Markerless Garment Capture

Derek Bradley[1]   Tiberiu Popa[1]   Alla Sheffer[1]   Wolfgang Heidrich[1]   Tamy Boubekeur[2]
1) University of British Columbia,   2) TU Berlin

Figure 1: *Left to right: an actor performing in the capture setup; one of sixteen views from the camera array; reconstructed T-shirt geometry; the real T-shirt is replaced by a rendering of the captured geometry with different appearance characteristics.*

Abstract

A lot of research has recently focused on the problem of capturing the geometry and motion of garments. Such work usually relies on special markers printed on the fabric to establish temporally coherent correspondences between points on the garment's surface at different times. Unfortunately, this approach is tedious and prevents the capture of off-the-shelf clothing made from interesting fabrics.

In this paper, we describe a marker-free approach to capturing garment motion that avoids these downsides. We establish temporally coherent parameterizations between incomplete geometries that we extract at each timestep with a multiview stereo algorithm. We then fill holes in the geometry using a template. This approach, for the first time, allows us to capture the geometry and motion of unpatterned, off-the-shelf garments made from a range of different fabrics.

CR Categories: I.3.3 [COMPUTER GRAPHICS]: Picture/Image Generation—Digitizing and scanning; I.3.5 [COMPUTER GRAPHICS]: Computational Geometry and Object Modeling—Geometric algorithms, languages, and systems; I.4.1 [IMAGE PROCESSING AND COMPUTER VISION]: Digitization and Image Capture—Imaging geometry.

Keywords: Object Scanning/Acquisition, Cloth Modeling, Image Processing, Motion Capture, Surface Reconstruction

1 Introduction

Capturing the geometry of moving garments and other cloth is a problem that has recently seen a significant amount of research interest. One goal of such research is to provide a data-driven alternative to cloth simulation in much the same way motion capture provides an alternative to character animation. Perhaps more importantly, cloth capture promises to become an indispensable tool for postprocessing and augmenting live action sequences with computer rendered content.

Consider, for example, a live action sequence in which the texture or material properties of a garment are to be changed in post production (see Figure 1). In order for synthetic surface texture to move correctly with the real cloth geometry, it is not only necessary to capture the 3D shape of the garment in each time step, but also to establish a temporally consistent parameterization of the surface that prevents the texture from floating on it.

For this reason, the state of the art in cloth and garment capture is to use markers printed on the garment to simultaneously track geometry and parameterization. The need for such markers unfortunately makes existing cloth capture approaches unattractive for several reasons. First, the effort to create garments with unique marker patterns can be prohibitive for some applications. Second, the types of fabrics on which markers can be printed are typically limited, making it difficult to capture fabric-specific differences in garment motion. However, the characteristic differences between, say, silk, cotton, and fleece, are precisely the reason why one would want to capture cloth as opposed to simulating it. A key goal of our work is thus to capture the motion of specific existing garments, with a specific design, and made from a specific fabric.

In this paper, we present a marker-free approach to garment capture that realizes this goal of capturing off-the-shelf garments. We use a multiview stereo approach for obtaining initial garment geometry for each time step. Although this initial geometry can have a significant number of holes, by leveraging low fabric stretch and utilizing the topology of the scanned garments, we are able to produce a parameterization of the geometry that is consistent across time, without explicitly tracking motion. This temporally consistent parameterization helps us fill the holes in the geometry, producing a high quality final result. A detailed overview of our method is provided in Section 3.

2 Related Work

This section describes related methods for cloth capture and provides a short overview of geometry processing techniques related to our work.

2.1 Cloth and Garment Capture

Over the last several years, techniques for capturing the motion of cloth have evolved from methods for tracking a single cloth sheet to systems for reconstructing full garments under complex motion.

Sheets of cloth and partial garments: Early work on this topic focused on single sheets of cloth with existing textures [Pritchard and Heidrich 2003; Scholz and Magnor 2004], or small subsets of a full garment with patterns that were custom-printed onto the fabric [Guskov et al. 2003]. Hasler et al. [2006] use an analysis-by-synthesis approach that sets their method apart from the other work. They too, however, rely on markers in the form of patterned cloth.

This early work identifies the basic task of combining the geometry of the cloth with a temporally coherent parameterization and introduces the use of markers as a solution to this problem. Capturing full garments with large areas of occlusion is, however, significantly more challenging.

Full garments: A first solution to this problem was proposed by Scholz et al. [2005], who suggested a special marker pattern composed of a matrix of colored dots, in which each 3×3 neighborhood of dots is uniquely identifiable. A simple thin-plate approach was used to fill any holes in the geometry. White et al. [2007] improved on these results by thoroughly analyzing the design of suitable marker patterns. They also improve on the hole filling, using a new data-driven technique. This work is currently the state of the art in garment capture and produces excellent results. However, like the work by Scholz et al., White et al.'s method requires custom-tailored garments made from fabric on which the marker patterns have been printed. It thus cannot deal with arbitrary off-the-shelf clothing.

Markerless approaches: Recently, some markerless capture techniques were proposed, albeit with a somewhat different focus than ours. Bhat et al. [2003] capture the motion of sheet samples made from different fabric to estimate parameters for a cloth simulation. Rosenhahn et al. [2007] focus on estimating the pose of a human wearing a garment using an analysis-by-synthesis approach similar to Hasler et al. [2006]. While they do show some approximate skirt geometry, they do not discuss how to capture the exact shape of complicated garments in the presence of occlusions. In a recent paper, Hernandez et al. [2007] propose a photometric stereo approach for capturing deformable geometry using garments as an example. While their approach is technically marker-free, they require patterned cloth mesostructure such as knitting patterns to compute optical flow. They do not discuss how to handle occlusion, and due to the interaction of differently colored light sources it is not clear how to extend their method to $360°$ surround capture.

Face capture: The overall design of our garment capture system has similarities to facial capture systems such as [Zhang et al. 2004] or [Bickel et al. 2007]. However, garment capture is fundamentally a different problem due to inherent oclusions and chaotic folding and rippling nature of fabrics.

Unlike previous methods, our technique is able to reconstruct high resolution models of full garments under normal human motion without the need for printed on-surface markers or high-frequency surface texture. In this sense our effort is analogous to the push for markerless methods in traditional motion capture (e.g. [Cheung et al. 2005; de Aguiar et al. 2007]).

2.2 Related Geometry Processing Techniques

From a geometry perspective, garment motion capture can be posed as a problem of tracking a deformable surface over time. Recently, two solutions have been presented for this problem, starting from an initial set of point clouds [Wand et al. 2007; Mitra et al. 2007]. While Wand et al. reconstruct both the geometry and temporal correspondences, and Mitra et al. register the point clouds without computing explicit correspondences, both methods assume that the motion is locally rigid. Therefore, these methods would not perform well for cloth capture.

Establishing temporal correspondences between acquired frames of a moving surface requires a bijective mapping, commonly referred to as *cross-parameterization* [Kraevoy and Sheffer 2004] or *inter-surface mapping* [Schreiner et al. 2004], between the frames. In this paper, we use the term *consistent cross-parameterization* of multiple frames to mean a cross-parameterization from-frame-to-frame that maps a point in one frame to the same point, subject to motion, in all other frames. We also say that meshes are *compatible* if they have the same connectivity and an explicit vertex correspondence. Most methods for cross-parameterization and compatible remeshing of surfaces rely on accurate user-provided correspondences for a set of markers on the surfaces [Allen et al. 2003; Schreiner et al. 2004; Kraevoy and Sheffer 2004; Anguelov et al. 2005; Kraevoy and Sheffer 2005]. This manual positioning of markers is impractical for hundreds or thousands of frames produced by a capture process, which is the main reason why existing cloth capture approaches rely on printed markers.

In contrast, we establish a consistent cross-parameterization between acquired frames of a moving surface through the use of a small number of *off-surface anchors* computed on-the-fly, using user-assisted keyframing when necessary. Thus we no longer need markers printed on the fabric.

3 Overview

Our method for markerless garment capture consists of four key components, illustrated in Figure 2:

Acquisition: We deploy a unique $360°$ high-resolution acquisition setup using sixteen inexpensive high-definition consumer video cameras. Our lighting setup avoids strong shadows by using indirect illumination. The cameras are set up in a ring configuration in order to capture the full garment undergoing a range of motions.

Multiview Reconstruction: The input images from the sixteen viewpoints are fed into a custom-designed multiview stereo reconstruction algorithm [Bradley et al. 2008] to obtain an initial 3D mesh for each frame. The resulting meshes contain holes in regions occluded from the cameras, and each mesh has a different connectivity.

Consistent Cross-Parameterization: We then compute a consistent cross-parameterization among all input meshes. To this end, we use strategically positioned *off-surface anchors* that correspond to natural boundaries of the garment. Depending on the quality of boundaries extracted in the multiview stereo step, the anchors are placed either fully automatically, or with a small amount of user intervention.

Compatible Remeshing and Surface Completion: Finally, we introduce an effective mechanism for compatible remeshing and hole completion using a template mesh. The template is constructed from a photo of the garment, laid out on a flat surface. We cross-parameterize the template with the input meshes, and then deform it into the pose of each frame mesh. We use the deformed template as the final per-frame surface mesh. As a result, all the reconstructed per-frame meshes have the same, compatible, connectivity, and the holes are completed using the appropriate garment geometry.

In Sections 4 to 7 we elaborate on the components of our method. We then show results in Section 8, and conclude with a discussion of the method.

4 Acquisition

This section describes our acquisition setup, including the hardware, calibration, and synchronization.

Cameras. Our acquisition setup consists of sixteen high-definition Sony HDR-SR7 video cameras, mounted in a ring around
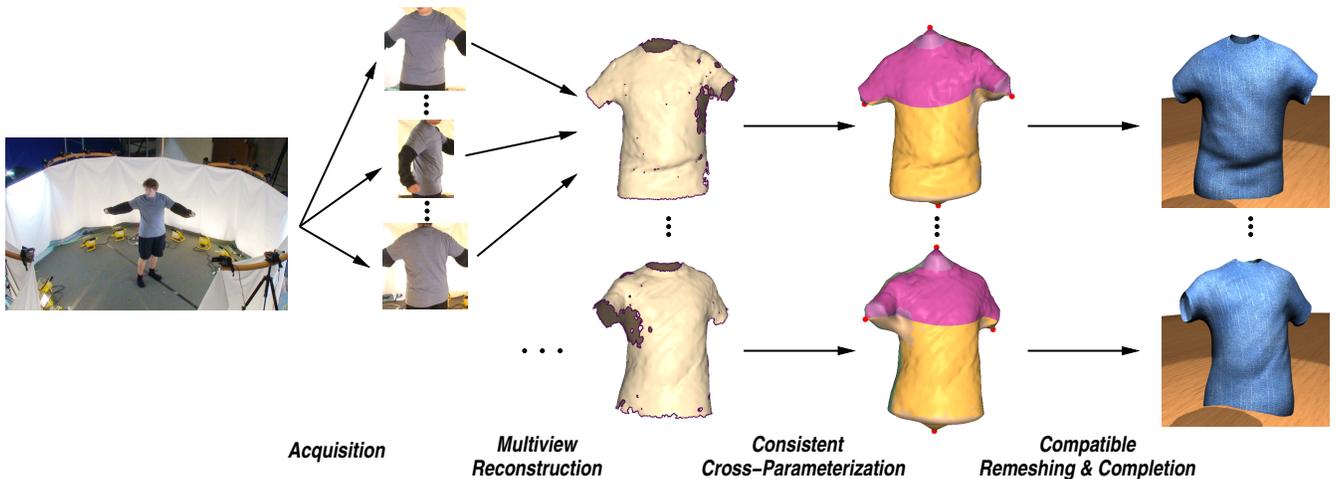
Figure 2: *Overview of our technique.*

the actor wearing the garment to be captured. External and internal camera parameters are calibrated using ARTag markers and the automatic calibration technique by Fiala and Shu [2005].

The cameras record 29.97 frames per second, stored as approximately sixty fields per second on camera-internal hard drives. We convert each field to a full resolution image using a simple spatio-temporal de-interlacing technique applied to the missing scanlines. Our method weights spatial interpolation higher in regions with fast motion, and uses simple temporal interpolation where there is no motion. A pixel $I(x,y,t)$ on a missing scanline is reconstructed as

$$I(x,y,t) = w \cdot \frac{I(x,y,t-1) + I(x,y,t+1)}{2} + (1-w) \cdot \frac{I(x,y-1,t) + I(x,y+1,t)}{2},$$

where w is a Gaussian function of the difference $I(x,y,t-1) - I(x,y,t+1)$, a choice that is inspired by the Bilateral filter [Tomasi and Manduchi 1998]. The final data stream is in the form of sixteen high-resolution (1920×1080) images, captured at sixty images per second.

Synchronization. Since all cameras operate independently, and cannot be synchronized using any external trigger, software synchronization needs to be applied in a post-process. We use a *moving-object* synchronization method. At the beginning of each capture sequence we rotate our ARTag calibration grid approximately $360°$ on a manually operated turn table over a period of ten seconds. Each camera auto-detects the grid in each frame and the amount of rotation is used to determine a temporal offset. Although this method yields offsets with sub-frame precision, we choose to round the offsets to the nearest integer field. At sixty fields per second, we observe that artifacts due to sub-field synchronization differences are not noticeable during moderately fast motion. For high speed motion, optical flow could be computed from frame to frame for each camera and the sub-frame offset could be used to interpolate between frames. We leave such interpolation for future work.

Lighting. Some care has to be taken in the lighting setup as well. As in all movie and video shots, bright illumination is necessary to avoid noise and motion blur by reducing both the camera gain and the exposure time. Moreover, direct illumination by a few bright spot-lights causes hard shadows that exceed the dynamic range of the cameras, thus preventing geometry reconstruction in the shadowed regions. The ideal illumination is therefore bright, diffuse lighting, not unlike that used on movie sets.

Figure 3: *Acquisition setup.*

To achieve this goal, we devised the indirect lighting setup depicted in Figure 3. The outer ring of our setup is lined with white cloth, which we use as a diffusing reflector. Sixteen inexpensive 500W halogen lights are placed inside the ring, pointing outwards at the cloth. This setup results in a bright, uniform illumination of the ring interior.

5 Multiview Reconstruction

The captured and processed images are used as input for a multiview stereo reconstruction algorithm that we developed for this purpose. Our method, which is described in full detail elsewhere [Bradley et al. 2008], was specifically designed to work well with the relatively small number of views available in our setup. According to the Middlebury multiview stereo evaluation [mview ; Seitz et al. 2006], this method is currently the best multiview algorithm for such datasets in terms of both precision and performance.

For the sake of completeness, we give a brief outline of our algorithm in the following. The algorithm has two stages, binocular stereo on image pairs, followed by surface reconstruction. Figure 4 shows a diagram of the individual stages.
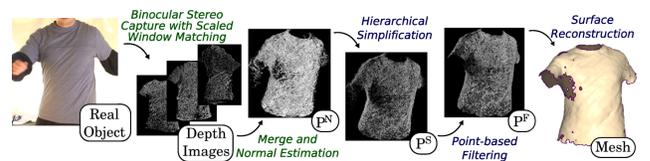

Figure 4: *Multiview reconstruction: starting from a set of depth images, a 3D point cloud is generated (P^N), simplified (P^S), filtered (P^F) and meshed.*

The binocular stereo part of our algorithm creates depth maps from pairs of adjacent viewpoints. After rectifying the image pairs, we

observe that the relationship between the views causes distortions of the comparison windows. We compensate for these distortions by employing a *scaled-window matching* technique. The compensation improves the quality, especially in high curvature regions and for sparse viewpoints (i.e., large baselines). The depth images from the binocular stereo pairs are converted to 3D points and merged into a single dense point cloud.

The second part of the algorithm aims at reconstructing a triangle mesh from the initial point cloud. Instead of implicit reconstructions, which smoothly interpolate over holes but introduce stretch in doing so, we prefer reconstruction methods that keep the holes so that they can be filled based on parameterization information generated further down our pipeline. We first use a hierarchical vertex clustering approach to eliminate excessive sampling density. After downsampling, the simplified point cloud still contains some noise. Unlike other methods, which integrate noise removal into the meshing algorithm, we control noise removal explicitly in a separate phase. The processed samples are finally triangulated using an improved version of the method by Boubekeur et al. [2005].

All stages of the algorithm are designed to scale well. The output of the reconstruction is a triangle mesh for each video frame. Even though our multiview stereo algorithm is the state of the art for sparse views according to the Middlebury benchmark, significant holes remain in occluded and saturated image regions. The subsequent stages of our cloth capture system therefore have to robustly deal with missing data in the cross-parameterization between the frames, as well as fill in the missing regions from a template. Figure 5 shows the multiview reconstructed frames from three different garments in motion, along with the final reconstruction result after all stages of our algorithm.

Figure 5: *Result of multiview reconstruction for different garments (middle row). One of the sixteen input images is shown above each result, and the final reconstruction after parameterization and completion is shown below.*

6   Consistent Cross-Parameterization

The next step of our method constructs a consistent cross-parameterization between the incomplete per-frame meshes. In order for cloth capture to be useful, the parameterization needs to capture the motion of the garment, mapping a point on the garment at one frame to the position of the same point on the subsequent frame. Inconsistencies in the mapping are likely to lead to visible artifacts such as "floating" texture on the garment surface. We have
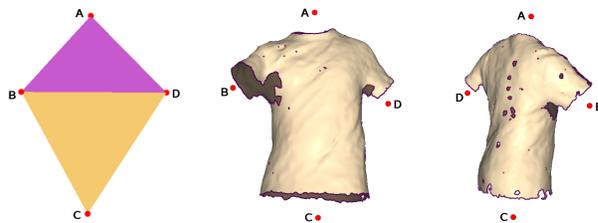

Figure 6: *Off-surface anchor vertices are used to establish boundary correspondences between the tetrahedral base mesh (left) and the different frames of the same garment (right).*

no explicit information on the motion of the garment, and estimating such motion is quite challenging. Instead we rely on geometric properties of the garments to compute a consistent parameterization. Many real fabrics exhibit very low stretch in regular use (see e.g. [Goldenthal et al. 2007]), and thus we can assume that a consistent parameterization of garments is (nearly) isometric. Moreover, since most garments exhibit no rotational symmetries, such an isometric parameterization is unique. A key insight of our work is therefore that we can obtain a consistent parameterization between frames simply by finding an as-isometric-as-possible, i.e. stretch-minimizing, parameterization between the per-frame meshes, instead of employing some form of tracking.

Several of the cross-parameterization methods reviewed in Section 2 can process incomplete meshes. However, in our application we face additional challenges due to the large number of frames to be processed. We cannot manually add markers in every frame to assist the parameterization, nor can we employ computation-intensive approaches such as the work by Anguelov et al. [2005] or Schreiner et al. [2004].

In order to parameterize the meshes in a reasonable time frame, we use the *base mesh* approach [Praun et al. 2001; Schreiner et al. 2004; Kraevoy and Sheffer 2004; Kraevoy and Sheffer 2005]. A base mesh (Figure 6, left) is a low-resolution mesh whose vertices correspond to a consistent set of marker vertices on the input meshes. In base mesh parameterization, each mesh is mapped to the base, such that the marker vertices map to the corresponding base vertices. This provides a cross parameterization between the inputs, where a map from one mesh to another is given by combining the map from the first mesh to the base with the inverse map from the base to the second mesh. In all previous methods, the on-surface marker vertices are manually specified by the user for each input mesh. As mentioned before, it would be infeasible in our setting to provide such markers manually for all frames. However, we observe that garments, in contrast to generic geometric models, have distinct, identifiable, topology with several boundary loops. For instance a T-shirt has four boundaries: a neckline, two sleeves and a waistline. Our algorithm identifies the corresponding boundaries across all frames, as discussed in Section 6.1, and associates a floating, off-surface anchor vertex with each boundary (see Figure 6). The anchors replace the on-surface markers in the base mesh construction and subsequent parameterization (Section 6.2).

In general, even when using a base mesh for stretch-minimizing cross-parameterization, we need to measure and optimize the stretch directly between the input meshes, a fairly time-consuming process. However, given two meshes that are isometric, mapping both of them using stretch-minimization to the same base mesh, we can expect the maps to be nearly identical, resulting in a nearly isometric map between the input meshes. Even in the presence of holes, careful design of a low-stretch base mesh parameterization algorithm, discussed in Section 6.2, allows us to compute cross-parameterizations that are consistent across the entire sequence by

parameterizing the meshes independently to the base. This makes our method fast enough to be used in practice and allows for parallel processing of individual frames.

6.1 Positioning Off-Surface Anchors

The off-surface anchors in our setup correspond to the natural boundaries of the processed garment and are mapped to the vertices of the base mesh during parameterization (Figure 6). The anchors are placed manually for the first frame of each video sequence and are updated over time to follow the motion. Each anchor is updated automatically, except where the multiview algorithm has failed to reconstruct a sufficient amount of geometry on the corresponding mesh boundary (see Figure 7, right). As mentioned in Section 5, missing geometry can be a problem due to occlusion and the relatively small number of cameras.
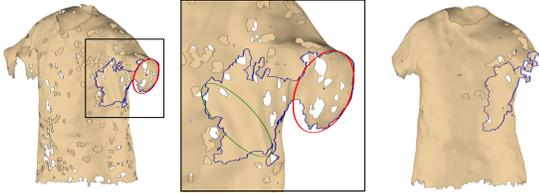
Figure 7: *Garment boundary tracking. Left: the automatic method correctly chooses the red ellipse based on confidence, size and temporal similarity. Center: using confidence alone, the green ellipse would be incorrectly chosen. Right: the boundary loop contains very little of the actual garment boundary and user-assisted tracking is required.*

Automatic boundary tracking. We observe that boundary loops of typical garments have an approximately elliptical shape. Given an elliptical approximation of each boundary, the corresponding anchor is placed a fixed distance away from the ellipse center with an associated normal vector pointing away from it. The normals are used in the hole filling process, as discussed in Section 6.2. To find the best fitting ellipse, we use a variant of the RANSAC algorithm [Fischler and Bolles 1981]: for each boundary loop of the mesh that corresponds to a garment boundary, we iteratively choose six boundary vertices at random and then fit an ellipse to these vertices. The *confidence* for any given ellipse is calculated using the number of boundary vertices that agree with that ellipse. The selection of a best fitting ellipse is further guided by size and temporal similarity to the position of the corresponding ellipse in the previous frame. The number of RANSAC iterations required in our setting is on the order of $\binom{n}{6}$, where n is the length of a typical boundary. Missing portions of the geometry such as sections under the arm, which are often not captured by the multiview stereo algorithm, can cause significantly prolonged boundary loops. As a result, 50K to 100K iterations may be required per boundary for such frames, taking fifteen to thirty seconds per frame.

Anchor positions and normals are smoothed in time using Gaussian smoothing. Figure 8 shows the resulting anchors for a short sequence of a garment undergoing motion. The anchors with normals are shown for each frame, and the garment geometry is shown for one frame of the sequence. This method can deal with significant missing geometry on the boundary, as shown in Figure 7 (left). However, for degenerate boundaries such as the one shown in Figure 7 (right), the automatic positioning fails. A short consecutive sequence of poor frames can still be handled gracefully, as the method will auto-detect and skip over the problematic boundaries when no suitable ellipse is found. However, this technique cannot recover from sequences with a large number of consecutive unrecognizable boundaries.
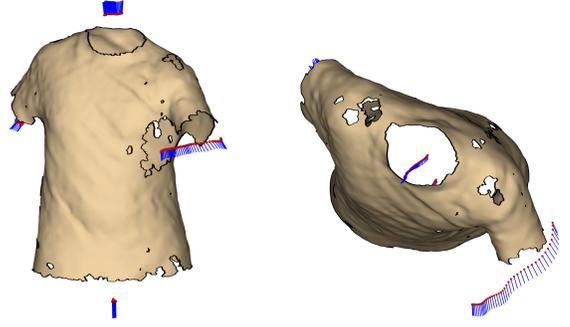

Figure 8: *Automatic off-surface anchor positioning across a sequence of frames. Anchor vertices are shown in red, while normals are shown in blue. Front view (left) and top view (right).*

User-assisted boundary tracking. For sequences where the automatic anchor placement fails for a larger number of consecutive frames, we have developed a semi-automatic alternative. For a number of keyframes, the user chooses an ellipse for each boundary from a selection of high-confidence ellipses determined by the RANSAC algorithm. Choosing an ellipse from a precomputed set typically takes about 10 seconds per boundary. In between keyframes, the anchor positions are interpolated using splines. The required density of keyframes depends on the speed of motion; for fast motions, a keyframe approximately every 30 frames (1/2 second) yields good results.

6.2 Base Mesh Parameterization

For each input garment sequence we first construct a base mesh and then map the per-frame meshes onto the base. The base is constructed by positioning the vertices at the locations of the corresponding anchors in a canonical pose (Figure 6, left). Given the common base, we compute a stretch-minimizing parameterization of each mesh onto the base using the computed anchors.

As an initialization step, a small cone is created around each anchor, with the apex pointing in the direction of the anchor normal (right). These cones are used later-on to facilitate smooth hole completion. Since our parameterization method mostly follows the work by Kraevoy and Sheffer [2005], we focus our description on the differences with their method and refer the interested reader to their paper for more details.
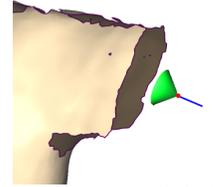
The algorithm first segments the mesh into charts corresponding to the base mesh facets and constructs an initial parameterization by mapping each chart to the corresponding facet. Since the anchor cones form separate connected components, we triangulate the gaps between them and the corresponding boundaries before computing the segmentation. Given the initial parameterization, the method iterates between re-triangulating the holes in the mesh, including the gaps between the anchor cones and the actual surface, and reparameterizing the mesh onto the base. The main goal of the process is to parameterize the mesh onto the base with minimal stretch.

During re-triangulation, Kraevoy and Sheffer [2005] first triangulate holes on the base, and then project the new vertices back to 3D using a Laplacian type formulation that minimizes

$$\sum_i (P_i - \sum_{j \in N(P_i)} \omega_{ij} P_j)^2, \qquad (1)$$

where P_i are the projected vertices with neighborhood $N(P_i)$, and ω_{ij} are the normalized mean value weights [Floater 2003] from the

base embedding. This formulation creates a membrane type surface, filling the holes in 3D with C^0 continuity along hole boundaries (Figure 9, center). The discontinuities along the boundaries lead to a sub-optimal estimation of the actual hole shape and area and increase the parameterization stretch. We rectify this problem by introducing additional terms $(P_i' - \sum_{j \in N(P_i')} \omega_{ij} P_j)^2$ into the minimized functional in Equation 1 for the vertices P_i' on the hole boundary. While the 3D positions of these boundary vertices remain fixed, the new terms influence the positions of the hole vertices adjacent to them. The additional terms effectively eliminate the discontinuities across hole boundaries, creating a C^1 effect (Figure 9, right).

Note that the membrane reprojection is suitable for establishing the base mesh parameterization, but is not always sufficient for final garment hole completion. Filling holes in complex geometric regions, such as armpits, using the generic membrane generates oversmoothed geometry (Figure 12, left). We address this problem in Section 7, with a template-based remeshing and surface completion technique.
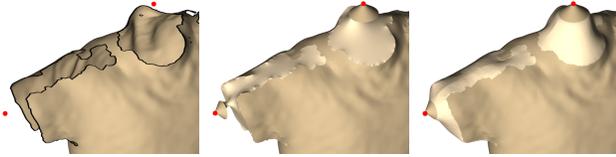

Figure 9: *Smooth hole completion during parameterization. Left: an incomplete T-shirt. Center: C^0 membrane. Right: C^1 membrane. The membranes are rendered in lighter tones for visualization.*

Since the geometry connecting the anchors to the mesh is iteratively updated as the parameterization improves, any inaccuracy in the anchor position is automatically corrected, thus the anchors need not be placed as accurately as on-surface markers. Therefore, our anchor-based parameterization is fairly robust to noise and artifacts in the input meshes.

The result of the base mesh parameterization method for two input frames is illustrated in Figure 10. Thanks to the near-isometry between the frames, the natural boundaries on the normal maps are well aligned. Once each mesh has been parameterized onto the common base, a trivial cross-parameterization is established between all frames.

7 Compatible Remeshing and Surface Completion

The acquired garment frames typically have numerous holes, which are filled with a C^1 continuous membrane during the parameterization. For final reconstruction, these membranes work well for flat or convex regions, but generate an oversmoothed surface in saddle-like regions (see Figure 12). One option for replacing the membrane with more realistic geometry is data-driven hole filling [White et al. 2007]. White et al. complete hole geometry using the mapping from other input frames that do not contain the hole. However, some regions of the garment may be occluded in all input frames. For this reason, we use a template mesh, which we generate by "inflating" a photo of the garment to provide a more faithful geometry completion. Additionally, we establish common, compatible connectivity by using the template triangulation for all frames.

Template construction. We construct a template mesh from a single photo of the garment, laid out on a flat surface (Figure 11,
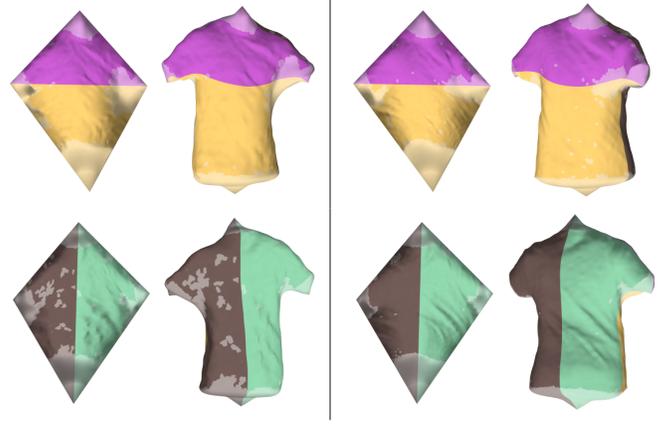

Figure 10: *Two frames parameterized onto the base mesh, illustrated as normal maps (top row: front view, bottom row: back view). The charts corresponding to the base mesh faces are shown on the garments. Again, hole regions are rendered in lighter tones.*
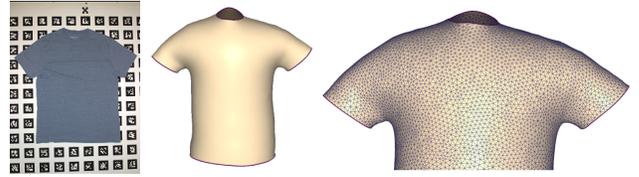

Figure 11: *Template construction. A photo of the garment (left) is inflated to 3D.*

left). The silhouette of the garment is extracted from the image and the interior is triangulated [Shewchuk 1996] creating a 2D mesh. The mesh is duplicated for the back surface, and garment boundaries are indicated by the user with brush strokes.

The 2D mesh is then inflated into 3D, using a variant of the iterative physical simulation approach of Mori et al. [2007]. Each iteration consists of two steps. First, every face is moved slightly in its normal direction, to mimic the effect of internal pressure. Then the length of each edge is adjusted to preserve the stretch of the material. In the work of Mori et al., material stretching is prevented while compression is tolerated, a desired effect for their application. For our garment template, we wish to prevent both stretch and compression. Thus, we modify the second step of the algorithm to include a compression prevention condition. In this step, the displacement d_{v_i} of a vertex v_i is computed as a weighted sum of the forces (t_{ij}) from the neighboring edges (E_i):

$$d_{v_i} = \frac{\sum_{e_{ij} \in E_i} \{A(e.leftface) + A(e.rightface)\} t_{ij}}{\sum_{e_{ij} \in E_i} \{A(e.leftface) + A(e.rightface)\}} \quad (2)$$

$$t_{ij} = \begin{cases} 0.5 \cdot (v_j - v_i) \cdot \frac{|v_i - v_j| - l_{ij}}{|v_i - v_j|} & \text{if } |v_i - v_j| \geq l_{ij} \\ 0.5 \cdot (v_i - v_j) \cdot \frac{l_{ij} - |v_i - v_j|}{l_{ij}} & \text{if } |v_i - v_j| < l_{ij} \end{cases}, \quad (3)$$

where $A(f)$ is the area of a face f and l_{ij} is the length of an edge e_{ij} in the initial 2D mesh. This formulation differs from Mori et al. in Equation 3, where the first condition is for preventing stretch and the second condition is for preventing compression.

A template is created only once for each garment. The constructed template of a T-shirt is shown in Figure 11.

Remeshing and completion The cross-parameterization of the input meshes will be used for consistent texture mapping and tem-
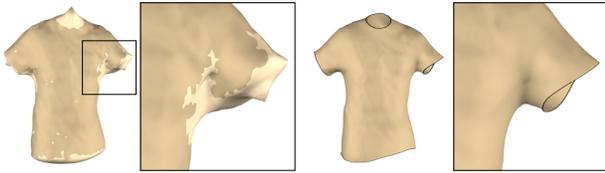
Figure 12: *Initial C^1 smooth membrane (left) and result of surface completion using the template from Figure 11 (right).*

poral smoothing of the geometry. Thus, we require that the meshes for all frames be compatible, i.e., have the same connectivity and explicit vertex correspondence. This is achieved by projecting the uniform triangulation of the template mesh onto each frame.

The template is parameterized onto the base mesh in the same way as the input frames establishing a mapping between them. This mapping is sufficient for remeshing each frame with the template connectivity by simply mapping the vertices of the template to each frame. However, for the hole regions in each frame we prefer to use the geometry of the template instead of the membrane. We therefore use the mapping from the template onto each frame to map only those vertices of the template that map to the original frame geometry. To obtain the position of the remaining vertices, we deform the template into the pose of the frame, using the mapped vertex positions as constraints. We use the deformation method of Sheffer and Kraevoy [2004], although other algorithms could be used instead. Figure 12 shows the result of the completion for one frame. This technique produces a compatible, uniform triangulation of the whole frame sequence, while completing the holes with surface patches that match the captured garment.

Finally, the frame sequence is smoothed in the temporal domain using Gaussian smoothing to remove temporal noise.

8 Results

The results of different garment captures are shown in Figures 13-18. We are able to render the captured geometry separately, as a replacement of the original garment, and as an augmentation to the original video frame. Our technique is, to our knowledge, the first method that is able to produce all three of these results from a single capture. In particular, augmentation would not be possible without markers remaining visible in most of the previous work. We encourage the reader to view the full sequences in the video.

Figure 13 shows different frames from a T-shirt sequence. This is the same T-shirt that was used in illustrations throughout the paper. Due to boundary issues discussed in Section 6, the user assisted anchor placement had to be employed for parts of this sequence. Note that this garment is rather tight, and the actor's stomach motion is correctly reconstructed.

For a fleece vest, shown in Figure 14, off-surface anchors were placed entirely using the automatic boundary tracking technique. We also note that the multiview algorithm produced only small holes for the vest, unlike the large holes caused by occlusion on the T-shirt. As a consequence, we were able to reconstruct the vest without creating a template mesh. Instead, a compatible remeshing was achieved by performing a uniform triangulation on the base mesh and then projecting the geometry to each frame. The fleece vest also illustrates that we can indeed capture garments made from different fabrics. While it would be relatively straightforward to print marker patterns onto the cotton fabric of the T-shirt, producing patterned fleece is not as easy, due to the fuzzy nature of this fabric. Note that this fuzziness does not prevent reconstruction of features such as folds with our method.
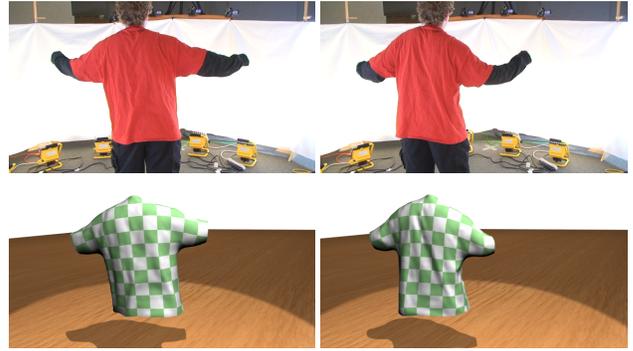

Figure 16: *Capture results for two frames of a large T-shirt. Input images (top) and reconstructed geometry (bottom).*
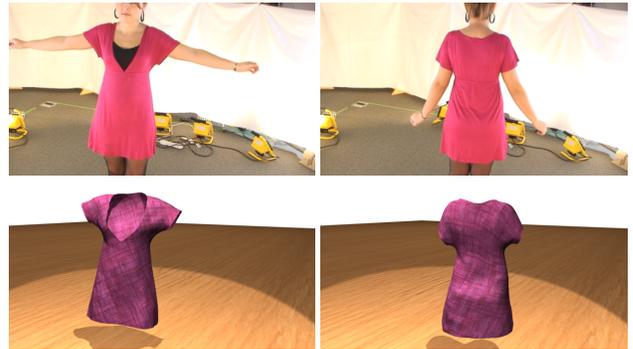

Figure 17: *Capture results for two frames of a pink dress. Input images (top) and reconstructed geometry (bottom).*

A different T-shirt is shown in Figure 16. The user assisted anchor placement was also used for this sequence. This garment is larger and more loosely fitting than the others, resulting in more ripples, which our capture method successfully reconstructs.

Two full-body garments were also reconstructed, demonstrating the versatility of our approach. A pink dress is shown in Figure 17, and a blue dress in Figure 15. For the pink dress, off-surface anchors were placed with the automatic boundary tracking method, while the user-assisted method was used for the blue dress.

Figure 18 shows another result of garments made from different fabric, in this case a long-sleeve nylon-shell down jacket. Anchor placement in this case was again partially manual.

Processing a sequence from captured video to the final result requires approximately one hour per frame, which can be easily parallelized. The majority of the time is spent in the multi-view reconstruction step, since our input data is in the form of high-definition video and we establish dense correspondences between images. A typical frame for the T-shirt data sets reconstructs over a million surface points, which are used to construct an initial triangulation with over 100K vertices. The very large resolution is required at this point in order to preserve high-frequency detail in the garments, particularly since the mesh reconstruction is non-uniform. The remeshed sequences typically have around 20K vertices.

9 Conclusion

In this paper, we have presented the first method for markerless capture of full garments and demonstrated its viability on a number of examples. Our acquisition and reconstruction methods generate per-frame meshes that capture the complex, changing structure of garments during human motion. Our parameterization and com-
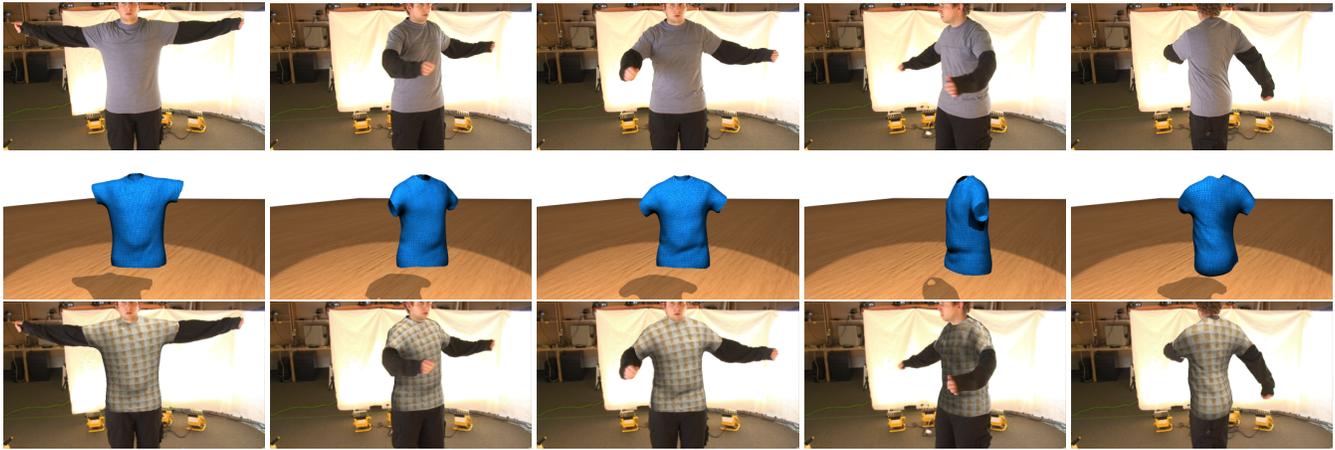
Figure 13: *Capture results for five frames of a T-shirt. From top to bottom: input images, captured geometry, T-shirt is replaced in the original images.*

Figure 14: *Capture results for a fleece vest. From left to right: one input frame, captured geometry, vest is replaced in the original image, the scene is augmented by adding a light source and making the vest more specular.*
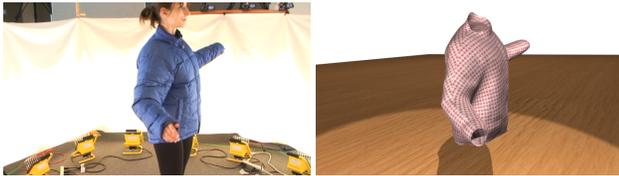

Figure 18: *Capture result for a long-sleeve nylon-shell down jacket*
.

pletion algorithms leverage fabric incompressibility and garment topology to consistently parameterize the frame sequences and generate realistic garment geometry and motion faithful to the input.

Our system has two main limitations. First, while our technique can handle significant occlusions, we do not handle situations where the garment surface comes in contact with itself, e.g. in case a sleeve touches the torso. The initial multiview reconstructions in these situations can contain incorrect surface topology, which is not handled in the subsequent process. For this reason, the actors were asked to keep their arms away from their torso. Second, the necessary spatial and temporal smoothing steps in order to produce noise-free results tend to remove the fine details in the geometry. As a result, the reconstructed garments appear to have fewer wrinkles than the original input images. Despite these limitations, we believe that our method greatly advances the current state-of-the-art in garment capture.

We used our method to reconstruct several different garments, with and without sleeves. In the future we would like to test it on more types of garments. For example, pants have only three boundary loops rather than the four boundaries of the examples in this paper. Such different types of garments will require a different structure for the base mesh. In the case of pants, the base mesh would be composed of two co-planar triangles, one for the front, and one for

the back. While we have not yet investigated such garments, we do not foresee any fundamental difficulties in processing them.

Another area of future work is the replacement of the boundary tracking method by a fully automatic method. We believe that it may be possible to use 2D image segmentation results to improve the reliability of the 3D anchor positions even in cases where the multiview stereo algorithm has not produced clean boundaries.

Acknowledgments

References

ALLEN, B., CURLESS, B., AND POPOVIĆ, Z. 2003. The space of human body shapes: reconstruction and parameterization from range scans. *ACM Trans. Graphics (Proc. SIGGRAPH)*, 587–594.

ANGUELOV, D., SRINIVASAN, P., KOLLER, D., THRUN, S., RODGERS, J., AND DAVIS, J. 2005. SCAPE: shape completion and animation of people. *ACM Trans. Graphics (Proc. SIGGRAPH)*, 408–416.

BHAT, K., TWIGG, C., HODGINS, J., KHOSLA, P., POPOVIC, Z., AND SEITZ., S. 2003. Estimating cloth simulation parameters from video. In *Proc. SCA*, 37–51.

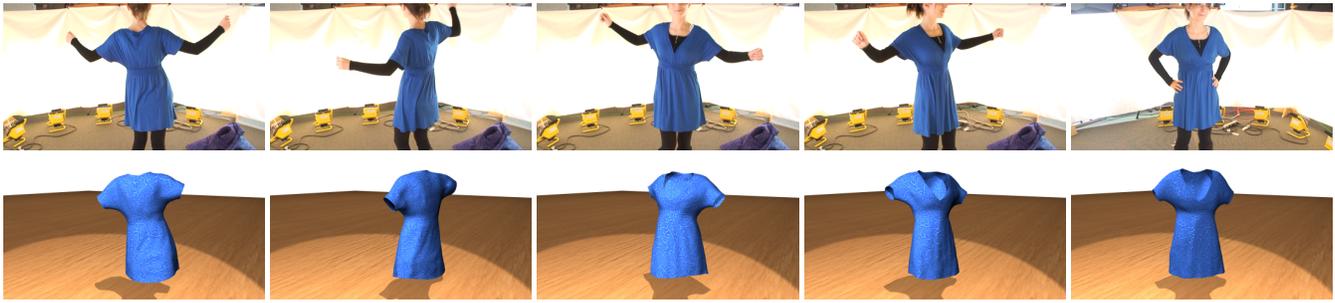BICKEL, B., BOTSCH, M., ANGST, R., MATUSIK, W., OTADUY, M., PFISTER, H., AND GROSS, M. 2007. Multi-scale capture

Figure 15: *Capture results for five frames of a blue dress. Input images (top) and reconstructed geometry (bottom).*

of facial geometry and motion. *ACM Trans. Graphics (Proc. SIGGRAPH)*, 33.

BOUBEKEUR, T., REUTER, P., AND SCHLICK, C. 2005. Visualization of point-based surfaces with locally reconstructed subdivision surfaces. In *Proc. Shape Modeling International*.

BRADLEY, D., BOUBEKEUR, T., AND HEIDRICH, W. 2008. Accurate multi-view reconstruction using robust binocular stereo and surface meshing. In *Proc. CVPR*.

CHEUNG, K., BAKER, S., AND KANADE, T. 2005. Shape-from-silhouette across time part ii: Applications to human modeling and markerless motion tracking. *IJCV 63*, 3, 225–245.

DE AGUIAR, E., THEOBALT, C., STOLL, C., AND SEIDEL, H.-P. 2007. Marker-less deformable mesh tracking for human shape and motion capture. In *Proc. CVPR*, 1–8.

FIALA, M., AND SHU, C. 2005. Fully automatic camera calibration using self-identifying calibration targets. Tech. Rep. NRC-48306 ERB-1130, National Research Council of Canada.

FISCHLER, M., AND BOLLES, R. 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM 24*, 6, 381–395.

FLOATER, M. 2003. Mean value coordinates. *Comput. Aided Geom. Des. 20*, 1, 19–27.

GOLDENTHAL, R., HARMON, D., FATTAL, R., BERCOVIER, M., AND GRINSPUN, E. 2007. Efficient simulation of inextensible cloth. *ACM Trans. Graphics (Proc. SIGGRAPH)*, 49.

GUSKOV, I., KLIBANOV, S., AND BRYANT, B. 2003. Trackable surfaces. In *Proc. SCA*, 251–257.

HASLER, N., ASBACH, M., ROSENHAHN, B., OHM, J.-R., AND SEIDEL, H.-P. 2006. Physically based tracking of cloth. In *Proc. Workshop on Vision, Modeling, and Visualization*, 49–56.

HERNANDEZ, C., VOGIATZIS, G., BROSTOW, G. J., STENGER, B., AND CIPOLLA, R. 2007. Non-rigid photometric stereo with colored lights. In *Proc. ICCV*.

KRAEVOY, V., AND SHEFFER, A. 2004. Cross-parameterization and compatible remeshing of 3d models. *ACM Trans. Graphics (Proc. SIGGRAPH)*, 861–869.

KRAEVOY, V., AND SHEFFER, A. 2005. Template-based mesh completion. In *Proc. SGP*, 13.

MITRA, N., FLÖRY, S., OVSJANIKOV, M., GELFAND, N., GUIBAS, L., AND POTTMANN, H. 2007. Dynamic geometry registration. In *Proc. SGP*, 173–182.

MORI, Y., AND IGARASHI, T. 2007. Plushie: an interactive design system for plush toys. *ACM Trans. Graphics (Proc. SIGGRAPH)*, 45.

MVIEW. http://vision.middlebury.edu/mview/.

PRAUN, E., SWELDENS, W., AND SCHRÖDER, P. 2001. Consistent mesh parameterizations. In *ACM Trans. Graphics (Proc. SIGGRAPH)*, 179–184.

PRITCHARD, D., AND HEIDRICH, W. 2003. Cloth motion capture. In *Proc. Eurographics*, 263–271.

ROSENHAHN, B., KERSTING, U., POWELL, K., KLETTE, R., KLETTE, G., AND SEIDEL, H.-P. 2007. A system for articulated tracking incorporating a clothing model. *Machine Vision and Applications 18*, 1, 25–40.

SCHOLZ, V., AND MAGNOR, M. A. 2004. Cloth motion from optical flow. In *Proc. Vision, Modeling and Visualization 2004*, 117–124.

SCHOLZ, V., STICH, T., KECKEISEN, M., WACKER, M., AND MAGNOR, M. 2005. Garment motion capture using color-coded patterns. In *Proc. Eurographics*, 439–448.

SCHREINER, J., ASIRVATHAM, A., PRAUN, E., AND HOPPE, H. 2004. Inter-surface mapping. *ACM Trans. Graphics (Proc. SIGGRAPH)*, 870–877.

SEITZ, S., CURLESS, B., DIEBEL, J., SCHARSTEIN, D., AND SZELISKI, R. 2006. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *Proc. CVPR*, 519–528.

SHEFFER, A., AND KRAEVOY, V. 2004. Pyramid coordinates for morphing and deformation. In *Proc. 3DPVT*, 68–75.

SHEWCHUK, J. 1996. Triangle: Engineering a 2D Quality Mesh Generator and Delaunay Triangulator. In *Applied Computational Geometry: Towards Geometric Engineering*, M. C. Lin and D. Manocha, Eds., vol. 1148 of *Lecture Notes in Computer Science*. Springer-Verlag, 203–222.

TOMASI, C., AND MANDUCHI, R. 1998. Bilateral filtering for gray and color images. In *Proc. ICCV*, 839.

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

WHITE, R., CRANE, K., AND FORSYTH, D. 2007. Capturing and animating occluded cloth. *ACM Trans. Graphics (Proc. SIGGRAPH)*, 34.

ZHANG, L., SNAVELY, N., CURLESS, B., AND SEITZ, S. M. 2004. Spacetime faces: high resolution capture for modeling and animation. *ACM Trans. Graphics (Proc. SIGGRAPH)*, 548–558.