Garment Modeling from a Single Image

Bin Zhou[1], Xiaowu Chen[†1], Qiang Fu[1], Kan Guo[1] and Ping Tan[2]

[1] State Key Laboratory of Virtual Reality Technology & Systems,
School of Computer Science & Engineering, Beihang University
[2] National University of Singapore

Figure 1: *Our method can model 3D garment shape from only a single image. From left to right: input image, 3D garments in the image view, 3D garments in another view.*

Abstract

Modeling of realistic garments is essential for online shopping and many other applications including virtual characters. Most of existing methods either require a multi-camera capture setup or a restricted mannequin pose. We address the garment modeling problem according to a single input image. We design an all-pose garment outline interpretation, and a shading-based detail modeling algorithm. Our method first estimates the mannequin pose and body shape from the input image. It further interprets the garment outline with an oriented facet decided according to the mannequin pose to generate the initial 3D garment model. Shape details such as folds and wrinkles are modeled by shape-from-shading techniques, to improve the realism of the garment model. Our method achieves similar result quality as prior methods from just a single image, significantly improving the flexibility of garment modeling.

Categories and Subject Descriptors (according to ACM CCS): I.3.5 [Computer Graphics]: Computational Geometry and Object Modeling—; I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—

1. Introduction

Reaslistic 3D garment models are required in various applications involing virtual characters. This work aims to build a simple and effective solution for garment modeling from just a single input image. Images carry abundant information and are used in various 3D modeling tasks before such as [WOQS05, TZW*07, XFT*08]. We seek to develop a user-friendly system with minimum input. With a single in-

† corresponding author: chen@buaa.edu.cn

put image, novice users can quickly obtain the desired garment model with some simple interactions.

Some prior techniques for garment modeling require a multiple-camera capture device, which is cumbersome to setup. Some sketch-based approaches, such as [TCH04, RMSC11], make restrictive assumptions about the mannequin pose and user input. The mannequin usually should be in a standing pose, and the user needs to mark out details such as cloth folds and wrinkles. In comparison, our method creates a realistic 3D garment model from a single image with a mannequin in general pose. We first develop a pose-aware garment outline interpretation method to model the large scale cloth shape based on the observation that garments are nearly piece-wise developable. We then exploit shading information to recover cloth shape details such as folds and wrinkles.

Given an input image, our method estimates the pose and shape of the mannequin from 2D joint positions and body contours. The garment outline is interpreted using an oriented facet decided by the mannequin pose to generate the initial garment. Unlike [RMSC11], our initialization works with general mannequin poses. This initialization generates a reasonable smooth 3D garment model, which does not have shape details and only captures the large scale (low frequency) shape characteristics. We then recover shape details by a shape-from-shading method [TS94] according to the shading information [SYJL11] in the input image. We transfer these details to the garment model by the weighted Laplacian method [SCOL*04]. This recovery of shape details is fully automatic, without any user interactions. The final results can capture the high-frequency folds and wrinkles, improving the visual realism of the recovered garment model.

Our main contributions are: 1) an effective method to create 3D garment models from a single image. To the best of our knowledge, 3D garment modeling from only a single image has not been addressed before. 2) a pose-aware garment outline interpretation method. This allows us to deal with images with general mannequin pose. 3) a shading-based details reconstruction method. This method achieves plausible results with only a single image, which significantly improves the effectiveness.

2. Related work

Some methods [GRH*12] create high quality garment models following the real-life design and tailoring process. The whole process includes 2D pattern creation, physically-based simulation, and iteratively optimization of invovled parameters to achieve the desired effects. The entire process is labor and computation intensive. It requires significant domain expertise, and is difficult to non-experts.

Sketch-based methods such as [TCH04, DJW*06, TWB*07] allow user to directly sketch on the mannequin figure to create garment models. Silhouettes and borders are marked out to model garment as an offset surface surrounding the mannequin. More recently, [RMSC11] proposed a context-aware sketch interpretation method, which improves results quality at loose regions. These methods generate good results and provide flexible user control with the sketch interface. We seek to use a real garment image to guide the user sketch to further improve the user experience. However, most of these prior sketch-based methods make restrictive assumptions about the mannequin pose and user sketches. Specifically, the mannequin usually should be in a standing pose with articulation largely limited to the 2D image plane. The users need to sketch out shape details such as folds and wrinkles. In a real image, the mannequin is often in general nature pose, and the clothes wrinkles are often too complicated to be sketched out manually. Our method focuses on solving these two problems to guide sketching with real images.

Markerless garment capture techniques [BPS*08] can capture 3D garment models with multiple synchronized cameras. These methods typically rely on multi-view stereo for 3D reconstruction. Temporal coherence, remeshing, and deformation techniques are used to produce the final result. Performance capture techniques [dAST*08, VBMP08, GSdA*09, SGdA*10] can also obtain approximate clothes of the captured actor, while they mainly focus on capturing the motion performance of actors. Both kinds of techniques require professional capture setup, and are computationally complicated.

Wrinkle details are important for the realism of garments [WHRO10, RPC*10, ZBO13]. [ATD*08, BHV*11] use active lighting to estimate wrinkles by a photometric stereo algorithm. However, the involved hardware setup, such as multi-spectral camera or calibrated light sources, is also complicated for amateur users. [PZB*09] estimate shape details from image edges of a multi-view garment video footage by space-time deformation. In our case, we seek to recover these details from a single input image, where both prior approaches cannot be applied.

3. Our Method

Figure 2 shows the main pipeline of our method. Given an input image, our method first estimates 3D pose and shape of the mannequin using 2D joints and body contours. Then, we construct an oriented facet for each mannequin bone according to its angles, and project the 2D garment outlines into corresponding facet. Using the projected outlines, we obtain an initial 3D garment model (Section 3.1). To enrich the initial model, we use the shading of the image to recovery shape details such as folds and wrinkles, and add these details to the initial garment (Section 3.2). The final garment shape is plausible and matches the input image well.

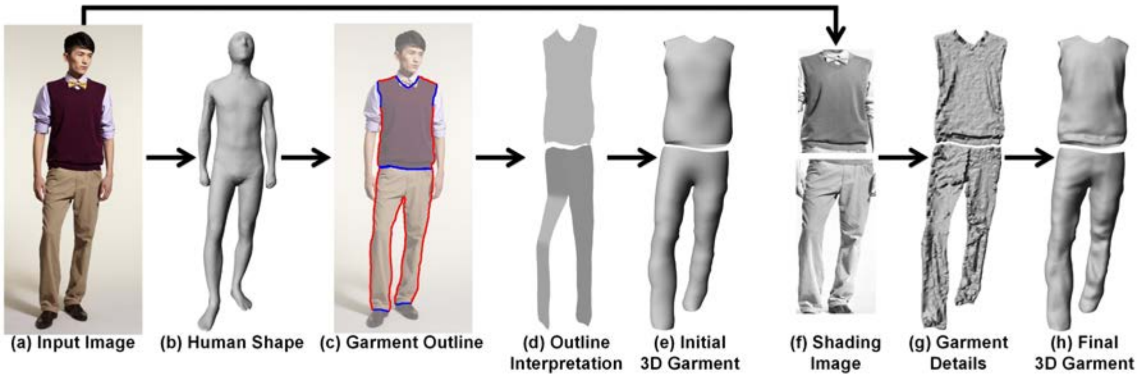

Figure 2: *Pipeline of our method.*

3.1. Garment Initialization

Our method firstly utilizes [ZFL*10, CGZZ] to estimate mannequin 3D pose and shape (Figure 2(b)) from the input image. The 3D pose is recovered by a semi-automatic pose estimation method [Tay00] using the user-specified 2D joints. The recovered 3D orientations and rotations of skeletal bones can be interactively refined by users. With the user-specified 2D body contours, the 3D shape is recovered by using the method similar to [GWBB09]. See [ZFL*10] for more details.

With the estimated 3D mannequin shape and the user-specified garment outlines, our method annotates the outlines as silhouettes and boundaries. The outlines are classified as silhouettes if they do not cross the mannequin and as boundaries if they do. As shown in Figure 2(c), red denotes the silhouettes, and blue denotes the boundaries. Meanwhile, we detect the related mannequin bones $\{l_1, \cdots, l_i, \cdots, l_n\}$ according to the overlay between the closed garment region and the projected mannequin.

We define an oriented facet F_{l_i} for each related bone l_i :

$$F_{l_i}: \quad \vec{L_J^{l_i}L} \cdot \vec{n}_{l_i} = 0, \qquad \vec{n}_{l_i} = R_{l_i} * \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}^T \tag{1}$$

where L is one point in the oriented facet F_{l_i}, $L_J^{l_i}$ is the 3D joints for bone l_i recovered by the semi-automatic pose estimation method [Tay00], and n_{l_i} is the normal of the the oriented facet F_{l_i}. R_{l_i} is the 3-by-3 rotation matrix of bone l_i calculated by using the absolute angles of the recovery pose. The normal n_{l_i} is decided by the component of rotation matrix R_{l_i} orthogonal to image plane. Two oriented facets (F_1 and F_2) are demonstrated in Figure 3(a).

Then, we calculate the line of intersection between two adjacent oriented facets, and project the line into the image plane. According to Eq. 1, the intersection line must go through the 3D joint corresponding to the two oriented facets. Thus, the projected line must go through the corresponding 2D joint, as demonstrated in Figure 3(a). Using the projected line and the 2D joint, we can obtain an interior line inside the garment region in the image plane, as demonstrated in Figure 3(a).

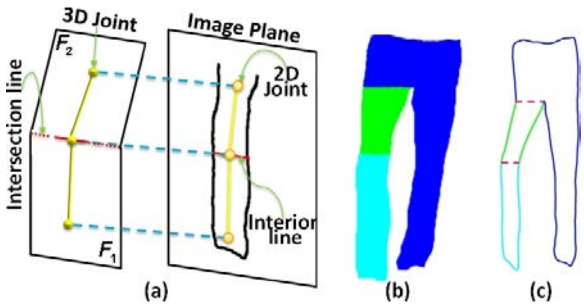

Figure 3: *Demonstration of garment outline interpretation.*

The interior lines divided the garment region, surrounded by the user-specified garment outlines, into different parts. As shown in Figure 3(b), different colors denote different divided parts, and each part is mapped to one oriented facet. For each part, we project the garment outlines and the interior lines into the mapped oriented facet. After that, we obtain an interpreted garment region in each oriented facet, as demonstrated in Figure 3(c). Each interpreted garment region consists of projected silhouettes and boundaries, as well as the projected interior lines.

We triangulate each garment region in the oriented facet, creating multiply 3D meshes. The 3D meshes are placed together using the sampled points in the projected interior lines as the common vertices. We assume the garment is symmetric in front and back to model in the invisible back part. Thus, we duplicate the 3D meshes with one for the front surface and the other for the back surface, expect for the vertices in the garment silhouettes. We take the duplicated 3D mesh as our garment outline interpretation, as shown in Figure 2(d).

We use a similar method as [RMSC11] to update the garment shape from the given outline interpretation, as shown in Figure 2(e). However, unlike [RMSC11], our method does not assume a simple standing mannequin. Our method makes use of the joint angles, and is pose-aware. This allows us to deal with more general data.

3.2. Shading based Garment Detail Recovery

The initialization generates a reasonable smooth 3D garment model, which does not have shape details and only captures the large scale (low frequency) shape characteristics. To enrich it, we use the shading of the image to recovery the depth map for the initial garment, and reintroduce the details from the initial garment to the updated garment shape. We make an assumption that the garment surface is approximated by Lambertian reflectance.

Firstly, we apply the intrinsic image decomposition algorithm [SYJL11] to the garment region of the input image, as shown in Figure 2(f). The algorithm decomposes the input image I into the product of a shading image S and a reflectance image R: $I_p = S_p R_p$, where p denotes a point in the image space. The shading image S consists of various lighting effects that include shadows and specular highlights in addition to shading. The reflectance image R contains the intrinsic color, or albedo, of surface points independent of the illumination environment. The garment region of the input image satisfies the assumption of intrinsic image decomposition well. The decomposed shading image S can benefit depth map recovery algorithm.

We recovery the depth map from the decomposed shading image S of the garment region using Shape-from-Shading (SfS) algorithm [TS94]. According to the correspondences between initial 3D garment and the input image, we assign the depth value to the initial 3D garment, forming an detailed initial 3D garment, as shown in Figure 2(g). The detailed initial 3D garment is denoted as S, while the initial 3D garment is denoted as $\tilde{S}$.

The surface $\tilde{S}$ can be regard as a low-frequency surface associated with S. We define the garment details to refer to the high-frequency surface S details. More precisely, the garment details are defined as the difference between S and $\tilde{S}$. We encode the garment details based on the Laplacian coordinates [SCOL*04]. Let δ_i and $\tilde{\delta}_i$ be the Laplacian coordinates of the vertex i in S and $\tilde{S}$, respectively. We define ξ_i to be the encoding of the garment details at vertex i defined by $\xi_i = \delta_i - \tilde{\delta}_i$.

In order to enrich the deformed garment shape, we need to transfer the garment details onto the deformed garment surface, denoted as U. We apply a weighted Laplacian formulation [SCOL*04] to solve this problem:

$$\arg\min_{V'} \sum_i ||L(v_i') - L(v_i) - w_i \xi_i||^2 \tag{2}$$

where $L(v_i)$ is the Laplacian coordinates of the vertex i in U, $L(v_i')$ is the Laplacian coordinates of the vertex i in the detailed garment, and is unknown. The weight $w_i = e^{-\frac{\mu}{d}}, \mu = 0.5$, and d is the distance from the vertex i to the closest silhouette vertex. The usage of weight w_i can eliminate the influence of noise in the silhouette vertex of garment detail.

After solving the linear system, we can obtain the final garment shape, as shown in Figure 2(h). The final garment shape is believable and with more details, such as folds and wrinkles.

4. Experimental Results

We evaluate our method on a variety of different images shown in the paper and the accompanying video. Our method can obtain realistic 3D garment models from only a single image, even for images with general character poses.

Figure 4 shows the comparison of our method and a recent garment capture technique [BPS*08], which is based on multi-view reconstruction. Four frames are selected from the public 'blue shirt' dataset. Our method models them respectively to create the result garment models. The character in these frames has different poses, and his arm is reaching out of the image plane in three frames. The results of [BPS*08] are generated from all the 16 videos as provided by the author. Our method produces convincing garment models with just a single image. Our results match well with the garment outline in the input image. That is because our method directly uses the garment outline as cue to model and morph the 3D garment shape. The position of garment outline in the input image serves as a strong constraint for our method. Our results contain more vivid wrinkle details than the results of [BPS*08] beacuse of the shading based garment detail recovery.

In Figure 5, we compare our method with [PZB*09], which estimates folds and wrinkles from image edges in multi-view garment video footages. We test out method on two frames from the public 'blue dress' dataset. Our method produces similar realistic-looking garment shapes as [PZB*09]. This proves the effectiveness of our method in detail modeling. Note that our method needs only a single image, while [PZB*09] requires a multi-view setup.

More results are provided in Figure 6. Our method can model a variety of different garments, including dresses, skirts, shirts, and pants, for female or male. The user interactions of our method are to specify the 2D joints positions and human body contours in the input image, and the 2D garment outlines, as well as to adjust the recovered 3D pose. These interactions are simple and intuitive, and do not require any domain knowledge in garment design. Our method is implemented using Matlab (single threaded) without special optimization such as GPU acceleration. All the experiments are done with a 32 bit desktop machine with a 3.0 GHz Intel Core (TM)2 Duo processor, 3.0 GB of memory.

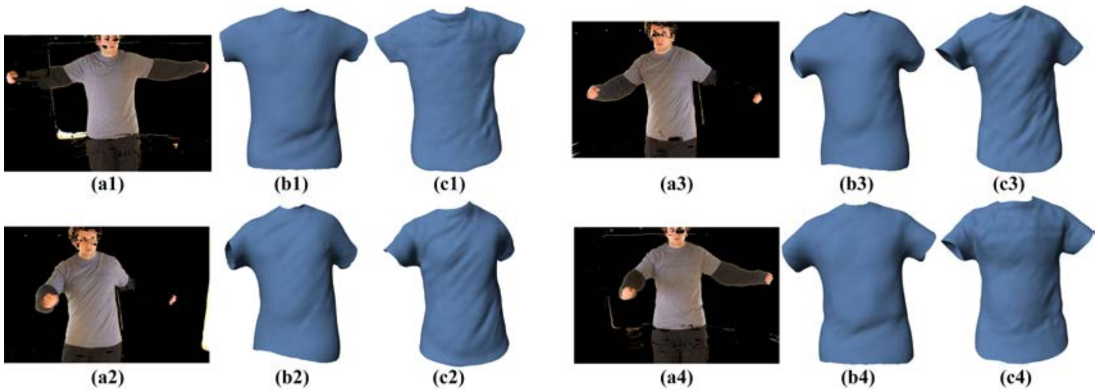

Figure 4: *Comparisons with garment capture method [BPS*08]. (a1)(a2)(a3)(a4) are the input images chosen from the public 'blue shirt' dataset.(a1): frame #57 of camera #9 . (a2): frame #97 of camera #12 . (a3): frame #114 of camera #12 . (a4): frame #245 of camera #9 . (b1)(b2)(b3)(b4) are the results of [BPS*08]. (c1)(c2)(c3)(c4) are the results of our method.*

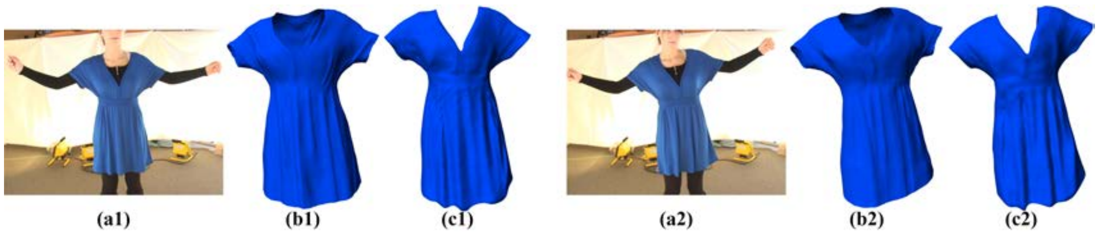

Figure 5: *Comparisons with [PZB*09]. (a1)(a2) are the input images chosen from the public 'blue dress' dataset. (a1): frame #36 of camera #9 . (a2): frame #270 of camera #9 . (b1)(b2) are the results of [PZB*09]. (c1)(c2) are the results of our method.*

The user interaction usually takes three to five minutes. For typical garments with five thousand triangles, our method takes one to two minutes to generate the results except user interactions.

Our recovered 3D garment model can be animated by physical-based simulation. We generate some animation results in Figure 7. Our estimated 3D character body shape is based on the low-dimensional parametric SCAPE model. Given a motion sequence, we generate an animated 3D human body shape by only changing the pose parameters and keeping the body shape parameters. For each frame in the sequence, we use Euler simulator as our physically-based simulators, taking the modeled 3D garment as template mesh and the results of previous frame as initialization. Please notice that we take the smooth 3D garment as the template mesh in animation. All the wrinkles are generated by simulation.

Discussion. We make an assumption that the garment is symmetric in front and back. Figure 1, 4 and 6(2nd to 4th rows) show the results when the garment is symmetric in front and back. Our method works well when the garments satisfy the symmetry assumption by using just a single image. Figure 5 and 6(1st rows) show the results when the garment is not symmetric in front and back. As shown in Figure 5, the difference in the backside of the garment is due to our assumption. According the symmetry assumption, we consider that the garment outline in back is same to that in front. Our method recovers the front side of the garment well (Figure 5(a1)(a2)). While in contrast to the results of [PZB*09](Figure 5(b1)(b2)), our method can not recovery the backside of the garment because a single image is inherently insufficient to cover the whole garment surface.

5. Conclusion and Limitations

We propose a method for modeling realistic 3D garments from a single image. Our method allows general pose for the image character, and recovers cloth shape details automatically from shading information. We show that our method achieves similar results as state-of-the-art garment capture methods with only a single image, significantly improving the effectiveness and flexibility.

Our method has some limitations, which point out the direction of future study. We assume almost all the garment outlines can be seen or labeled in the input image. We do not deal with occlusions cause by the character. This problem might be addressed by a more sophisticated garment outline

Figure 6: *More results of our method. For each results, from left to right: the input image, the recovered 3D garment in image view, the recovered 3D garment in another view.*

prediction according to the symmetry of the body and garment. Furthermore, we assume the garment surface is Lambertian, and the illumination is directional. The Lambertian assumption works well for many common fabrics including cotton and wool. But it tends to fail on silk, satin, polyester, etc. Otherwise, for garments with complicated texture or lots of heterogeneity (such as different colored regions), the intrinsic image decomposition tends to fail. Though there are interactive solutions like [BPD09], it requires further user input.

Acknowledgements

We would like to thank the anonymous reviewers for their help in improving the paper. This work was partially supported by NSFC (60933006), 863 Program (2012AA011504). Ping Tan was supported by the Singpoare A*STAR PSF Project R-263-000-698-305.

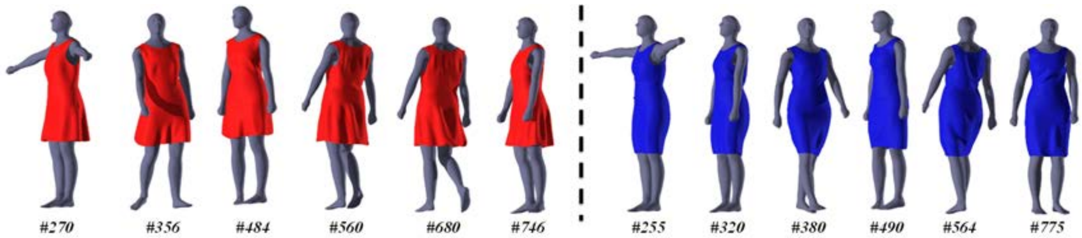

Figure 7: *Left: Animate the estimated red loose-fitting dress in Figure 6 (1st result in 2nd row). Right: Animate the estimated brown tight-fitting dress in Figure 6 (2nd result in 4th row). The frame numbers are list below each animated result.*

References

[ATD*08] AHMED N., THEOBALT C., DOBREV P., PETER SEIDEL H., THRUN S.: Robust fusion of dynamic shape and normal capture for high-quality reconstruction of time-varying geometry. In *Proc. of CVPR* (2008). 2

[BHV*11] BROSTOW G. J., HERNÁNDEZ C., VOGIATZIS G., STENGER B., CIPOLLA R.: Video normals from colored lights. *TPAMI 33*, 10 (2011), 2104–2114. 2

[BPD09] BOUSSEAU A., PARIS S., DURAND F.: User-assisted intrinsic images. *ACM Trans. Graph. (Proc. of Siggraph Asia) 28*, 5 (2009), 130:1–130:10. 6

[BPS*08] BRADLEY D., POPA T., SHEFFER A., HEIDRICH W., BOUBEKEUR T.: Markerless garment capture. *ACM Trans. Graph. (Proc. of Siggraph) 27*, 3 (2008), 99:1–99:9. 2, 4, 5

[CGZZ] CHEN X., GUO Y., ZHOU B., ZHAO Q.: Deformable model for estimating clothed and naked human shapes from a single image. *The Visual Computer*. doi:10.1007/s00371-013-0775-7. 3

[dAST*08] DE AGUIAR E., STOLL C., THEOBALT C., AHMED N., SEIDEL H.-P., THRUN S.: Performance capture from sparse multi-view video. *ACM Trans. Graph. (Proc. of Siggraph) 27*, 3 (2008), 98:1–98:10. 2

[DJW*06] DECAUDIN P., JULIUS D., WITHER J., BOISSIEUX L., SHEFFER A., CANI M.-P.: Virtual garments: A fully geometric approach for clothing design. *Comput. Graph. Forum (Proc. of Eurographics) 25*, 3 (2006), 625–634. 2

[GRH*12] GUAN P., REISS L., HIRSHBERG D. A., WEISS A., BLACK M. J.: Drape: Dressing any person. *ACM Trans. Graph. (Proc. of Siggraph) 31*, 4 (2012), 35:1–35:10. 2

[GSdA*09] GALL J., STOLL C., DE AGUIAR E., THEOBALT C., ROSENHAHN B., SEIDEL H.-P.: Motion capture using joint skeleton tracking and surface estimation. In *Proc. of CVPR* (2009), pp. 1746–1753. 2

[GWBB09] GUAN P., WEISS A., BĂLAN A. O., BLACK M. J.: Estimating human shape and pose from a single image. In *Proc. of ICCV* (2009), pp. 1381–1388. 3

[PZB*09] POPA T., ZHOU Q., BRADLEY D., KRAEVOY V., FU H., SHEFFER A., HEIDRICH W.: Wrinkling captured garments using space-time data-driven deformation. *Comput. Graph. Forum (Proc. of Eurographics) 28*, 2 (2009), 427–435. 2, 4, 5

[RMSC11] ROBSON C., MAHARIK R., SHEFFER A., CARR N.: Context-aware garment modeling from sketches. *Comput. and Graph. (Proc. of SMI)* (2011), 604–613. 2, 4

[RPC*10] ROHMER D., POPA T., CANI M.-P., HAHMANN S., SHEFFER A.: Animation wrinkling: Augmenting coarse cloth simulations with realistic-looking wrinkles. *ACM Trans. Graph. (Proc. of Siggraph Asia) 29*, 6 (2010), 157:1–157:8. 2

[SCOL*04] SORKINE O., COHEN-OR D., LIPMAN Y., ALEXA M., RÖSSL C., SEIDEL H.-P.: Laplacian surface editing. In *Proc. of SGP* (2004), pp. 179–188. 2, 4

[SGdA*10] STOLL C., GALL J., DE AGUIAR E., THRUN S., THEOBALT C.: Video-based reconstruction of animatable human characters. *ACM Trans. Graph. (Proc. of Siggraph Asia) 29*, 6 (2010), 139:1–139:10. 2

[SYJL11] SHEN J., YANG X., JIA Y., LI X.: Intrinsic images using optimization. In *Proc. of CVPR* (2011), pp. 3481–3487. 2, 4

[Tay00] TAYLOR C. J.: Reconstruction of articulated objects from point correspondences in a single uncalibrated image. *CVIU 80*, 10 (2000), 349–363. 3

[TCH04] TURQUIN E., CANI M.-P., HUGHES J. F.: Sketching garments for virtual characters. In *Proc. of SBIM* (2004), pp. 175–182. 2

[TS94] TSAI P.-S., SHAH M.: Shape from shading using linear approximation. *Image Vision Comput. 12*, 8 (1994), 487–498. 2, 4

[TWB*07] TURQUIN E., WITHER J., BOISSIEUX L., CANI M.-P., HUGHES J.: A sketch-based interface for clothing virtual characters. *Comput. Graph. Appl. 27*, 1 (2007), 72–81. 2

[TZW*07] TAN P., ZENG G., WANG J., KANG S. B., QUAN L.: Image-based tree modeling. *ACM Trans. Graph. (Proc. of Siggraph) 26*, 3 (2007), 87:1–87:7. 1

[VBMP08] VLASIC D., BARAN I., MATUSIK W., POPOVIĆ J.: Articulated mesh animation from multi-view silhouettes. *ACM Trans. Graph. (Proc. of Siggraph) 27*, 3 (2008), 97:1–97:9. 2

[WHRO10] WANG H., HECHT F., RAMAMOORTHI R., O'BRIEN J. F.: Example-based wrinkle synthesis for clothing animation. *ACM Trans. Graph. (Proc. of Siggraph) 29*, 4 (2010), 107:1–107:10. 2

[WOQS05] WEI Y., OFEK E., QUAN L., SHUM H.-Y.: Modeling hair from multiple views. *ACM Trans. Graph. (Proc. of Siggraph) 24*, 3 (2005), 816–820. 1

[XFT*08] XIAO J., FANG T., TAN P., ZHAO P., OFEK E., QUAN L.: Image-based façade modeling. *ACM Trans. Graph. (Proc. of Siggraph Asia) 27*, 5 (2008), 161:1–161:10. 1

[ZBO13] ZURDO J. S., BRITO J. P., OTADUY M. A.: Animating wrinkles by example on non-skinned cloth. *IEEE Trans. Vis. Comput. Graph. 19*, 1 (2013), 149–158. 2

[ZFL*10] ZHOU S., FU H., LIU L., COHEN-OR D., HAN X.: Parametric reshaping of human bodies in images. *ACM Trans. Graph. (Proc. of Siggraph) 29* (2010), 126:1–126:10. 3