Temporally Coherent Completion of Dynamic Shapes

HAO LI
Columbia University, EPFL, and ETH Zurich
LINJIE LUO
Princeton University
DANIEL VLASIC
Massachusetts Institute of Technology
PIETER PEERS
The College of William & Mary and Institute for Creative Technologies/University of Southern California
JOVAN POPOVIĆ
Adobe Systems Inc., University of Washington, and Massachusetts Institute of Technology
MARK PAULY
EPFL
and
SZYMON RUSINKIEWICZ
Princeton University

We present a novel shape completion technique for creating temporally coherent watertight surfaces from real-time captured dynamic performances. Because of occlusions and low surface albedo, scanned mesh sequences typically exhibit large holes that persist over extended periods of time. Most conventional dynamic shape reconstruction techniques rely on template models or assume slow deformations in the input data. Our framework sidesteps these requirements and directly initializes shape completion with topology derived from the visual hull. To seal the holes with patches that are consistent with the subject's motion, we first minimize surface bending energies in each frame to ensure smooth transitions across hole boundaries. Temporally coherent dynamics of surface patches are obtained by unwarping all frames within a time window using accurate interframe correspondences. Aggregated surface samples are then filtered with a temporal visibility kernel that maximizes the use of nonoccluded surfaces. A key benefit of our shape completion strategy is that it does not rely on long-range correspondences or a template model. Consequently, our method does not suffer error accumulation typically introduced by noise, large deformations, and drastic topological changes. We illustrate the effectiveness of our method on several high-resolution scans of human performances captured with a state-of-the-art multiview 3D acquisition system.

Categories and Subject Descriptors: I.3.7 [**Computer Graphics**]: Three-Dimensional Graphics and Realism—*Animation*; I.4.8 [**Image Processing and Computer Vision**]: Scene Analysis—*Surface Fitting*

General Terms: Algorithms

Additional Key Words and Phrases: Animation reconstruction, shape completion, nonrigid registration, dynamic surface reconstruction, dynamic-shape acquisition, visual hull, 3D video, temporal coherence

ACM Reference Format:

Li, H., Luo, L., Vlasic, D., Peers, P., Popović, J., Pauly, M., and Rusinkiewicz, S. 2012. Temporally coherent completion of dynamic shapes. ACM Trans. Graph. 31, 1, Article 2 (January 2012), 11 pages.
DOI = 10.1145/2077341.2077343
http://doi.acm.org/10.1145/2077341.2077343

This project was supported by the SNF fellowship for prospective researchers, SNF grant 20001-112122, NSF grant ISS-1016703, the University of Southern California Office of the Provost, and the U.S. Army Research, Development, and Engineering Command (RDECOM). The content of the information does not necessarily reflect the position or the policy of the U.S. Government, and no official endorsement should be inferred.

Authors' addresses: H. Li (corresponding author), Columbia University, EPFL, and ETH Zurich; email: hao@hao-li.com; L. Luo, Princeton University; D. Vlasic, Massachusetts Institute of Technology; P. Peers, The College of William and Mary and Institute for Creative Technologies/University of Southern California; J. Popović, Adobe Systems Inc., University of Washington, and Massachusetts Institute of Technology; M. Pauly, EPFL; S. Rusinkiewicz, Princeton University.

DOI 10.1145/2077341.2077343
http://doi.acm.org/10.1145/2077341.2077343

1. INTRODUCTION

Advances in real-time 3D capture have enabled the acquisition of dynamically deforming shapes at sustained "video" rates, either from a single viewpoint (e.g., Zhang et al. [2003], Davis et al. [2005], and Weise et al. [2007]) or, more recently, from multiple viewpoints (e.g., Vlasic et al. [2009]). Real-time captured data plays

an increasing role in fields ranging from film-making and gaming to engineering and medicine. For example, in visual effects, the use of densely acquired geometry is often preferred over conventional marker-based motion capture systems for digitizing highly compelling human facial animations. In oncology, when cancer patients undergo radiation therapies, the locations of preidentified malignant tumors need to be constantly updated using surface capture to ensure accurate treatment. Having a complete and accurate digital 3D representation of deforming objects is therefore vital for these applications.

Although resolution and accuracy are constantly improving with each new generation of image sensors and scanning techniques, acquisition systems are generally unable to capture the full surface at once. Even though multiple sensors can be placed around the subject, most scanned shapes are likely to exhibit large holes due to occlusions and low surface albedo. We therefore argue that increasing 3D scan coverage is on a fundamentally different "technology curve," and is unlikely to be solved by improvements in scanning technology. While many shape completion techniques for static scans have been developed, surface patches that fill holes in dynamic data must deform coherently and according to the subject's general shape.

We consider the problem of obtaining temporally coherent *watertight* 3D meshes from high-resolution scan sequences of dynamic performances recorded from multiple viewpoints. We assume that the input scans have reasonable coverage and that most noise and outliers are suppressed, either by using an improved scanning technology or by effectively postprocessing the data.

In human performance capture, large holes are typically observed between legs, regions occluded by arms, and those parts exhibiting significant grazing angles to the cameras. While a deforming shape can expose newly observed regions over time, these holes are usually so large that full coverage is only possible after extended recording. Most current techniques for temporally consistent shape completion assume that the dynamic subject is represented by a single deformable surface (*template*). The template model is usually obtained by a separate rigid reconstruction step (e.g., Li et al. [2008], de Aguiar et al. [2008], and Vlasic et al. [2008]) or by globally aggregating all surface samples through time (e.g., Wand et al. [2009], Mitra et al. [2007], and Süßmuth et al. [2008]). Both approaches rely on establishing full interframe correspondences of surface points across entire recordings for the template. However, we do not wish to restrict the degree of deformation or fix the topology. Deformations that involve topology changes or interactions between multiple disconnected components cannot be accurately modeled with a single template (e.g., gliding cloth, exposing new body parts, etc.). Thus the correct shape is unlikely to be recovered by simply propagating geometry across long sequences without knowledge of full interframe correspondences in occluded regions. Moreover, error accumulation is likely to occur when correspondences need to be repeatedly determined between pairs of input scans. Consequently, none of these techniques can guarantee drift-free reconstruction for complex deformations and largely incomplete input data.

Our proposed method does not require globally consistent correspondences or a template model. The key insight is that only accurate pairwise correspondences are needed for temporally consistent shape completion, as the relevance of surface information decreases with time. For example, a fold on a dress observed in one frame is likely to disappear or completely change its shape at a later time. To establish dense pairwise correspondences, we employ a novel two-stage registration algorithm that: (1) performs a coarse nonrigid registration algorithm of Li et al. [2009] equipped with deformation graph prediction and sparse texture-based constraints for higher accuracy and robustness, and (2) refines this coarse correspondence computation using an improved version of a fine-scale alignment algorithm [Brown and Rusinkiewicz 2007]. Because surface correspondences only reside within a subset of two consecutive pairs of incomplete scans, more coverage leads to improved alignment quality. We maximize coverage by accumulating newly observed surfaces using an interleaved registration/merging method in a forward-and-backward fashion.

Given the original scanned surfaces and their pairwise correspondences, our shape completion approach starts by filling the holes in each frame independently using the visual hull as a topological prior. We further optimize vertex positions to satisfy spatial smoothness across hole boundaries [Liepa 2003]. The use of the visual hull as a topological prior helps to resolve ambiguous hole-filling strategies (e.g., when the arm is close to the body). To minimize temporal flicker, we unwarp all watertight shapes within a time window into the current frame using the precomputed dense pairwise correspondences. The aggregation of nearby frames forms a temporally coherent shape which we reconstruct by weighted integration of surface samples [Kazhdan et al. 2006]. We design our weighting scheme to act similarly to a temporal bilateral filter, but instead of preserving motion discontinuities, we maximize the aggregation of nonoccluded regions. However, fine-scale geometrical details tend to be blurred out by the integration of the unwarped shapes. To resynthesize these fine-scale details, high-frequency details from partial input scans or user-provided normal maps are reapplied to the integrated surface using the method of Nehab et al. [2005].

Our framework is designed to handle input data with large occlusions, topological changes, and complex deformations. Because an interleaved registration/merging scheme is employed, only a few adjacent meshes are needed simultaneously, leading to modest memory requirements. This also makes our method well suited for very high-resolution input data. We illustrate our method on the meshes of Vlasic et al. [2009] and compare our results with recent work on space-time reconstruction. While the absence of globally corresponded meshes precludes certain applications, our method is the first to enable free-viewpoint video of watertight and temporally-coherent high-resolution dynamic geometries.

2. PREVIOUS WORK

A large body of prior work has investigated hole filling for static geometries. Various strategies exist that either operate explicitly on polygonal meshes [Held 1998; Liepa 2003] or implicitly via a volumetric representation [Curless and Levoy 1996; Carr et al. 1997; Davis et al. 2002; Kazhdan et al. 2006]. Regardless of the heuristic used to fill in the missing geometry, the end result is a hole-free surface. We refer to Ju [2009] for an in-depth discussion of various hole-filling strategies and heuristics. While these methods generate excellent hole-free static surfaces, directly applying them to every frame in a dynamic performance separately can result in incorrect topology and temporally incoherent surfaces.

A common approach to producing temporally consistent watertight dynamic geometry is to employ a template prior. Early methods [Carranza et al. 2003; Starck and Hilton 2003; Zhang et al. 2004; Corazza et al. 2006; Sand et al. 2003; Anguelov et al. 2005; Allen et al. 2002] deform a generic or user-specified template geometry to match a dynamic performance. While the general animation can be captured, geometric details are limited to those in the

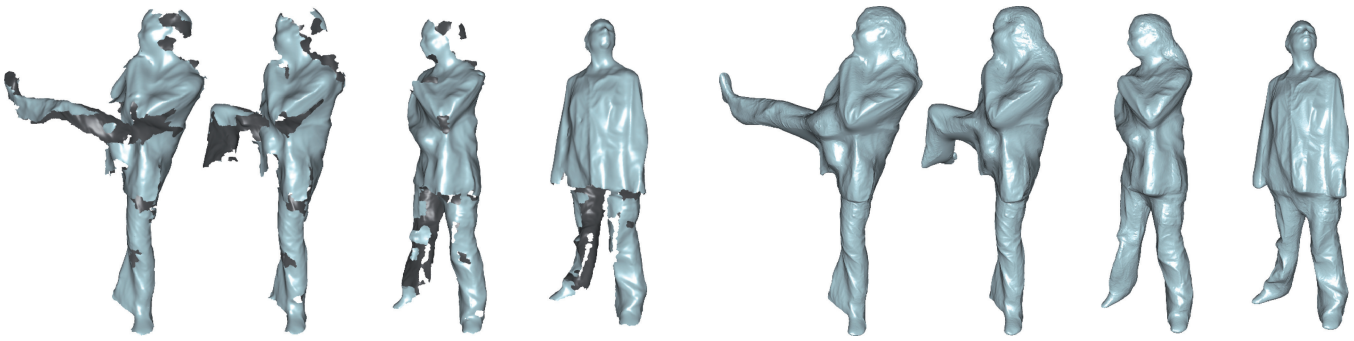


Fig. 1. Left: Real-time 3D acquired dynamic performance geometry typically exhibits holes that are often temporally persistent. Right: Hole-filled, temporally coherent, and detailed sequence of watertight surfaces reconstructed using our method.

template. Recently, detailed person-specific static 3D scans have been directly considered as a template geometry [de Aguiar et al. 2007, 2008; Theobalt et al. 2007; Bradley et al. 2008; Vlasic et al. 2008], resulting in richer details than the input data sequence. While convincing results can be produced, they fail in modeling fine-scale dynamics as the details are baked in the template. Biresolution approaches presented by Ahmed et al. [2008] and Li et al. [2009] deform a smooth template to match large-scale motion to scanned geometry of a dynamic performance. Small-scale details, such as wrinkles and folds, are synthesized on top of the deformed template to provide a match at higher resolution. General problems with template-based methods are that they cannot deal with changing topology and that appropriate template geometry must be available (either from a database or through a surface reconstruction process).

Pekelny and Gotsman [2008] assume that the dynamic performance consists of articulations of rigid parts. Starting from a manual segmentation, an optimal rigid motion is computed for each part. Finally, information is accumulated (forward in time) for each rigid part to fill holes and improve the quality of the reconstructed surface. Chang and Zwicker [2009] propose a method that does not require any manual segmentation or template. However, their method is limited to subjects that exhibit articulated motion. Zheng et al. [2010] automatically extract a consensus skeleton to derive a consistent temporal topology. However, it assumes that the underlying shape is clearly articulated, which is not always the case for subjects wearing loose clothing.

Wand et al. [2009] globally solve for an optimal deforming template and minimize the effects of drift by employing a hierarchical scheme to register pairs of surfaces. Given the computed correspondences, they accumulate the frames to form a single representative shape. As demonstrated by Li et al. [2009], their method fails with drastic topology changes.

Explicitly computing correspondences over long sequences is an error-prone process. To avoid these issues Mitra et al. [2007] cast the problem of computing hole-free surfaces from unregistered dynamic performance geometry as a spatio-temporal 4D interpolation problem. Similarly, the method proposed by Wand et al. [2007] uses a statistical framework to solve for the dynamic shape under an as-rigid-as-possible motion and impose temporal smoothness. Süssmuth et al. [2008] improve on robustness by first fitting an implicit 4D surface before optimizing motion. However, extracting a dynamic manifold in the space-time domain imposes smooth deformation of the subject and fails for large deformations between adjacent frames.

Sharf et al. [2008] propose to reconstruct surfaces by relaxing the as-rigid-as-possible motion by a less restrictive volume preservation condition to better deal with noisy data. However, this introduces noticeable flickering in the reconstructions. Moreover, the deformation of most real-world objects does not exactly preserve volume (e.g., loose clothing).

Our method differs in that we only rely on correspondences within a small time window and exploit topology information obtained from the visual hull. A hole-free spatio-temporally coherent surface is obtained by temporally filtering the unwarped frames originating within a small time window. We also improve tracking with an interleaved registration/merge scheme and process the data in a forward-and-backward fashion to reliably obtain pairwise correspondences.

3. TEMPORALLY COHERENT HOLE FILLING

The proposed shape completion method employs a three-step algorithm to synthesize temporally coherent watertight surfaces from scanned sequences of nonrigidly deforming shapes.

(1) We start by filling the holes in each frame separately, employing the visual hull as a topological prior. Furthermore, to promote temporal smoothness and avoid unnatural discontinuities across hole boundaries, we optimize the hole-filled vertex positions by minimizing a bending energy fairness functional.

(2) We proceed with a weighted surface integration scheme that reconstructs a temporally coherent watertight surface from adjacent time frames, thus minimizing temporal artifacts. We warp the resulting shapes using pairwise correspondences computed in a preprocessing step (detailed in Section 4).

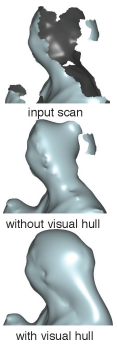
(3) Finally, we resynthesize the details lost during warping and integration onto the final temporally coherent watertight mesh.

The complete three-step process is schematically depicted in Figure 2.

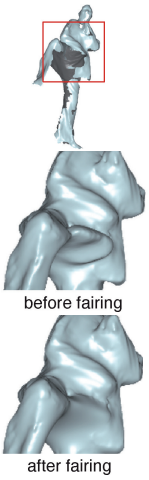
3.1 Single Frame Hole Filling

As illustrated in Figure 1, scanning human performances typically results in large holes which persist in close proximity over many frames. Filling these holes can become ambiguous when two separate incomplete surfaces get close.

Visual Hull Prior. As suggested in Vlasic et al. [2009], the visual hull provides a robust estimate for obtaining watertight shapes. We therefore initialize our hole filling by combining the vertices of the original partial scans with those of the visual hull. We set a weight $w = 1$ for each surface sample located on the scans and $w = \epsilon$ for visual hull samples. A hole-free mesh is then obtained by Poisson surface reconstruction [Kazhdan et al. 2006] using the weighted oriented surface samples. As each frame is being completed independently, considerable flickering artifacts are likely to occur in hole-filled regions for dynamic input geometry.



Surface Fairing. To enforce smooth transitions with the surroundings of a hole-filled mesh region, we solve for new vertex positions by minimizing a fairness functional constrained by the hole boundaries, similar to Liepa [2003]. In particular, we minimize the linearized bending energy of the patched mesh's nonboundary vertices using the standard cotangent bi-Laplacian [Botsch and Sorkine 2008]. Since only limited views are provided for computing the visual hull, optimizing surface fairness in hole regions yields spatially smooth and more plausible reconstructions for curved surfaces such as folds in a garment. While spatial smoothness for hole regions can be directly obtained by carefully estimating sample weights at hole boundaries during Poisson reconstruction, this extra fairing step avoids the need for additional feathering parameters. While the fairing significantly reduces strong discontinuities across boundaries, flickering still persists as each frame is processed independently.



3.2 Temporal Filtering

Temporal flicker is present both in the original data (due to independent per-frame reconstruction) and our hole-filled surfaces (due to visual-hull-based optimization). We address this with a temporal filter that combines each frame with its neighbors, and only requires knowledge of pairwise correspondences between neighboring frames in the original sequence. The correspondences are computed in a preprocessing step (detailed in Section 4).

Our temporal filtering process starts with the incomplete reconstructed mesh (original data) and the hole-filled regions at each frame. We warp the hole-filled regions into the neighboring frames using a mesh deformation based on the pairwise correspondences and Laplacian coordinates [Alexa 2003], where the reconstructed meshes define the constraints. At this point, we have the reconstructed meshes from the current and the neighboring frames, as well as the hole-filled regions from those three frames, all aligned to a common pose. We combine them all using Poisson surface reconstruction [Kazhdan et al. 2006] with the following weights: 100 for the reconstructed mesh of the current frame, 10 for the reconstructed mesh of the neighboring frames (deformed to the current frame), 2 for the hole-filled regions of the current frame, and 1 for the hole-filled regions of the neighboring frames (also deformed to the current frame). This imposes a mild temporal filter on the reconstructed surfaces, and a strong filter on the hole-filled regions.

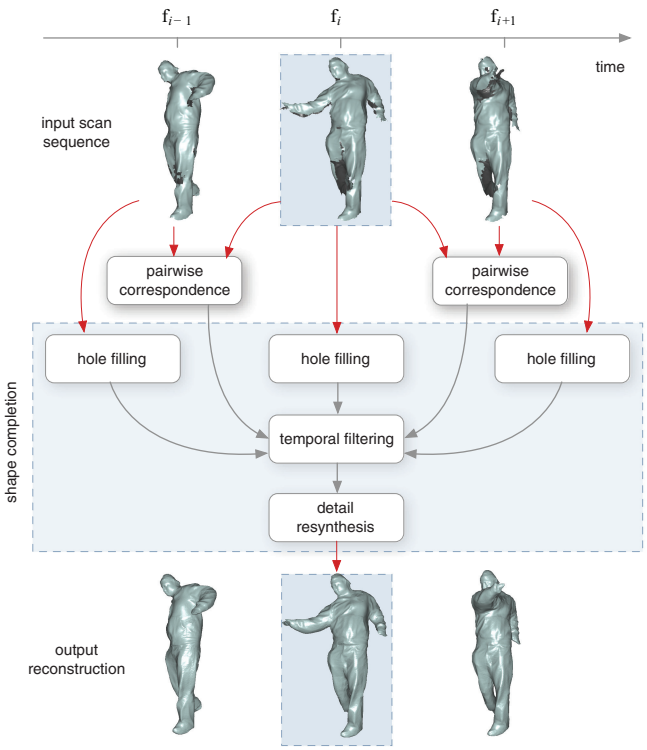


Fig. 2. Overview of our system.

This step reduces the temporal flicker, and propagates some of the reconstructed surface detail from the neighboring frames onto the current frame (this stems from the neighboring reconstructed mesh weight being larger than any hole-filled region weight).

3.3 Detail Resynthesis

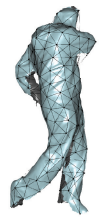
While the weighted temporal filtering approach reduces flicker between the hole-filled meshes, it also tends to remove some fine geometric details mainly due to the Poisson surface reconstruction step. Since our input data is only affected by very little noise, the stability of the high-frequency details in nonboundary regions allows us to reintroduce details and compensate for this loss. We employ the method of Nehab et al. [2005] to resynthesize high-frequency detail, which can either come directly from the original input scans, or alternatively from measured normal maps. In our case, stable normal information is available in the form of normal maps [Vlasic et al. 2009].

4. PAIRWISE CORRESPONDENCES

A crucial component in the proposed shape completion algorithm are the accurate pairwise correspondences between consecutive frames of a dynamic performance. Several short- and long-range correspondence algorithms exist (e.g., Zheng et al. [2010], Wand et al. [2009], Li et al. [2009], and Sharf et al. [2008]). However, we found that none of these methods gave the necessary accuracy to obtain high-quality shape completions (see Section 5 for a qualitative comparison). In this work, we develop a novel two-scale approach. We start by computing coarse correspondences that are globally coherent and capture large-scale deformations (Section 4.1). Next,

we refine these coarse correspondences to accurately align the fine-scale geometric details (Section 4.2).

4.1 Registration Based on Deformation Graphs


deformation graph

To compute the pairwise coarse-scale registration, a state-of-the-art nonrigid ICP algorithm [Li et al. 2009] is extended with: (1) a prediction-based initialization and (2) sparse positional constraints computed from input video data. The proposed improvements increase robustness to large deformations and minimize tangential drift, improving accuracy over short time windows (as validated in Section 5).

The underlying subspace deformation technique uses a graph with nodes that are uniformly sampled on the scan surface to warp the scan mesh vertices via linear blend skinning. Each vertex has a weight inversely proportional to its Euclidean distance to the $k = 4$ closest nodes. The optimization solves for the affine transformations on the graph nodes and is regularized with an energy term E_{rigid} that maximizes rigidity. Another energy term, E_{smooth}, ensures consistency between node transformations that are connected with an edge. Nonrigid ICP iteratively computes the combined closest point and minimizes point-to-point and point-to-plane distance E_{fit} in the optimization. In addition, we add an energy term E_{tex} for sparse 3D positional constraints obtained from sparse texture correspondences. At each deformation step we solve a nonlinear optimization with the objective function

$$E_{\text{tot}} = \alpha_{\text{fit}} E_{\text{fit}} + \alpha_{\text{tex}} E_{\text{tex}} + \alpha_{\text{rigid}} E_{\text{rigid}} + \alpha_{\text{smooth}} E_{\text{smooth}}, \qquad (1)$$

where $\alpha_{\text{fit}} = 1$ and $\alpha_{\text{tex}} = 100$. Similarly to Li et al. [2009], robustness is assured against suboptimal local minima by starting the registration with a high regularization ($\alpha_{\text{rigid}} = 100$ and $\alpha_{\text{smooth}} = 10$) and successively halving the weights whenever the deformation step converges. We stop the optimization when $\alpha_{\text{smooth}} = 0.01$.

Graph Prediction. While effective for a large range of deformations, the preceding registration technique is likely to converge to an incorrect local minimum when there is significant motion between consecutive frames (e.g., a fast kick) or in regions with few geometric features. Convergence to the *correct* deformation can be promoted by employing a prediction that provides an initial deformation close to the desired deformation. The deformation graph in frame $f + 2$ is predicted by linear extrapolation from frames f and $f + 1$. In short, for each edge of the deformation graph, we extract the smallest $3D$ transformation that deforms that edge from frame f to $f + 1$. We then transform each vertex of the deformation graph in frame $f + 1$ by the average of all the transformations corresponding to its incident edges.

Sparse Texture-Based Constraints. So far, E_{fit} is used to bring the source scan closer to the target. However, this does not preclude tangential drifts (even with the preceding prediction). For regions with very little detail, using only geometric constraints can yield suboptimal alignment (e.g., sliding versus stretching). Thus, we add texture constraints (obtained from image recordings that are projected onto the mesh) and use them as sparse positional constraints for the optimization.

To determine these sparse features we compute 2D feature descriptors from the video recordings of 8 different camera positions between consecutive frames. In our implementation we used SURF feature descriptors [Bay et al. 2008], though many other 2D descriptors can be employed. In the case of SURF, features tend to be concentrated at the silhouette of the subject, and do not represent true surface features. Therefore, only those features that lie away from some preset distance (8 pixels) of the silhouette are considered.

Next, we match each detected feature point to the best corresponding feature point in the subsequent frame. To speed-up detection and minimize false positive matches, we restrict the search space by employing an optical flow-based prediction [Brox et al. 2004] and search for the best matching SURF descriptor in a small neighborhood around this predicted feature point location. We discard the pairwise match if the error on the feature descriptors exceeds a certain threshold. We search in a radius of $\min\{10, d\}$ around the predicted point, with d being the distance of the predicted displacement. We reject matches with a descriptor error above 0.2. To improve robustness, we only consider correspondences that can be reliably tracked for at least 3 consecutive frames.

Finally, we project every tracked 2D feature back on the original geometries to obtain 3D positional constraints. Section 5 validates that the found texture-based correspondences (up to 1000 per frame) greatly improve the registration quality.

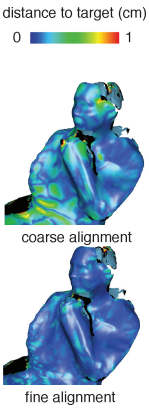
4.2 Fine-Scale Alignment

After the coarse nonrigid alignment, we perform fine-scale registration using a nonrigid locally weighted ICP algorithm based on Brown and Rusinkiewicz [2007]. This improves the alignment of small geometric details. Our algorithm improves on Brown and Rusinkiewicz [2007] by taking the following two observations into account.

(1) The main goal is to locally improve the alignment, hence the weight distribution function should have local support. Otherwise, far-away points can bias the local alignment. We use a Compactly Supported Radial Basis Function (CSRBF).
(2) Gelfand et al. [2003] showed that the stability of the ICP matching algorithm depends on the local geometry. If the matched geometry does not contain enough surface detail, drift may occur. Ideally, the size of the matched geometry should adapt to the local feature size.

These observations define the following three-step algorithm.

Sampling. We start by sampling an optimal set of feature points on the deformed mesh according to the alignment error which is defined by the distance between source mesh and nearest point on the target mesh after nonrigid ICP. This step ensures that regions with median alignment error gain average sampling weights, while the influence of large outliers is decreased.



Matching. Next, we find correspondences using a local ICP algorithm based on Brown and Rusinkiewicz [2007]. However, we differ in that we employ a CSRBF for point selection near a feature point and iteratively select the best radius of CSRBF according to the local geometric stability. Specifically, we use a quadratically decreasing CSRBF $f(x) = \max\{1-(x/r)^2, 0\}$, where r is an adaptively selected support radius. To optimally select the support radius, we iteratively apply ICP, reducing the radius at each step as long as the alignment error decreases and the stability of the sampled points is above 0.02, a threshold that empirically prevents drifting. The iterative scheme proves robust since the relatively large initial CSRBF radii avoid suboptimal local matching. We further

improve robustness by rejecting correspondences whose nearest vertices are on the mesh boundaries.

Warping. We employ the RBF deformation model proposed by Kojekine et al. [2002] to avoid known numerical instabilities of thin-plate splines as described in Sibson and Stone [1991]. The resulting linear system is sparse, due to the local support of the CSRBF, and can thus be solved efficiently.

4.3 Shape Accumulation

The preceding two-scale registration algorithm is capable of producing accurate correspondences between mutually visible surface regions since the target geometry is fully defined (no partial data). However, when correspondences map to a hole in the target scan, their positions are solely determined by the deformation model used during pairwise registration. Due to discontinuous changes between observed and occluded regions, correspondences in those areas are less reliable and may result in flickering surfaces. To improve accuracy of correspondences in hole regions we maximize surface coverage by merging previously observed surfaces to new frames.

We propose an interleaved registration/merging shape accumulation approach. Pairwise correspondences between consecutive frames $\mathbf{f}'_i$ (merged) and $\mathbf{f}_{i+1}$ (original) are used to warp $\mathbf{f}'_i$ and merge it with $\mathbf{f}_{i+1}$, yielding an accumulated shape $\mathbf{f}'_{i+1}$. We repeat this process for every frame starting from the first frame going to the last frame and perform a second pass backwards in time, from the last frame to the first.

As the scanned subject moves and deforms over time, newly visible surface regions are being exposed at each frame while certain other regions become occluded. To maximize coverage, we accumulate the deformed incomplete mesh $\mathbf{f}'_i$ and its target $\mathbf{f}_{i+1}$ after each pairwise alignment. As we allow our subject to change topology, tracking the entire recording with a single consistent mesh, as with template-based approaches, is not possible. Merging the deformed mesh $\mathbf{f}'_i$ with its target $\mathbf{f}_{i+1}$ would not only improve computational efficiency (since the vertices will not be duplicated), but it would also allow source sample positions of the correspondences to adapt to the topology of the current frame.

We employ a mesh deformation based on Laplacian coordinates [Alexa 2003] to warp frame $\mathbf{f}'_i$ to frame $\mathbf{f}_{i+1}$ and use the correspondences computed in the pairwise registration step as soft point constraints. The warped scans are then merged by accumulating vertices of both meshes, followed by the Poisson surface reconstruction method of Kazhdan et al. [2006] with equally weighted surface samples. Note that holes from $\mathbf{f}_{i+1}$ are reintroduced in the watertight Poisson reconstruction. Because of incomplete shapes, finding correspondences in unobserved surface regions for extended periods can result in accumulation of errors. As a result, the geometry of these areas can deteriorate over time and nearby surfaces can erroneously merge into a single surface. We therefore perform visual hull-based pruning by disregarding vertices that fall outside the visual hull. Furthermore, we only use the accumulated surfaces for correspondence computations, and do not use them for hole filling due to possible error accumulation.

5. RESULTS

We demonstrate our method on three of the sequences (Saskia, Abhijeet, and Jay) made publicly available by Vlasic et al. [2009]. Those high-resolution scans were captured from 8 cameras placed around a human body and cover, on average, approximately 75% of the entire surface. For efficiency, we operate on down-sampled meshes, and up-sample when resynthesizing the detail. The statistics of our input and output data are as follows (we measure size of holes as the ratio between hole area over the area of the completed mesh).

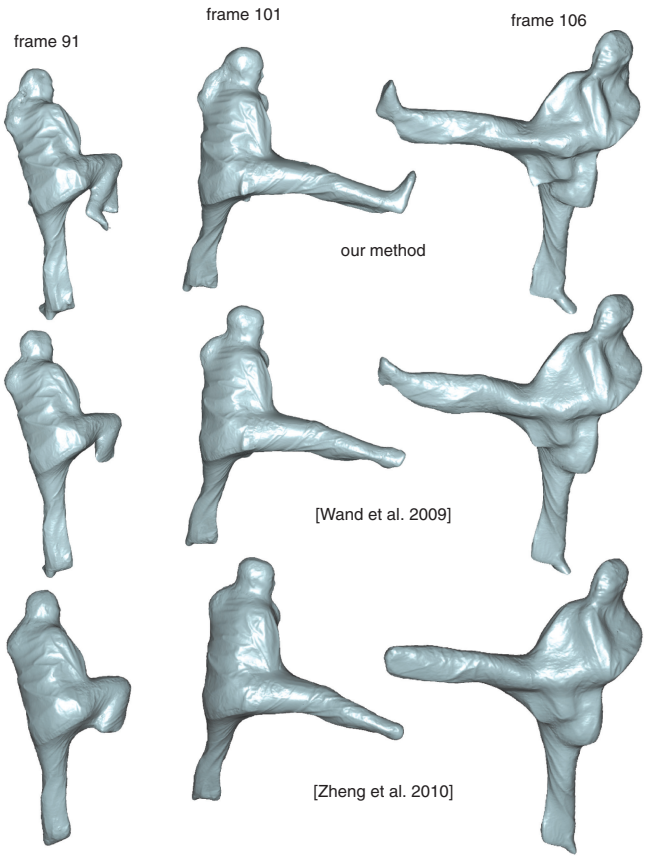


Fig. 3. Although two recent methods (second and third row, with our resynthesized detail for fair comparison) produce single topologies over the complete motion, our method (first row) is able to recover more faithful per-frame surfaces.

dataset	#frames	#input vert	#output vert	size of holes
Saskia	113	132k∼140k	353k∼380k	25%∼27%
Jay	187	95k∼119k	278k∼335k	27%∼38%
Abhijeet	112	142k∼153k	369k∼412k	20%∼29%

Figure 6 and the accompanying video show intermediate results from those sequences at different stages of our pipeline. In addition, our reconstructions are suitable for free viewpoint video applications and can be seamlessly integrated into a virtual scene with different illumination as demonstrated in Figures 7 and 8. We obtain the full albedo of the watertight subject by blending the textures observed from each view. To enforce smooth texture transitions, we solve a Poisson equation constrained by averaged color gradients as detailed in Chuang et al. [2009].

Compared to the original data, our meshes are complete and watertight, exhibit less temporal noise, and contain an equivalent or increased amount of surface detail. Naively closing the holes with visual hulls (as mentioned in Vlasic et al. [2009]) produces watertight surfaces, but introduces even more temporal noise. More

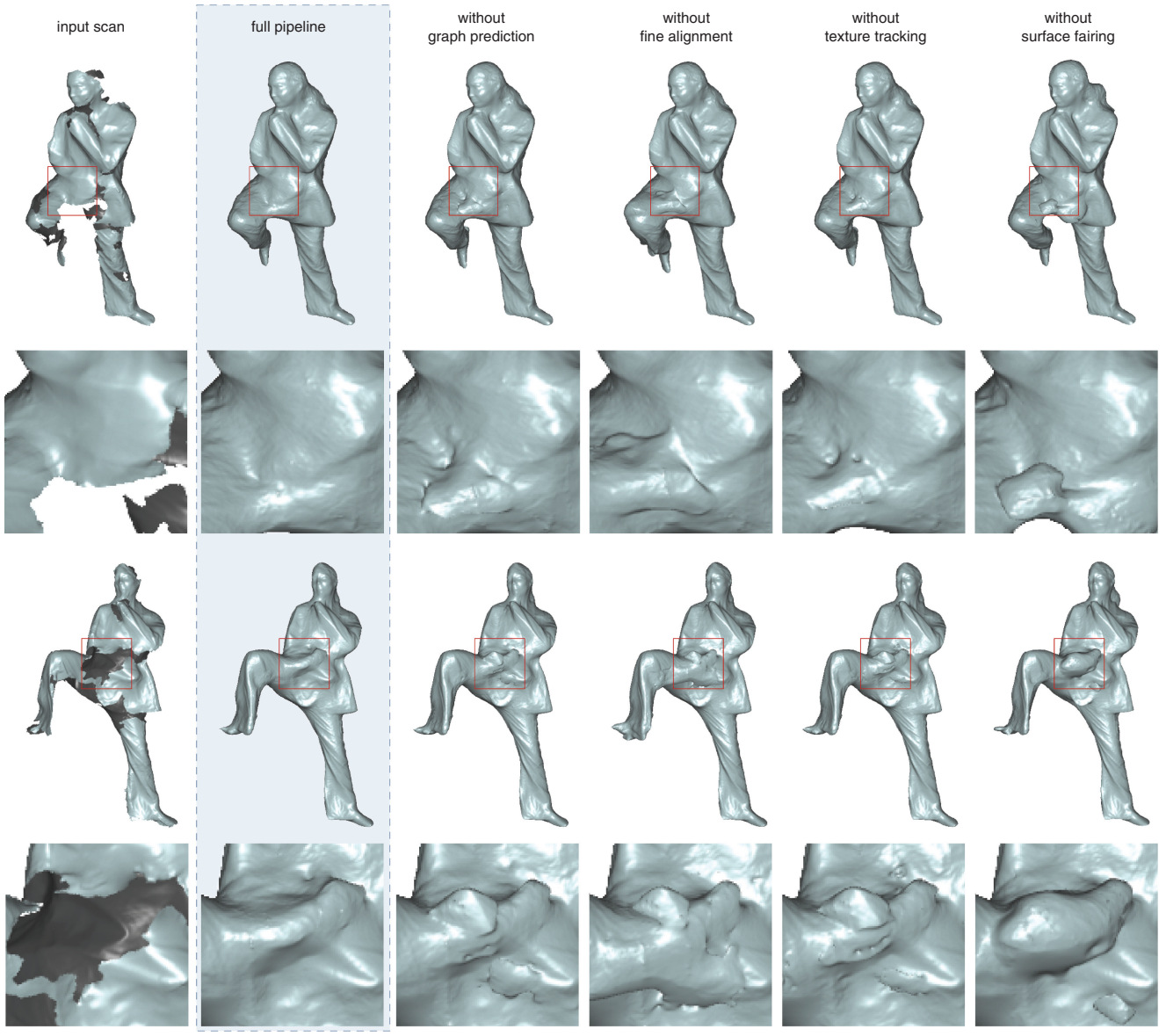


Fig. 4. Comparison between full pipeline and leaving out individual stages of the correspondence computation. The last column clearly shows the importance of surface fairing.

sophisticated methods [Wand et al. 2007] attempt to accumulate surface information over time. However, they have a hard time finding correspondences over many frames of nonrigid incomplete surfaces (second row in Figure 3). Consensus skeleton [Zheng et al. 2010] may be used to determine a consistent topology throughout the whole motion, but we observe similar issues with our data, as it assumes clearly articulated and well-sampled underlying shapes (third row in Figure 3). Sharf et al. [2008] can accumulate surface over time from sparse data such as ours, but may exhibit artifacts with flowing clothes that violate their volume-preserving assumption.

Timing. Ignoring data transfer, the whole pipeline runs at about 9 minutes per frame on a modern machine. The per-frame hole-filling (Section 3.1) takes 40 seconds, Laplacian deformation and

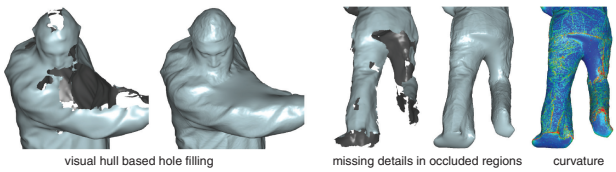


Fig. 5. Our method produces temporally coherent watertight meshes, but the quality of filled holes depends on the visual hull (left) and the unobserved regions have no surface detail (right).

Poisson reconstruction (Section 3.2) add 50 seconds, final detail resynthesis (Section 3.3) 90 seconds, coarse frame-to-frame alignment (Section 4.1) 45 seconds, and the fine-scale alignment (Section 4.2) an additional 320 seconds. The process can be run in

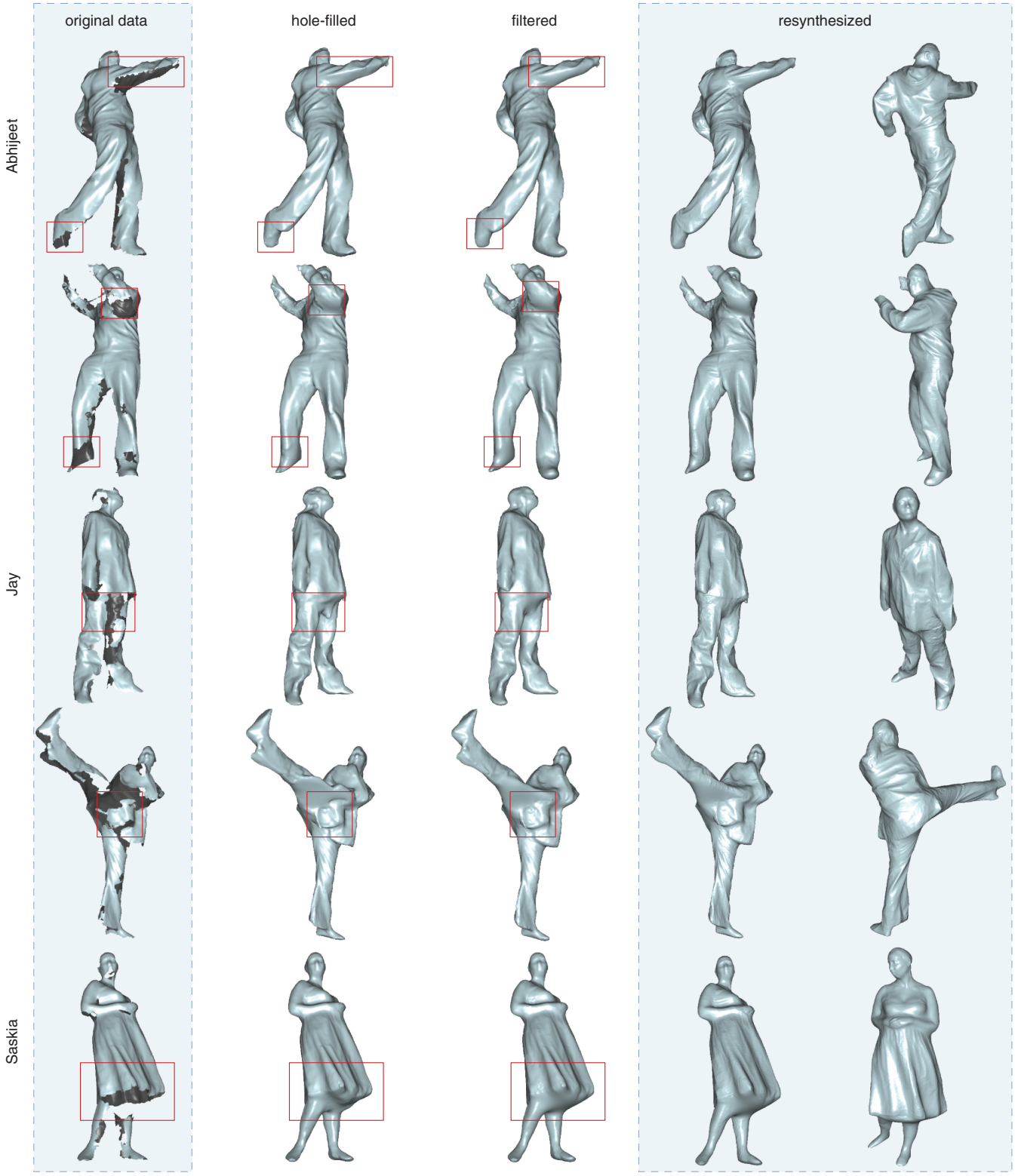


Fig. 6. Left to right: original mesh, hole-filled mesh, temporally filtered mesh, two views of the final mesh with resynthesized detail.

Fig. 7. Our surfaces in conjunction with texture blending [Chuang et al. 2009] are suitable for free viewpoint video. Top: example with gliding cloth, impossible to faithfully reproduce with conventional template-based methods. Bottom: complete digital models produce correct shadows.

parallel for each frame independently, which makes processing many frames of motion reasonable.

Limitations. Our method produces detailed watertight meshes that are smooth over time, but also lends to some limitations. First, the topology of our meshes will always match the (changing and sometimes incorrect) topology of the visual hull since we use it as the initial guess for shape completion (Figure 5 left). Ideally, we would like to extract a single consistent topology for the whole motion. Second, our temporal correspondences are valid between nearby frames, but they cannot be accurately propagated throughout the whole motion. This stands in the way of producing congruent moving meshes that are useful for analysis and editing, and should be addressed with a global approach. Third, the unobserved regions in each frame have no geometric details in them (Figure 5 right). With correspondences throughout the whole motion, the detail could be transferred from frames where those regions are visible. Nevertheless, we see our method as the next logical step towards the ultimate goal of dynamic shape capture, which is to acquire a single moving mesh, consistently parameterized over time, that exhibits all the observed detail and propagates it to the occluded regions throughout the whole motion.

Fig. 8. Reconstructed human performance with full albedo integrated into a virtual scene with different illuminations.

6. CONCLUSIONS

Due to the rapid advances in real-time 3D acquisition technology, the importance of obtaining temporally coherent watertight mesh sequences will be undeniable for many applications involving digitization of dynamic objects. We present the first framework to automatically fill holes with temporal coherent patches without relying on a geometrical template. We have shown that the maturity of nonrigid registration techniques enables us to compute accurate and reliable correspondences for our purpose of filling holes in dynamic shapes. As opposed to other approaches, our method is specifically designed to handle changes in topology. Another advantage is that we can process scan sequences of arbitrary lengths without error accumulation because our correspondence computations are temporally localized. All presented results were produced from high-resolution captured data of real-life performances that are publicly available [Vlasic et al. 2009]. Our key contribution is the interleaved registration/merging scheme which is propagated in a forward-and-backward fashion, the weighted temporal filtering of patches filled using the visual hull, and the integration of the all these components into a complete shape completion framework.

In considering shape completion of dynamic scans as a crucial step in digitization of real-world objects, we anticipate several challenges for future research. Since our proposed approach is purely geometric, a more accurate reproduction of deformations in occluded regions could possibly take into account physical properties that are either user guided or even learned from the captured data. Ultimately, we would like to address the problem of finding dense global correspondences through entire recordings and we postulate that determining them using hole-free surfaces is a simpler problem than using incomplete ones.

ACKNOWLEDGMENTS

We thank the anonymous reviewers, Paul Debevec, Bill Swartout, Randy Hill, Randolph Hall for the constructive feedback and discussions, Saskia Mordijck, Jay Bush, and Abhijeet Ghosh for the input performances, Krystle de Mesa for proofreading, and Duygu Ceylan for helping with the video. We are grateful to Martin Bokeloh, Michael Wand, Qian Zheng, and Baoquan Chen for providing the comparisons to prior work.

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 *Proceedings of the IEEE Computer Vision and Pattern Recognition Conference*.

ALEXA, M. 2003. Differential coordinates for local mesh morphing and deformation. *Vis. Comput. 19,* 2, 105–114.

ALLEN, B., CURLESS, B., AND POPOVIĆ, Z. 2002. Articulated body deformation from range scan data. *ACM Trans. Graph. 21,* 3, 612–619.

ANGUELOV, D., SRINIVASAN, P., KOLLER, D., THRUN, S., RODGERS, J., AND DAVIS, J. 2005. Scape: Shape completion and animation of people. *ACM Trans. Graph. 24,* 3, 408–416.

BAY, H., TUYTELAARS, T., AND VAN GOOL, L. 2008. SURF: Speeded up robust features. *Comput. Vis. Image Underst. 10,* 3, 346–359.

BOTSCH, M. AND SORKINE, O. 2008. On linear variational surface deformation methods. *IEEE Trans. Vis. Comput. Graph. 14,* 1, 213–230.

BRADLEY, D., POPA, T., SHEFFER, A., HEIDRICH, W., AND BOUBEKEUR, T. 2008. Markerless garment capture. *ACM Trans. Graph. 27,* 3, 99.

BROWN, B. J. AND RUSINKIEWICZ, S. 2007. Global non-rigid alignment of 3-d scans. *ACM Trans. Graph. 26,* 3, 21.

BROX, T., BRUHN, A., PAPENBERG, N., AND WEICKERT, J. 2004. High accuracy optical flow estimation based on a theory for warping. In *Proceedings of the 8th European Conference on Computer Vision*. 25–36.

CARR, J. C., FRIGHT, W. R., AND BEATSON, R. K. 1997. Surface interpolation with radial basis functions for medical imaging. *IEEE Trans. Med. Imag. 16*, 96–107.

CARRANZA, J., THEOBALT, C., MAGNOR, M. A., AND SEIDEL, H.-P. 2003. Free-viewpoint video of human actors. *ACM Trans. Graph. 22,* 3, 569–577.

CHANG, W. AND ZWICKER, M. 2009. Range scan registration using reduced deformable models. *Comput. Graph. Forum 28,* 2, 447–456.

CHUANG, M., LUO, L., BROWN, B. J., RUSINKIEWICZ, S., AND KAZHDAN, M. 2009. Estimating the Laplace-Beltrami operator by restricting 3D functions. In *Proceedings of the Symposium on Geometry Processing*.

CORAZZA, S., MÜNDERMANN, L., CHAUDHARI, A., DEMATTIO, T., COBELLI, C., AND ANDRIACCHI, T. P. 2006. A markerless motion capture system to study musculoskeletal biomechanics: Visual hull and simulated annealing approach. *Ann. Biomed. Engin. 34,* 6, 1019–1029.

CURLESS, B. AND LEVOY, M. 1996. A volumetric method for building complex models from range images. In *Proceedings of SIGGRAPH. Computer Graphics Proceedings*, Annual Conference Series, 303–312.

DAVIS, J., MARSCHNER, S. R., GARR, M., AND LEVOY, M. 2002. Filling holes in complex surfaces using volumetric diffusion. In *Proceedings of the Symposium on 3D Data Processing, Visualization, and Transmission*. 428–438.

DAVIS, J., NEHAB, D., RAMAMOORTHI, R., AND RUSINKIEWICZ, S. 2005. Spacetime stereo: A unifying framework for depth from triangulation. *IEEE Trans. Pattern Anal. Mach. Intell. 27,* 2, 296–302.

DE AGUIAR, E., STOLL, C., THEOBALT, C., AHMED, N., SEIDEL, H.-P., AND THRUN, S. 2008. Performance capture from sparse multi-view video. *ACM Trans. Graph. 27,* 3, 98.

DE AGUIAR, E., THEOBALT, C., STOLL, C., AND SEIDEL, H.-P. 2007. Markerless deformable mesh tracking for human shape and motion capture. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*.

GELFAND, N., RUSINKIEWICZ, S., IKEMOTO, L., AND LEVOY, M. 2003. Geometrically stable sampling for the icp algorithm. In *Proceedings of the International Conference on 3D Digital Imaging and Modeling*. 260.

HELD, M. 1998. Fist: Fast industrial-strength triangulation. Tech. rep.

JU, T. 2009. Fixing geometric errors on polygonal models: A survey. *J. Comput. Sci. Technol. 24,* 1, 19–29.

KAZHDAN, M., BOLITHO, M., AND HOPPE, H. 2006. Poisson surface reconstruction. In *Proceedings of the Symposium on Geometry Processing*.

KOJEKINE, N., SAVCHENKO, V., SENIN, M., AND HAGIWARA, I. 2002. Real-time 3d deformations by means of compactly supported radial basis functions. In *Proceedings of Eurographics Short Papers*. 35–43.

LI, H., ADAMS, B., GUIBAS, L. J., AND PAULY, M. 2009. Robust single-view geometry and motion reconstruction. *ACM Trans. Graph. 28,* 5.

LI, H., SUMNER, R. W., AND PAULY, M. 2008. Global correspondence optimization for non-rigid registration of depth scans. *Comput. Graph. Forum 27,* 5, 1421–1430.

LIEPA, P. 2003. Filling holes in meshes. In *Proceedings of the Symposium on Geometry Processing*. 200–205.

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

NEHAB, D., RUSINKIEWICZ, S., DAVIS, J., AND RAMAMOORTHI, R. 2005. Efficiently combining positions and normals for precise 3d geometry. *ACM Trans. Graph. 24,* 3, 536–543.

PEKELNY, Y. AND GOTSMAN, C. 2008. Articulated object reconstruction and markerless motion capture from depth video. *Comput. Graph. Forum 27,* 2, 399–408.

SAND, P., MCMILLAN, L., AND POPOVIĆ, J. 2003. Continuous capture of skin deformation. *ACM Trans. Graph. 22,* 3, 578–586.

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 Trans. Graph. 27,* 5, 1–10.

SIBSON, R. AND STONE, G. 1991. Computation of thin-plate splines. *SIAM J. Sci. Statist. Comput. 12,* 6, 1304–1313.

STARCK, J. AND HILTON, A. 2003. Model-based multiple view reconstruction of people. In *Proceedings of the International Conference on Computer Vision*. 915–922.

SÜSSMUTH, J., WINTER, M., AND GREINER, G. 2008. Reconstructing animated meshes from time-varying point clouds. In *Proceedings of the Symposium on Geometry Processing*. *27,* 5, 1469–1476.

THEOBALT, C., AHMED, N., LENSCH, H., MAGNOR, M., AND SEIDEL, H.-P. 2007. Seeing people in different light-joint shape, motion, and reflectance capture. *IEEE Trans. Vis. Comput. Graph. 13,* 4, 663–674.

VLASIC, D., BARAN, I., MATUSIK, W., AND POPOVIĆ, J. 2008. Articulated mesh animation from multi-view silhouettes. *ACM Trans. Graph. 27,* 3, 97.

VLASIC, D., PEERS, P., BARAN, I., DEBEVEC, P., POPOVIĆ, J., RUSINKIEWICZ, S., AND MATUSIK, W. 2009. Dynamic shape capture using multi-view photometric stereo. In *SIGGRAPH Asia'09 ACM* SIGGRAPH Asia 2009 Papers. 1–11.

WAND, M., ADAMS, B., OVSJANIKOV, M., BERNER, A., BOKELOH, M., JENKE, P., GUIBAS, L., SEIDEL, H.-P., AND SCHILLING, A. 2009. Efficient reconstruction of nonrigid shape and motion from real-time 3d scanner data. *ACM Trans. Graph. 28,* 2, 15.

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

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

ZHANG, L., CURLESS, B., HERTZMANN, A., AND SEITZ, S. M. 2003. Shape and motion under varying illumination: Unifying structure from motion, photometric stereo, and multi-view stereo. In *Proceedings of the International Conference on Computer Vision*. 618.

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

ZHENG, Q., SHARF, A., TAGLIASACCHI, A., CHEN, B., ZHANG, H., SHEFFER, A., AND COHEN-OR, D. 2010. Consensus skeleton for non-rigid space-time registration. *Comput. Graph. Forum 29,* 2, 635–644.

Received October 2010; revised April 2011; accepted July 2011