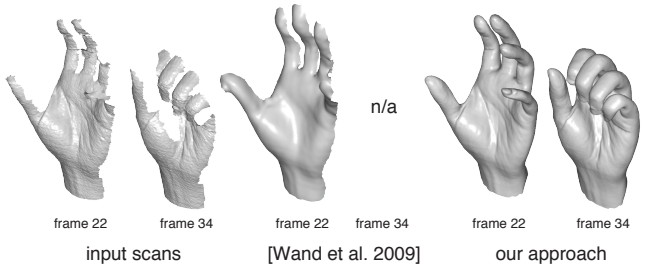

Figure 14: *Reconstruction without a template is particularly challenging for single-view acquisition. The results in the center have been produced by the authors of [Wand et al. 2008].*

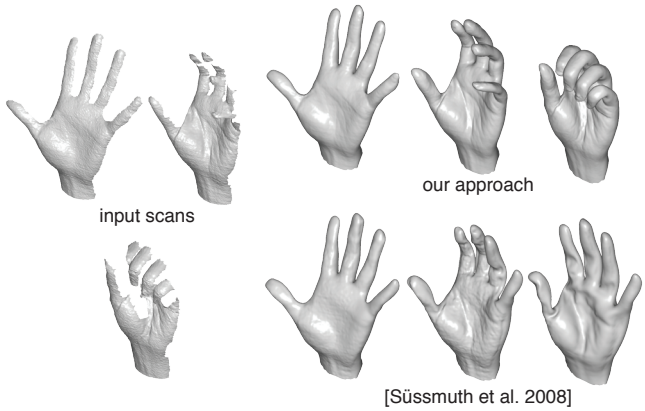

Figure 15: *Comparison of two template based reconstruction methods. The results in the bottom right have been produced by the authors of [Süssmuth et al. 2008].*

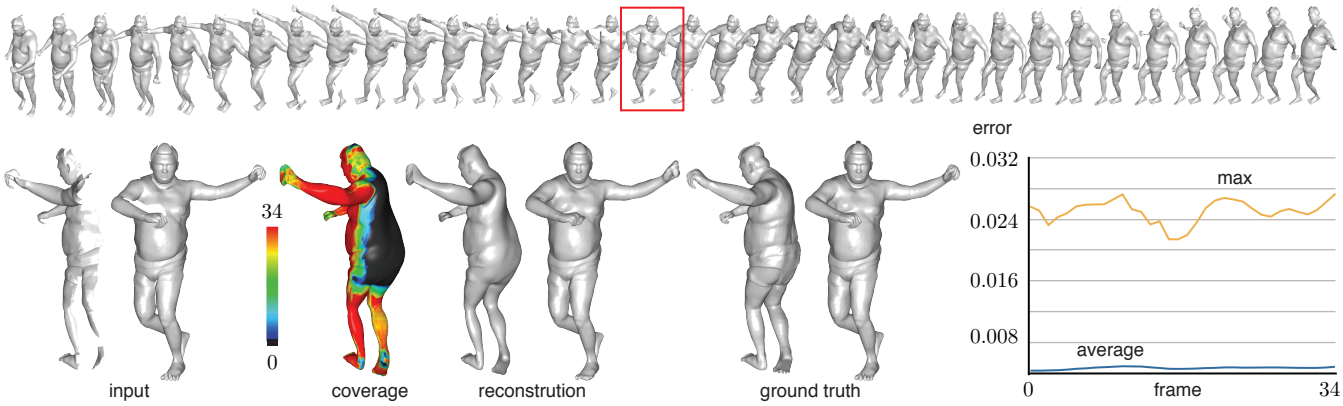

Figure 16: *Ground truth comparison for a synthetic full-body example with fast motion. The top row shows every frame of the input sequence. The color-coded image indicates the number of frames in which a certain part of the shape is covered by the scans. The graph shows the maximum and average error distance between the ground truth and the reconstruction for each frame.*

manually and the registration re-started with user assistance. While none of our sequences required such manual intervention, the acquisition of longer sequences was inhibited by this limitation of our scanning system.

Global aspects, such as the loop closure problem well-known in rigid scanning [Pulli 1999] are currently not considered in our system. To address these limitations, more sophisticated feature tracking would be required in order to establish reliable correspondences across larger spatial and temporal distances. We currently do not prevent global self-intersections of the reconstructed meshes. However, as shown in Figure 17 (b), our method robustly recovers, mainly due to the use of geodesic distances on the template mesh and the correspondence pruning strategy based on normal consistency and visibility. Avoiding self-intersections entirely would require an additional self-collision handling step in the shape deformation optimization algorithm, which would add a significant overhead to the overall reconstruction pipeline. Our method does not discover topological errors in the template, as shown in Figure 17 (c). In the template reconstruction the pinky has been erroneously connected to the paper bag, which leads to artifacts in the final frames of the sequences, where the finger is lifted off the bag.

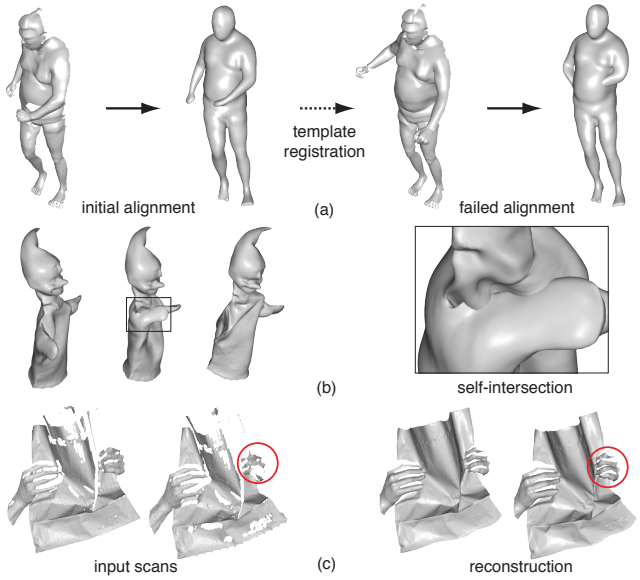

Figure 17: *Limitations: (a) registration can fail if the frame-rate is too low relative to the motion of the scanned object; (b) self-intersections are not prevented during template alignment; (c) wrong template topology leads to artifacts when the finger is lifted off the paper bag.*

8 Conclusion

We have presented a robust algorithm for geometry and motion reconstruction of dynamic shapes. One of the main benefits of our method is simplicity. Our scanning system requires no specialized hardware or complex calibration or synchronization, and can be readily deployed in different acquisition scenarios. We do not require silhouette or feature extraction, manual correction of correspondences, or the explicit construction of a shape skeleton. Our system demonstrates that even for single-view acquisition, high-quality results can be obtained for a variety of scanned objects, with a realistic reconstruction of shape dynamics and fine-scale features. Key to the success of our algorithm is the robust template tracking based on an adaptive deformation model. Our novel detail synthesis method exploits the accurate registration to aggregate and propagate geometric detail into occluded regions. As future work we plan to resolve aforementioned limitations and incorporate global self-collision handling. Moreover, we want to evaluate the algorithm in a multi-view setting where larger parts of the object are seen at the same or alternating time instances. As our current acquisition system only allows us to scan within a working volume of $40 \times 30 \times 60$ cm^3, we wish to extend our scanning setup to allow acquisition of larger objects such as full human body performances. The tests on synthetic data indicate that our reconstruction algorithm should perform well for such cases. Finally, the proposed registration algorithm can be used to acquire and learn material behavior (such as the crumpling of paper or folding of skin). Such information can be used to improve the realism of physically-based simulation algorithms.

Acknowledgements. The authors would like to thank Thibaut Weise for providing the real-time 3D scanner, Carsten Stoll for his performance capture data, Sang Il Park and Jessica Hodgins for the animated sumo. Special thanks go to Johannes Schmid for helping with the video editing, Michael Wand, Martin Bokeloh, and Jochen Süssmuth for performing the comparisons, Qi-Xing Huang and Maks Ovsjanikov for the feedbacks and discussions. This work is supported by SNF grant 200021-112122, NSF grants ITR 0205671, FRG 0354543, FODAVA 808515, as well as NIH grant GM-072970 and the Fund for Scientific Research, Flanders (F.W.O.-Vlaanderen).

References

AHMED, N., THEOBALT, C., DOBREV, P., SEIDEL, H.-P., AND THRUN, S. 2008. Robust fusion of dynamic shape and normal capture for high-quality reconstruction of time-varying geometry. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2008)*, 1–8.

ALLEN, B., CURLESS, B., AND POPOVIĆ, Z. 2003. The space of human body shapes: reconstruction and parameterization from range scans. *ACM Transactions on Graphics 22*, 3, 587–594.

AMBERG, B., ROMDHANI, S., AND VETTER, T. 2007. Optimal step nonrigid icp algorithms for surface registration. In *Proceedings of IEEE CVPR*.

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

AURICH, V., AND WEULE, J. 1995. Non-linear gaussian filters performing edge preserving diffusion. In *Mustererkennung 1995, 17. DAGM-Symposium*, Springer-Verlag, 538–545.

BLANZ, V., AND VETTER, T. 1999. A morphable model for the synthesis of 3D faces. In *Proceedings of ACM SIGGRAPH 99*, ACM Press / ACM SIGGRAPH, 187–194.

BOTSCH, M., AND SORKINE, O. 2008. On linear variational surface deformation methods. *IEEE Transactions on Visualization and Computer Graphics 14*, 1, 213–230.

BRADLEY, D., POPA, T., SHEFFER, A., HEIDRICH, W., AND BOUBEKEUR, T. 2008. Markerless garment capture. *ACM Transactions on Graphics 27*, 3, 99:1–99:9.

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

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

BROWN, B. J., AND RUSINKIEWICZ, S. 2007. Global non-rigid alignment of 3-d scans. *ACM Transactions on Graphics 26*, 3, 21:1–21:10.

CHANG, W., AND ZWICKER, M. 2008. Automatic registration for articulated shapes. *Computer Graphics Forum (Proc. SGP) 27*, 5, 1459–1468.

CHANG, W., AND ZWICKER, M. 2009. Range scan registration using reduced deformable models. *Computer Graphics Forum (Proceedings of Eurographics 2009), to appear*.

DE AGUIAR, E., STOLL, C., THEOBALT, C., AHMED, N., SEIDEL, H.-P., AND THRUN, S. 2008. Performance capture from sparse multi-view video. *ACM Transactions on Graphics 27*, 3, 98:1–98:10.

GUENNEBAUD, G., AND GROSS, M. 2007. Algebraic point set surfaces. In *ACM Transactions on Graphics*, ACM, New York, NY, USA, vol. 26, 23:1–23:10.

HUANG, Q., ADAMS, B., WICKE, M., , AND GUIBAS, L. J. 2008. Non-rigid registration under isometric deformations. *Computer Graphics Forum (Proc. of SGP) 27*, 5, 1459–1468.

IKEMOTO, L., GELFAND, N., AND LEVOY, M. 2003. A hierarchical method for aligning warped meshes. In *Proceedings of 4th Int. Conference on 3D Digital Imaging and Modeling*, 434–441.

KIMMEL, R., AND SETHIAN, J. A. 1998. Computing geodesic paths on manifolds. In *Proc. Natl. Acad. Sci. USA*, 8431–8435.

LI, H., SUMNER, R. W., AND PAULY, M. 2008. Global correspondence optimization for non-rigid registration of depth scans. *Computer Graphics Forum (Proc. SGP) 27*, 5, 1421–1430.

MITRA, N. J., FLORY, S., OVSJANIKOV, M., GELFAND, N., GUIBAS, L., AND POTTMANN, H. 2007. Dynamic geometry registration. In *Symposium on Geometry Processing*, 173–182.

PARK, S. I., AND HODGINS, J. K. 2006. Capturing and animating skin deformation in human motion. *ACM Transactions on Graphics 25*, 3, 881–889.

PARK, S. I., AND HODGINS, J. K. 2008. Data-driven modeling of skin and muscle deformation. *ACM Transactions on Graphics 27*, 3, 96:1–96:6.

PAULY, M., MITRA, N. J., GIESEN, J., GROSS, M., AND GUIBAS, L. J. 2005. Example-based 3d scan completion. In *Symposium on Geometry Processing*.

PULLI, K. 1999. Multiview registration for large data sets. In *Second Int. Conf. on 3D Dig. Image and Modeling*, 160–168.

ROBERTS, S. 1959. Control chart tests based on geometric moving averages. *Technometrics1*, 239–250.

RUSINKIEWICZ, S., HALL-HOLT, O., AND LEVOY, M. 2002. Real-time 3D model acquisition. *ACM Transactions on Graphics 21*, 3, 438–446.

SHARF, A., ALCANTARA, D. A., LEWINER, T., GREIF, C., SHEFFER, A., AMENTA, N., AND COHEN-OR, D. 2008. Space-time surface reconstruction using incompressible flow. *ACM Transactions on Graphics 27*, 5, 110:1–110:10.

SUMNER, R. W., SCHMID, J., AND PAULY, M. 2007. Embedded deformation for shape manipulation. *ACM Transactions on Graphics 26*, 3, 80:1–80:7.

SÜSSMUTH, J., WINTER, M., AND GREINER, G. 2008. Reconstructing animated meshes from time-varying point clouds. *Computer Graphics Forum (Proceedings of SGP 2008) 27*, 5, 1469–1476.

VLASIC, D., BARAN, I., MATUSIK, W., AND POPOVIĆ, J. 2008. Articulated mesh animation from multi-view silhouettes. *ACM Transactions on Graphics 27*, 3, 97:1–97:9.

WAND, M., JENKE, P., HUANG, Q., BOKELOH, M., GUIBAS, L., AND SCHILLING, A. 2007. Reconstruction of deforming geometry from time-varying point clouds. In *Symposium on Geometry processing*, 49–58.

WAND, M., ADAMS, B., OVSJANIKOV, M., BERNER, A., BOKELOH, M., JENKE, P., GUIBAS, L., SEIDEL, H.-P., AND SCHILLING, A. 2009. Efficient reconstruction of non-rigid shape and motion from real-time 3d scanner data. *ACM Transactions on Graphics*. (to appear).

WEISE, T., LEIBE, B., AND GOOL, L. V. 2007. Fast 3d scanning with automatic motion compensation. In *IEEE Conference on Computer Vision and Pattern Recognition*, 1–8.

ZHANG, L., AND SEITZ, S. M. 2000. Image-based multiresolution shape recovery by surface deformation. SPIE, S. F. El-Hakim and A. Gruen, Eds., vol. 4309, 51–61.

ZHANG, L., SNAVELY, N., CURLESS, B., AND SEITZ, S. M. 2004. Spacetime faces: high resolution capture for modeling and animation. *ACM Transactions on Graphics 23*, 3, 548–558.