HAL
open science

ClothCap: Seamless 4D Clothing Capture and Retargeting

Gerard Pons-Moll, Sergi Pujades, Sonny Hu, Michael J Black

To cite this version:

Gerard Pons-Moll, Sergi Pujades, Sonny Hu, Michael J Black. ClothCap: Seamless 4D Clothing Capture and Retargeting. ACM Transactions on Graphics, 2017, 36 (4), pp.1-15. 10.1145/3072959.3073711 . hal-02162166

HAL Id: hal-02162166

https://inria.hal.science/hal-02162166v1

Submitted on 21 Jun 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ClothCap: Seamless 4D Clothing Capture and Retargeting

GERARD PONS-MOLL*, Max Planck Institute for Intelligent Systems, Tübingen, Germany
SERGI PUJADES*, Max Planck Institute for Intelligent Systems, Tübingen, Germany
SONNY HU, Body Labs, New York, NY, USA
MICHAEL J. BLACK, Max Planck Institute for Intelligent Systems, Tübingen, Germany

Fig. 1. **ClothCap.** From left to right: (1) An example 3D textured scan that is part of a 4D sequence. (2) Our multi-part aligned mesh model, layered over the body. (3) The estimated *minimally clothed shape* (MCS) under the clothing. (4) The body made fatter and dressed in the same clothing. Note that the clothing adapts in a natural way to the new body shape. (5) This new body shape posed in a new, never seen, pose. This illustrates how ClothCap supports a range of applications related to clothing capture, modeling, retargeting, reposing, and try-on.

Designing and simulating realistic clothing is challenging. Previous methods addressing the capture of clothing from 3D scans have been limited to single garments and simple motions, lack detail, or require specialized texture patterns. Here we address the problem of capturing regular clothing on fully dressed people in motion. People typically wear multiple pieces of clothing at a time. To estimate the shape of such clothing, track it over time, and render it believably, each garment must be segmented from the others and the body. Our *ClothCap* approach uses a new multi-part 3D model of clothed bodies, automatically segments each piece of clothing, estimates the minimally clothed body shape and pose under the clothing, and tracks the 3D deformations of the clothing over time. We estimate the garments and their motion from 4D scans; that is, high-resolution 3D scans of the subject in motion at 60 fps. ClothCap is able to capture a clothed person in motion, extract their clothing, and retarget the clothing to new body shapes; this provides a step towards virtual try-on.

CCS Concepts: • **Computing methodologies** → **Computer graphics**; **Shape modeling**; *Animation*; *Mesh models*;

Additional Key Words and Phrases: Clothing, animation, 3D shape, 3D segmentation, try-on, performance capture.

ACM Reference format:
Gerard Pons-Moll, Sergi Pujades, Sonny Hu, and Michael J. Black. 2017. ClothCap: Seamless 4D Clothing Capture and Retargeting. *ACM Trans. Graph.* 36, 4, Article 73 (July 2017), 15 pages.
DOI: http://dx.doi.org/10.1145/3072959.3073711

*Equal contribution

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
© 2017 Copyright held by the owner/author(s). 0730-0301/2017/7-ART73 $15.00
DOI: http://dx.doi.org/10.1145/3072959.3073711

1 INTRODUCTION

Dressing virtual avatars and animating them with high quality, visually plausible, results is a challenging task. Highly realistic physical simulation of clothing on human bodies in motion is complex: clothing models are laborious to construct, patterns must be graded so that they can be sized to different characters, and the physical parameters of the cloth must be known. Instead, we propose a data-driven clothing capture (ClothCap) approach; we capture dynamic clothing on humans from 4D scans and transform it to more easily dress virtual avatars.

We proceed by capturing the garment geometry in motion on a body, estimate the body shape and pose under clothing, and segment and extract the clothing pieces. We then retarget the captured clothing to new body shapes and poses. To that end, we develop a novel multi-part mesh model and show how to segment, track, and recover garment shape from sequences of 3D scans (see Fig. 1).

Previous methods for 3D garment capture are not sufficiently accurate or detailed to compete with physical simulation. Existing capture methods suffer from low resolution, static shapes, simple body motions, capture only one clothing piece, or do not segment the clothing from the body. The key problems to solve include high-quality capture, segmentation, tracking of surface shape, as well as body shape and pose estimation.

We address these problems by introducing a novel multi-mesh representation that we fit to sequences of 3D scans captured with a high-resolution 4D scanner (Fig. 1 (1)). Our method requires as input a point cloud with texture, here we used the same active stereo system as in (Pons-Moll et al. 2015a). The multi-mesh representation is consistent with how clothing is worn, modeling the changing visibility of surfaces over time. It further allows us to address the

segmentation and tracking problems in a coherent framework. In doing so we *automatically* segment clothing on the 4D scans using both appearance and shape information (Fig. 1 (2)) using a Markov random field. Given segmented garments, we fit a set of multi-part template meshes to the scans, putting them into correspondence across time (Fig. 1 (2)) and estimate the underlying *minimally clothed shape* (MCS) (Fig. 1 (3)). Rendering the segmented clothing pieces on top of the MCS produces plausible clothing geometry with the appropriate layered appearance. Segmentation of the clothing is critical for realism in many applications.

Our philosophy shifts the focus from garment simulation to garment capture. There are many applications for this. For example, for clothing shopping, the physical garment already exists. Traditional virtual try-on involves getting the pattern from the manufacturer (which can be hard) and performing simulation on different bodies. Since the garment exists, our approach is to scan a person wearing the garment and then generalize this to new bodies.

The automatically extracted clothing meshes correspond to garments and are naturally associated with an articulated body model underneath (SMPL (Loper et al. 2015)). This association enables us to change the shape of the body (Fig. 1 (4)), or transfer the clothing to a new body automatically. While the acquired wrinkle pattern may not be physically realistic, the visual appearance is sufficient for many virtual try-on applications. Given our captured garments, we can plausibly dress any real body shape. Furthermore, using the pose blend shapes of the underlying body model (SMPL (Loper et al. 2015)) we can repose the garments (Fig. 1 (5)). Again, while the wrinkles may not vary with pose in a physically realistic way, the results are sufficient for many applications. Additionally, the extracted clothing provides a foundation for future work on modeling clothing dynamics.

In summary, ClothCap lays the foundation for garment modeling from data by addressing the challenging problems of garment segmentation and tracking from 4D scans. Precisely we contribute: a) an automatic method to segment 3D scan sequences leveraging a body model (Sec. 5.2), b) a multi-mesh template tracking method (Sec. 5.3) and c) a technique to retarget dynamic cloth to a new body shape (Sec. 5.4). We demonstrate the full method by extracting several types of clothing including long pants, t-shirts, and shorts on a variety of people performing complex motions. We also show the method working for skirts, which have a different topology than the body, but this requires an additional manual step to establish the rough garment topology. We illustrate the use for virtual try-on by dressing new people, including subjects from the CAESAR dataset (Robinette et al. 2002).

2 RELATED WORK

Cloth simulation. There is an extensive literature on, and many commercial solutions for, clothing simulation. At its best, clothing simulation can be photo realistic but these high-quality results come at great expense in both labor and computation.

Much of the recent work in the field is focused on how to make simulation more computationally efficient (Goldenthal et al. 2007), particularly by adding realistic wrinkles to low-resolution simulation (Gillette et al. 2015; Kavan et al. 2011; Kim et al. 2013; Wang et al. 2010). Other approaches have focused on taking off-line simulations and "compiling" them into efficient approximate methods appropriate for real-time rendering in video games (de Aguiar et al. 2010; Guan et al. 2012; Kim et al. 2013).

Rogge et al. (2014) match move a 3D body model to a monocular video sequence, simulate clothing on the body, and then project the simulated garment back into the video. The method involves significant manual intervention at several steps.

Capturing cloth parameters. Accurate clothing simulation requires realistic physical parameters. Several authors measure cloth parameters and then use these in traditional simulation (Miguel et al. 2012; Wang et al. 2011). Recent work explores getting such parameters from perceptual judgements of humans (Sigal et al. 2015).

Several methods attempt to extract physical cloth parameters from video sequences of simple cloth pieces (Bhat et al. 2003; Bouman et al. 2013). Bhat et al. (2003) then show clothing animated with these parameters. These methods do not look at clothing moving on, and interacting with, the body.

Rosenhahn et al. (2007) formulate the problem of tracking a person and their clothing in an integrated way. Assuming a known piece of clothing they track the person and clothing while estimating a few physical cloth parameters.

The most relevant here is the work of Stoll et al. (2010). Using a multi-camera system they capture coarse 3D meshes of dressed bodies in motion. They segment the clothing into regions of rigid and non-rigid motion, fit the articulated motion of the body underneath, and estimate the physical parameters for the non-rigid clothing region. For new motions, they simulate the non-rigid region at low resolution and then deform the original high-resolution mesh using the low-resolution simulation. The results look appealing but are only visualized from the view of the capture cameras; issues of occlusion and layering are not illustrated. While their work is the similar to ours, it differs in several ways. Most importantly they do not segment garments from each other and the body, meaning that, the clothes cannot be removed and transferred to new bodies. Their segmentation only attempts to find highly non-rigid regions for the purpose of simulating their deformations. In ClothCap we focus on segmentation of the physical garments and retargeting to new body shapes and poses.

Transfer to new body shapes. Another problem for simulation approaches is that sizing (grading) garments to new characters remains labor intensive and requires expertise in garment design. A few methods have addressed the transfer of hand-designed and simulated garments to new body shapes (Brouet et al. 2012; Guan et al. 2012). For example, DRAPE (Guan et al. 2012) uses simulated garments on many subjects in many poses to learn an approximate model of garments that can be automatically applied to new body shapes and motions. The results, however, are oversmoothed and lack the detail of realistic clothing or high-quality simulation.

There are also many systems that address virtual try-on. Many address simple texture overlay (Hilsmann and Eisert 2009) or overlay of garments in relatively static poses (Sekine et al. 2014). None of the above address 4D clothing capture.

Shape under clothing. Putting garments on new body shapes often first involves removing the clothing from scans of people. Like Stoll et al. (2010), Hasler et al. (2009) estimate the shape and pose of a body under clothing but do so only from a single static 3D scan. They do not extract the clothing or retarget it to new bodies. Balan and Black (2008) estimate 3D body shape under clothing using several poses but do so from 2D image data.

Neophytou and Hilton (2014) take sequences of 3D scans in correspondence and estimate an underlying body shape and pose as well as a deformation from this shape; they describe the shape, pose, and the clothing deviation as being three "layers." Given a sequence they learn a model of how clothing deviates from the body and use this to cloth new body shapes. Like Neophytou and Hilton, we think of the minimally clothed shape as a layer, but unlike them, we segment the clothing into distinct meshes and have multiple clothing pieces that can overlap. This improves realism.

Other recent work estimates parametric body shape under clothing from multiple scans (Wuhrer et al. 2014) or scan sequences (Yang et al. 2016). These methods lack detail and accuracy. Recent work solves not only for a parametric model but allows constrained deviations from this (Zhang et al. 2017). This produces significantly more accurate and realistic results.

Performance capture. Surface capture methods capture time varying geometry. There is a large literature in this area but most methods treat the body and clothing as a single deforming mesh. Vlasic et al. (2008) propose a multi-pass approach that involves fitting a skeleton to visual hulls, deforming the template according to the skeleton, and finally deforming the template to fit image silhouettes. In related work, Gall et al. (2009) require an initial model from a laser scan and manually insert a skeleton. De Aguiar et al. (2008) forego the skeleton and do performance capture of the full body in clothing from multiple cameras. They require a laser scan of the person in the clothing and then deform it. Although impressive results are achieved, most of the detail is derived from the initial laser scan. Other works also use a free-form surface (Collet et al. 2015; Innmann et al. 2016; Newcombe et al. 2015; Wang et al. 2016) but can not separate the clothing from the person and can not retarget it. To make tracking stable one option is to learn the correspondences between frames (Dou et al. 2016) or to a common template (Pons-Moll et al. 2015b). Wu et al. (2012) capture fine wrinkle details using shading, obviating the need for a laser scan. Tejera et al. (2013) capture 4D data (3D meshes over time) and align the surfaces in time (SurfCap). For a subject they build a model of their space of deformations. Given new mocap data they animate the person. They do not retarget clothing to new bodies and the meshes are of fairly low spatial resolution, so that wrinkles are not prominent. Similarly, Huang et al. (2015) leverage captured 4D data and corresponding video to retrieve new animations using a skeleton to query the database. Novel view textures are rendered using the approach of Casas et al. (2014). While this approach interpolates motion for a captured subject, it does not allow the body shape to be easily changed. Robertini et al. (2014) use a novel photo-consistency objective function to add fine-scale cloth details into a deforming mesh captured with multiple video cameras. The results are quite smooth and the clothing is not segmented.

There is also work on 4D capture and modeling of soft-tissue deformations for bodies in limited clothing (Bogo et al. 2017; Loper et al. 2015; Pons-Moll et al. 2015a). Clothing is more complex to capture and model because of the multiple pieces, layering, changing occlusion relationships, material properties, wrinkles and folds.

Garment capture. As an alternative to simulation, several methods attempt to extract garments directly from 3D or 4D data. Bradley et al. (2008) capture single garments worn by a person performing slow careful movements. The method requires manual intervention and does not capture multiple garments and the body pose. The capture system results in relatively low wrinkle detail but provides a first proof of concept. Popa et al. (2009) use image edges to estimate where wrinkles are and then put them back into the 3D mesh. We go beyond their work to capture the full body, multiple garments, the body shape underneath, and show how to retarget garments.

White et al. (2007) use a custom pattern printed on garments to make video-based 3D capture easier, effectively providing dense markers. We do not require a special texture and capture multiple garments at once. They show manual editing of the captured garments but not transfer to new shapes. They show animation of pants from motion capture but rig the recovered model manually.

Zhou et al. (2013) capture garment shape from a single image using manual intervention and shape-from-shading to get fine wrinkle detail. They also compute the shape and pose of the body and show animation of captured garments on new bodies.

Chen et al. (2015) describe a system for capturing and modeling garments using a Kinect-based scan of a person. The system uses shape and appearance to parse (segment) a mesh into clothing pieces. They then find similar parts in a database of synthetic clothing parts and stitch a 3D garment together from the parts. The result does not look exactly like the scan but captures the style. Subjects remain still and are rotated on a turntable during capture; they do not capture clothing in motion. Recent work leverages synthetic renderings from physical simulation to learn mappings from images to 3D garments using deep learning (Danecek et al. 2017). A specific model per garment type needs to be trained and image segmentation is required. Such methods would benefit from the more realistic training data produced by ClothCap.

Clothing segmentation. There are several methods for segmenting clothing in static images (called "parsing" in the computer vision community). An example of recent work exploits a training set of labeled garments and uses convolution neural networks for segmentation (Liang et al. 2015). This work does not address 3D meshes.

Pixels instead of vertices. An alternative to 3D capture is video retrieval and warping (Xu et al. 2011) where they look up video with the right motions and use an image-based approach. Jain et al. (2010) fit an unclothed parametric body model to multi-camera and monocular image data. They then reshape the body and warp the image texture of the foreground appropriately results look realistic. They do not capture the 3D shape of wrinkles and cannot change the camera view. That is, they do not capture or model 3D clothing shape.

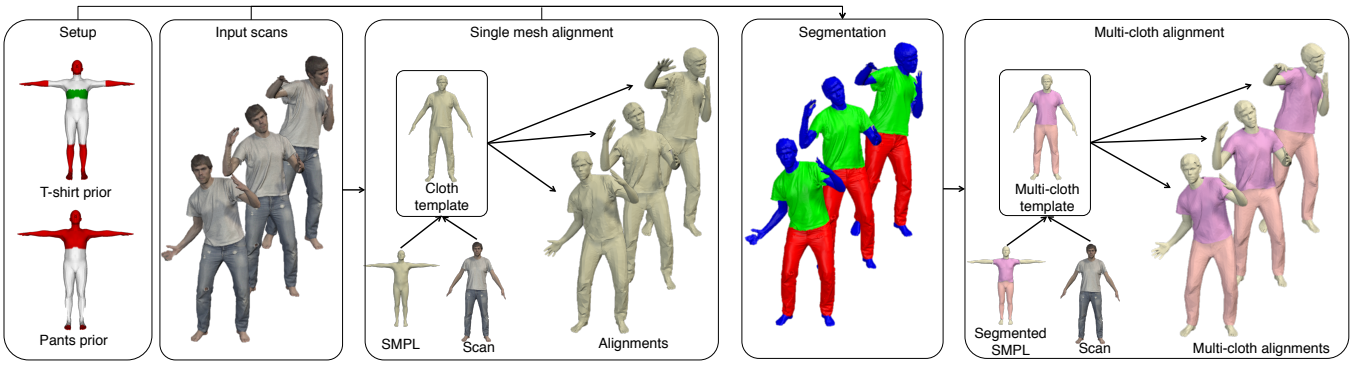

Fig. 2. Outline. Three main steps are needed to obtain multi-cloth alignments: single mesh alignment, scan segmentation and multi-cloth alignment. Multi-cloth alignment garments can be easily retargeted to novel shapes.

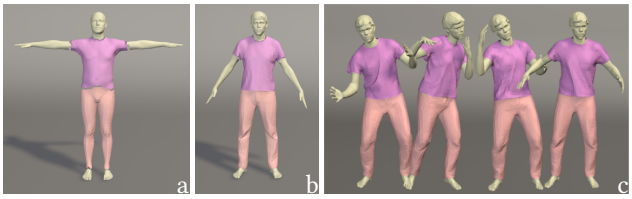

Fig. 3. Concepts. (a) *automatically* segmented cloth template: defines the garment topology only; i.e., which vertices belong to each cloth part. (b) multi-cloth template captures both geometry and topology. This is computed for every new subject and garment and captures the geometry of each garment. (c) multi-cloth alignments are deformations of the multi-cloth template to fit each frame of the sequence.

3 OUTLINE

Our method has four basic steps that we summarize here and then describe in detail in the following sections.

Step 0: Setup. To capture a particular class of garments, we have to define the number of garments N_{garm} and a very weak spatial prior about where garments could be on the body, see Fig. 2 Setup. We show in Section 5.2 that this prior is very crude and easy to define. For now we focus on clothing that has a topology that is easily mapped to the body. We later extend this to more challenging topologies, such as skirts. Here we use the SMPL body model (Loper et al. 2015), which we describe in more detail in the following section. After this setup step, all other steps are automatic.

For a particular subject we estimate a *minimally clothed shape* (MCS); i.e. the shape of the subject without clothes. This MCS could be estimated using an existing technique (Zhang et al. 2017), but, since we are capturing a 4D scan of the subject, capturing an additional minimally clothed scan requires little extra time or effort. We then align the body model to the minimally clothed scan to obtain the subject's MCS. Finally we assume that all scan sequences begin roughly with an "A" pose.

Step 1: Segmentation. The 4D scans include geometry and color information but, of course, contain noise and missing data. Since segmenting these raw scans is difficult, we break the problem into two substeps.

Step 1a: Single mesh alignment. As a first step we align an undressed SMPL body model to the scan sequences (Section 5.1). We deform the body model to fit the observed data and initialize each frame with the result of the previous frame. The output of this step is a sequence of lower-resolution, registered, meshes (Fig. 2).

Step 1b: Segmentation. We then use our segmentation prior and a Markov random field to segment each scan frame in the sequence (Section 5.2). This gives a segmentation of the single mesh alignments, and a per vertex segmentation of the scans. We use single mesh alignment segmentation of the first frame, (which is roughly in an "A" pose) to define a *segmented cloth template* that determines the garment piece topology but not the geometry (Fig. 3a). We then discard the single mesh alignments.

Step 2: Multi-cloth alignment. We deform the segmented cloth template to fit the segmented scan of the first frame and obtain a multi-cloth template (Fig. 3b). The multi-cloth template captures both the topology and the geometry of the garments and the skin (Section 5.3).

The alignment is similar to the single-mesh alignment but here each garment piece is tracked to match the segmented scan data (Fig. 2). Furthermore, the segmentation allows us to define part specific terms in our objective function. This optimization involves a term to regularize the boundaries of the garment, e.g. at the arms, neck, waist, and ankles. In addition, we include a smoothness term for the cloth parts. This step outputs multi-cloth aligned scans (Fig. 3c), which are deformations of the multi-cloth template. Therefore, every garment in every frame is aligned to a common template.

Step 3: Retargeting. Retargeting a garment involves two steps, *preparation* and *dressing*. First we need a source sequence of a person moving in a garment that we want to retarget. From this sequence we estimate the MCS and compute a multi-cloth alignment but this is not enough for retargeting. In the preparation step, we estimate a time-varying *undressed shape* that may deviate from the MCS and we estimate how the clothing is displaced from this body shape in the

reference T-pose. We do this for every frame in the source sequence. In the dressing step, given a new target body shape, we dress it performing the same motion as the source body by simply applying the clothing displacements to the new body in the reference T-pose and posing it as in the source sequence. The following sections provide the details of the full method.

4 BODY MODEL

ClothCap is built on the skinned vertex based model, SMPL (Loper et al. 2015). The SMPL model is a function that takes pose and shape parameters and produces a watertight triangulated mesh with 6890 vertices and 13,776 triangles.

To fix ideas and define notation, meshes and surfaces are denoted using capital calligraphic math, (e.g. $\mathcal{M}$). Matrices and vectors of concatenated mesh vertex coordinates are denoted with capital and bold face (e.g. $\mathbf{M}$). In this way, a mesh is defined by a vector of N vertices and a matrix of F faces $\mathcal{M} = \{\mathbf{M} \in \mathbb{R}^{3N}, \mathbf{F} \in \mathbb{R}^{F\times 3}\}$. Single vertices are denoted by lower case bold face $\mathbf{x} \in \mathbb{R}^3$. Finally, functions are denoted using capital letters (e.g. $M(\cdot)$). Following this notation, the SMPL body model is a function $M(\boldsymbol{\beta}, \boldsymbol{\theta})$, parameterized by shape and pose. The output of this function are the deformed vertices of a triangulated surface. The shape parameters $\boldsymbol{\beta}$ are PCA coefficients of a low-dimensional shape space, learned from thousands of registered scans. In this work we use 50 coefficients: $\boldsymbol{\beta} \in \mathbb{R}^{50}$. The pose of the body is determined by angular rotations in a kinematic structure containing $K = 23$ joints. Every relative rotation between parts is parameterized using the axis-angle representation. Hence, the full pose $\boldsymbol{\theta} \in \mathbb{R}^{72}$ consists of $23 \times 3 + 3$ parameters, 3 parameters per joint plus 3 for the global orientation. The global translation $\mathbf{t}$ adds 3 additional parameters. SMPL relies on a linear blend skinning (LBS) skinning function, $W(\bar{\mathbf{V}}, \mathbf{J}, \boldsymbol{\theta}, \mathbf{W}) : \mathbb{R}^{3N\times 3K\times|\boldsymbol{\theta}|\times|\mathbf{W}|} \mapsto \mathbb{R}^{3N}$, that takes the unposed vertices in the rest pose (or zero pose), $\bar{\mathbf{V}}$, joint locations $\mathbf{J}$, a pose $\boldsymbol{\theta}$, and the blend weights $\mathbf{W}$, and returns the posed vertices. Here, and in the remainder of the paper, unposed vertices are denoted with a bar on top. SMPL effectively parameterizes the skinning function with pose and shape by

$$M(\boldsymbol{\beta}, \boldsymbol{\theta}) = W(T_P(\boldsymbol{\beta}, \boldsymbol{\theta}), J(\boldsymbol{\beta}), \boldsymbol{\theta}, \mathbf{W}) \quad (1)$$

$$T_P(\boldsymbol{\beta}, \boldsymbol{\theta}) = \bar{\mathbf{T}} + B_S(\boldsymbol{\beta}) + B_P(\boldsymbol{\theta}) \quad (2)$$

where $B_S(\boldsymbol{\beta}) \in \mathbb{R}^{3N}$ and $B_P(\boldsymbol{\theta}) \in \mathbb{R}^{3N}$ are vectors of vertices representing offsets from the mean shape $\bar{\mathbf{T}}$. We refer to these as shape and pose blend shapes respectively. The joint locations are inferred using a learned sparse regressor matrix $\mathbf{J}_{\text{reg}} \in \mathbb{R}^{3K\times 3N}$ from the unposed shape, that is $J(\boldsymbol{\beta}) = \mathbf{J}_{\text{reg}}(\bar{\mathbf{T}} + B_S(\beta))$. The SMPL hyper-parameters of SMPL are learned from roughly 4,000 body scans of different people from the CAESAR dataset (Robinette et al. 2002), and another 1600 scans of people in a wide variety of poses. For more details see (Loper et al. 2015).

In our work, the SMPL body model will be used to explain both human body meshes (with no clothes on), as well as clothed meshes. All concepts in Fig. 2 rely on the SMPL model. Thus here we effectively extend SMPL to also model, manipulate, and pose garments.

5 CLOTH CAPTURE: METHODS

5.1 Single mesh alignment

Our goal is to bring all the temporal 3D scans, $\mathcal{S}_k^j$, for a subject j in a particular garment into correspondence by aligning (registering) the SMPL model to all of them. We do this in two stages. The first stage aligns (or registers) SMPL to one scan of the subject in clothing. This will provide a rough template shape that captures the clothing geometry. We use this in the second stage to track the shape over a sequence of scans.

In the first stage, a subject-specific clothed *single mesh*, $\mathcal{A}^j$, is computed based on the first frame of the sequence, which we assume is in an "A" pose. The SMPL model is deformed to explain the first frame. This provides a subject-specific template mesh, including the geometry of the clothing.

Second, we use the sequence-specific template from the first stage and align it to all scans, $\mathcal{S}_k^j$ in the sequence. The process is fully automatic.

5.1.1 Stage 1: Subject-specific single mesh model. Each subject j, in a particular garment, is captured performing a variety of movements starting with an A-pose, obtaining a set of scans $\mathcal{S}^j$. We take the first frame of each sequence, $\mathcal{S}_1^j$. We register a common template to these scans using a method similar to (Pons-Moll et al. 2015a) that regularizes the aligned templates to be similar to a SMPL model.

Specifically we simultaneously solve for the low-D subject-specific body shape parameters, $\boldsymbol{\beta}$, the scan-specific pose parameters, $\boldsymbol{\theta}$, and a deformed template, $\mathcal{A}$, that minimize

$$E(\boldsymbol{\beta}, \boldsymbol{\theta}, \mathcal{A}; \mathcal{S}) = w_g E_g + w_c E_c + w_{\boldsymbol{\theta}} E_{\boldsymbol{\theta}} + w_{\boldsymbol{\beta}} E_{\boldsymbol{\beta}} \quad (3)$$

where E_g is the data term and E_c, $E_{\boldsymbol{\theta}}$ and E_β are priors.

Data term: this encourages the deformed template, $\mathcal{A}$, to be close to the scan surface. For all points, $\mathbf{x}_s$, on the surface of the scan, $\mathcal{S}$, we minimize the point to surface distance dist$(\cdot)$ to the alignment $\mathcal{A}$; where $\rho(e) = e^2/(\sigma^2 + e^2)$ is a robust Geman-McClure penalty function with bandwidth σ

$$E_g(\mathcal{A}; \mathcal{S}) = \sum_{x_s \in \mathcal{S}} \rho(\text{dist}(\mathbf{x}_s, \mathcal{A})). \quad (4)$$

Coupling term: this encourages the alignment deformations to be close to the best SMPL model and vice versa; it plays an important role in regularizing the deformation of the template $\mathcal{A}$. While one could enforce that the vertices of the alignment remain close to the vertices of SMPL, this penalizes heavily tangential motion along the surface, which happens often with clothing. Hence, we enforce instead the edges of the alignment $\mathcal{A}_{t,e}$ (e denotes an edge of triangle t) to remain close to the edges of SMPL. By an abuse of notation, we denote as as $M_{t,e}(\boldsymbol{\beta}, \boldsymbol{\theta})$ the edges of the SMPL mesh corresponding to the vertices $M(\boldsymbol{\beta}, \boldsymbol{\theta})$. The coupling energy is

$$E_c(\mathcal{A}, \boldsymbol{\beta}, \boldsymbol{\theta}) = \sum_{t,e} w_t \|\mathcal{A}_{t,e} - M_{t,e}(\boldsymbol{\beta}, \boldsymbol{\theta})\|_F^2, \quad (5)$$

where the scalar weight, w_t, is set empirically to increase the coupling strength for parts like hands and feet where the scans are noisy and no clothing is occluding the body shape.

Pose prior: This enforces the pose to be close to an A pose at this stage. This is easily achieved by fitting a Gaussian distribution $\mathcal{N}(\mu_{\boldsymbol{\theta}}, \Sigma_{\boldsymbol{\theta}})$ to the poses of the pose dataset that are labeled as A-poses. This term then penalizes large Mahalanobis distances

$$E_{\boldsymbol{\theta}}(\boldsymbol{\theta}) = D_M(\boldsymbol{\theta}; \mu_{\boldsymbol{\theta}}, \Sigma_{\boldsymbol{\theta}}). \tag{6}$$

Shape prior: SMPL provides a Gaussian prior over body shapes defined by the diagonal covariance matrix $\Sigma_{\boldsymbol{\beta}}$. The Mahalanobis distance then constrains the shape to be within the space of human bodies:

$$E_{\boldsymbol{\beta}}(\boldsymbol{\beta}) = D_M(\boldsymbol{\beta}; 0, \Sigma_{\boldsymbol{\beta}}). \tag{7}$$

Optimization is performed using different weights. We first set the regularizer weights such that the optimization initially performs model-only alignment to fit the SMPL parameters to the scan. Initially, the regularizer weights $w_c, w_{\boldsymbol{\theta}}, w_{\boldsymbol{\beta}}$ are large to help prevent the solution finding local optima. In particular, we initially set the coupling weight w_c to infinite, effectively performing a model-only alignment first. This converges to a reasonable initial body shape and pose underneath the cloth. Then the regularizer weights are decreased and the data weight increased, allowing the aligned mesh, $\mathcal{A}$, to deform away from the body shape to explain the outer cloth surfaces.

Since the objective in Eq. (3) is highly non-convex, initialization plays an important role. Here we initialize the pose, $\boldsymbol{\theta}$, to $\mu_{\boldsymbol{\theta}}$, the shape to the mean shape ($\boldsymbol{\beta} = 0$), and the registration edges, $\mathcal{A}_{t,e}$, to the model edges, $M_{t,e}(\boldsymbol{\beta}, \boldsymbol{\theta}_k)$.

Cloth template: From the optimization, we obtain a registered surface $\mathcal{A} = \{\mathbf{A}, \mathbf{F}\}$ in a given pose $\boldsymbol{\theta}$. However, an aligned mesh $\mathcal{A}$ can not be reposed so it can not be used to regularize the fitting process for the rest of the sequences. Hence, we compute a clothed template (in "T"-pose), by solving the following optimization problem:

$$\bar{\mathbf{A}} = \arg\min_{\bar{\mathbf{A}}} \|W(\bar{\mathbf{A}} + B_P(\boldsymbol{\theta}), \mathbf{J}_{\text{reg}}\bar{\mathbf{A}}, \boldsymbol{\theta}, \mathbf{W}) - \mathbf{A}\|_F^2. \tag{8}$$

This computes the *unposed* mesh $\bar{\mathbf{A}}$ that, when posed, matches the alignment $\mathbf{A}$; see (Loper et al. 2015).

The *single mesh* model $C(\bar{\mathbf{A}}, \boldsymbol{\theta}) = W(\bar{\mathbf{A}} + B_P(\boldsymbol{\theta}), \mathbf{J}_{\text{reg}}\bar{\mathbf{A}}, \boldsymbol{\theta}, \mathbf{W})$ is a function that takes as input a pose and the unposed vertices and produces posed vertices. This function is effectively the SMPL model with the shape fixed and set to $\bar{\mathbf{A}}$, which captures the clothing.

5.1.2 Stage 2: subject-specific sequence registration. Given subject specific single mesh models, $C(\bar{\mathbf{A}}, \boldsymbol{\theta})$, we align the template to the full sequences of scans $\mathcal{S}_k$, while regularizing the solution to the subject-specific model by minimizing

$$E(\mathcal{A}_k, \boldsymbol{\theta}_k; \mathcal{S}_k) = w_g E_g + w_c E'_c \tag{9}$$

$$E'_c(\mathcal{A}_k, \boldsymbol{\theta}_k; \bar{\mathbf{A}}) = \sum_{t,e} w_t \|\mathcal{A}_{t,e,k} - C_{t,e}(\bar{\mathbf{A}}, \boldsymbol{\theta}_k)\|_F^2, \tag{10}$$

where $E_g(\cdot)$ is the same as in Eq. (3) and $E'_c(\mathcal{A}_k, \boldsymbol{\theta}_k)$ is a modified coupling term that regularizes the edges of the solution to be close to the edges of the single mesh model. Note that for tight clothing the single mesh model is already a reasonable approximation but for loose apparel it can differ significantly from reality. We initialize the optimization at each frame with the pose from the previous frame and therefore the pose prior $E_{\boldsymbol{\theta}}(\cdot)$, is not needed. For the first frame, we use the aligned mesh obtained during the subject-specific single mesh model creation step. Results of single mesh alignments can be seen in Figure 5. Since the template is single closed mesh, cloth boundaries are not properly tracked and hence results lack realism. Furthermore, they suffer from surface sliding: this means motion tangential to the surface is not well tracked. Below we segment the scans and the templates to track the boundaries and constrain tracking.

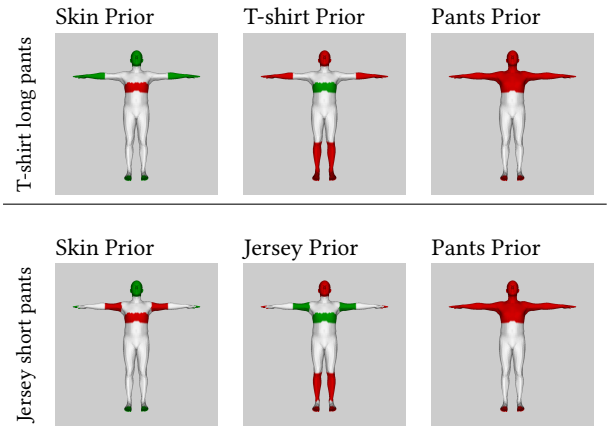

Fig. 4. Manually defined clothing segmentation prior on the body model. A cloth configuration consists of different garments. For each garment, each body model vertex is labeled as: likely to be (green), agnostic (gray) or unlikely to be (red). First row: priors of the configuration "t-shirt long pants" consisting of "Skin", "T-shirt" and "Pants". Second row: priors of the configuration "jersey short pants", consisting of "Skin", "Jersey" and "Pants". Notice how the priors are coarsely defined.

5.2 Body model aided segmentation

We argue that a segmentation of the scan data into garment parts makes the subsequent task of alignment easier, allowing us to track the garment boundaries more accurately. Unfortunately, clothing segmentation itself is a hard problem (Fig. 6). Fortunately, segmentation can benefit from 3D shape information. The single mesh alignments, while inaccurate, provide additional information about the body parts and how to segment the mesh into garments. To that end, instead of solving the segmentation problem in the image domain, where no shape information is available, we perform segmentation directly on the graph given by the alignment itself. This also has the strong advantage of not suffering from self occlusions; e.g., the arm occluding the shirt is not a problem in our formulation since the arm and the shirt are topologically very far apart.

More formally, let us introduce one random variable v_i, for every vertex in our template mesh $i \in \mathcal{T}$. These random variables can take discrete values $s_i \in \{0 \dots N_{\text{garm}}\}$, where 0 corresponds to the skin and N_{garm} are the number of garments the person is wearing. With some abuse of notation, we solve for the collection of random variables $\mathbf{v} = \{v_i \mid i \in \mathcal{T}\}$ that minimize the following cost function

$$E(\mathbf{v}) = \sum_{i \in \mathcal{T}} \varphi_i(v_i) + \sum_{(i,j) \in \mathcal{T}} \psi_{ij}(v_i, v_j) \tag{11}$$

Fig. 5. The relevance of a multi-cloth model. We show 2 triplets with: scan, single-mesh alignment and multi-cloth alignment. The first frame is accurately segmented by both Steps, but in a later frame, the single mesh alignment can not accurately track the garment. The t-shirt (in lavender) is explained by the wrong set of vertices. The single-mesh alignment could be improved using an appearance based term. However, a single mesh is not able to model the gap between the trousers and the t-shirt. Using a different mesh to track each garment better matches the physical situation.

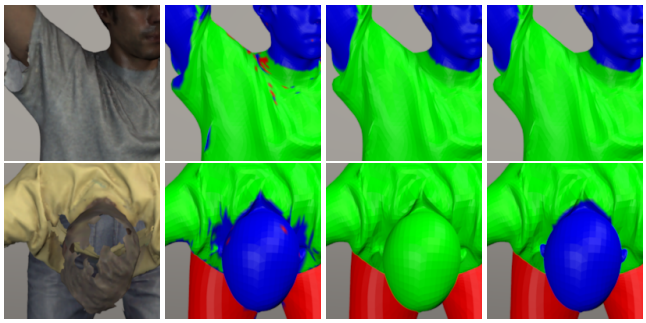

Fig. 6. Closeups of segmentation results. From left to right: input scan with color, segmentation using only the Gaussian mixture models (segmentation is noisy), segmentation with MRF but no prior (the head is wrongly labeled as shirt), segmentation with prior and MRF (best result).

where φ_i is a node-dependent unary term and ψ_{ij} is a binary term.

Unary term: The unary term encodes the negative log likelihood of node i taking label v_i

$$\varphi_i(v_i) = \sum_{j \in \mathcal{B}_i(\mathcal{S})} -\log(p_j(v_i)) + \epsilon_i(v_i) \tag{12}$$

where the first term is the *data likelihood term* and the second term ϵ is a *garment prior* over body parts.

Data likelihood term: We fit a Gaussian mixture model (GMM) to the appearance of each of the garments. Then, for every scan point $\mathbf{x}_j$ in the neighborhood $\mathcal{B}_i(\mathcal{S})$ of node i, we evaluate the likelihood under the fitted GMM:

$$p_j(s) = \sum_{m=0}^{N} \pi_s^m \mathcal{N}(I(\mathbf{x}_j)|\mu_s^m, \Sigma_s^m) \tag{13}$$

where $I(\mathbf{x}_j)$ is the HSV appearance of scan point $\mathbf{x}_j$, and μ_s^m, Σ_s^m are the mean and covariance of mixture mode m of segmentation class s. Note that in order to be more robust to illumination changes we fit the GMM in HSV space instead of RGB. To train the GMM, we use the first frame in the "A" pose. The segmentation for this frame is obtained automatically by replacing the GMM probabilities p_j with a simpler unsupervised K-means voting scheme.

The appearance model is sensitive to noise, shadows and illumination changes. See the obtained results using only the appearance term, second column of Fig. 6. Consequently, we add a rough garment prior term that encodes prior knowledge of plausible garment segmentations.

Garment prior: This encodes intuitive information such as: the torso nodes are likely to be T-shirt while hands and head have to be unclothed (skin). To that end, we leverage the underlying segmentation of the body into parts that is provided by SMPL. This prior is manually defined per garment configuration (see Fig. 4). Formally, we define two kinds of priors,

(1) a node i is likely to be label $l_i \in \{0 \dots N_{\text{garm}}\}$ in which case $\epsilon_i(s) = 1 - \delta(s - l_i)$
(2) a node i can not take a certain label in which case $\epsilon_i(s) = \delta(s - l_i)$.

In particular, for the case of a t-shirt and long trousers (first row of Fig. 4), the nodes of the head, hands and feet should be labeled as "skin," and the nodes of the torso should be labeled "shirt". This is a very conservative prior. Notice that we do not make any hard assumption about the top of the trousers since the belly of the subject could be visible. This sort of intuitive prior knowledge effectively improves segmentation as illustrated in Fig. 6.

Pairwise term: This smoothness term encourages neighboring pixels to have the same label. This term is extremely simple; given the adjacency matrix $\mathbf{Z}$ of size $N \times N$ of our template mesh, where N is the number of vertices in $\mathcal{T}$, the term is

$$\psi_{ij}(v_i, v_j) = \mathbf{Z}_{ij}\,(1 - \delta(v_i - v_j)) \tag{14}$$

taking cost 1 if nodes i and j are neighbors and take different labels, and cost 0 otherwise. We minimize the energy in Eq. (11) using alpha expansion. The resulting segmentation is a per node label, s_i, for each random variable, $v_i = s_i$, indicating to which garment a node belongs.

During the multi-cloth alignment it is very important that the boundaries of the garments match correctly. Therefore, we split the single mesh alignment into garment pieces, and for every connected mesh we detect the boundaries by identifying the edges of the mesh with only one face. Furthermore, we can easily identify and label the different boundaries of every garment. We use again the

Fig. 7. Segmentation results: rows of scan geometry segmentations, shirt (top row), pants (bottom row). After solving the discrete MRF on the body topology we propagate the obtained labels back to the scan. This results in segmented scans $\mathcal{S}_k$. Boundaries (shown in white) can be easily identified and are propagated to the scans. Such labels provide important information for the multi-cloth alignment step. They allow ClothCap to estimate consistent geometry with smooth boundaries.

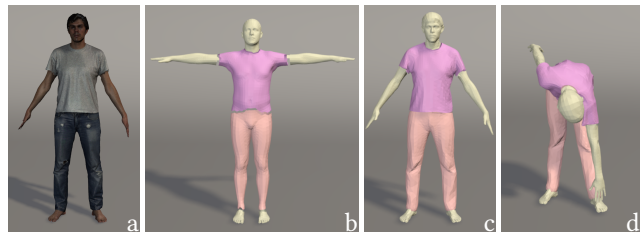

Fig. 8. Multi-cloth template extraction. Using the segmentation labels on the alignment for the first frame we break the mean shape body into a segmented cloth template (b). The segmented cloth template is then deformed to explain each cloth piece and the skin of the first frame scan (a). This results in the multi-cloth template (c). Since every vertex of the multi-cloth template is associated with one body model vertex, we can easily repose it (d) using SMPL. This reposing is critical for tracking the garments accurately over time. The extracted multi-cloth template is used to register all the dynamic sequences. In this way we align all scans to a common template.

garment prior information, which relates garments to the body. Using the body parts, we automatically assign labels all the garment boundaries (e.g. right sleeve, top of the t-shirt or bottom). This will facilitate the matching process in the multi-cloth alignment stage. Finally, the segmentation labels in the single mesh alignment are mapped back to the scan by finding the closest scan points to each mesh vertex.

The result is a per scan point label indicating wether it is skin or one of the garments. Figure 7 shows some results of segmentation of the scan vertices, where the red and green parts correspond to different garments. If the scan point corresponds to a boundary vertex in the model, then it gets a label indicating which garment boundary it corresponds to; these are all shown as white vertices in Fig. 7. Hence we obtain a per frame vector of labels denoted as $\mathbf{v}_{s,k}$. Using those labels we segment the scan into garment parts (Fig. 7). These labels provide very important additional information that we use to reconstruct and track the garment geometry together with the body over time.

5.3 Multi-cloth alignment

We reformulate the alignment problem in a more constrained way. The input scan has been segmented into $N_{\text{garm}} + 1$ meshes $\mathcal{S} = \{\mathcal{S}^0, \ldots, \mathcal{S}^{N_{\text{garm}}}\}$, see Fig. 7. As we did in the single mesh alignment (Sec. 5.2) we perform the multi-cloth alignment in two stages that use the same objective function. In a first stage we compute a single subject-specific multi-cloth template using one frame, which has also $N_{\text{garm}} + 1$ parts as shown in Fig. 8 c. In a second stage we use the multi-cloth template to regularize the temporal alignments. Each template part is used to align and track the corresponding scan vertices. The main difference with the single mesh alignment (Sec. 5.1) is that now the template and the scan are segmented into parts; furthermore we know the correspondence between parts in the template and scan making alignment easier. From now on we refer to the "Skin" part with the super-index 0.

5.3.1 Stage 1: Subject-specific multi-cloth template. We illustrate the computation of a multi-cloth template $\bar{\mathbf{L}}$ in Fig. 8. We first break the SMPL mean shape body into a segmented cloth template (Fig. 8 b) using the segmentation labels on the single mesh alignment of the first frame of a sequence. The segmented cloth template is then deformed to explain each cloth piece and the skin of the first frame scan (Fig. 8 a). This results in the multi-cloth template (Fig. 8 c). The multi-cloth template is computed only once per subject and clothing configuration. Then it is kept fix to align all sequences. Let $C(\bar{\mathbf{L}}, \boldsymbol{\theta}) = \{C^0(\bar{\mathbf{L}}, \boldsymbol{\theta}), \ldots, C^{N_{\text{garm}}}(\bar{\mathbf{L}}, \boldsymbol{\theta})\}$ denote the multi-cloth model function that takes a multi-cloth template $\bar{\mathbf{L}}$, and poses it in a given pose $\boldsymbol{\theta}$. An example result can be seen in Fig. 8 d. Let us introduce the multi-cloth mesh alignments $\mathcal{L}$ which consist of a set of parts $\mathcal{L} = \{\mathcal{L}^0, \ldots, \mathcal{L}^{N_{\text{garm}}}\}$. A multi-cloth alignment has the same topology as C but its vertices are free variables to optimize. We note these vertices with $\mathbf{L}$. Our goal now is to compute multi-cloth alignments $\mathcal{L}$ that best fit the scan data. We minimize the following objective:

$$E(\Phi_k) = w_g E'_g + w_b E_{\text{bound}} + w_c E'_c + w_s E_{\text{lap}} + w_a E_{\text{smth}}. \quad (15)$$

Similar to the single mesh alignment optimization, E'_g, and E_b are data terms and $E'_c, E_{\text{stretch}}, E_{\text{smth}}$ are priors. At every frame k the variables of the optimization are $\Phi_k = [\boldsymbol{\theta}_k, \mathbf{L}_k]$. Note that we optimize the vertices $\mathbf{L}_k$ of the the multi-cloth alignment $\mathcal{L}_k$ but the topology remains constant during tracking.

To simplify notation we drop the frame index k in the following. We modify the data and coupling terms in Eq. (3) to include segmentation information.

Data term: We enforce each alignment mesh $\mathcal{L}^l$ to match with its corresponding segmented scan mesh $\mathcal{S}^l$ using Eq. (4) for each

part:

$$E'_g(\mathcal{L};\mathcal{S}) = \sum_{l=0}^{N_{\text{garm}}} E_g(\mathcal{L}^l;\mathcal{S}^l). \tag{16}$$

Boundary term: The vertices on the surface boundaries provide a very useful constraint on how the garments can deform. Each garment surface provides rings at its extremity points. On the scan, the rings are obtained from the labels $\mathbf{v}_{s,k}$, whereas on the segmented cloth template, the mesh ring topology is constant and consists of all edges with two vertices on a boundary. Let $B_r(\cdot)$ be a function that takes a mesh and returns the corresponding mesh ring r. Then we can force the multi-cloth alignment boundaries to match the scan boundary vertices by minimizing

$$E_{\text{bound}}(\mathcal{L};\mathcal{S}) = \sum_{l=1}^{N_{\text{garm}}} \sum_{r=0}^{R_l} E_g(B_r(\mathcal{L}^l), B_r(\mathcal{S}^l)). \tag{17}$$

Note that the correspondence between the R_l different garment rings r (*e.g.*, bottom of shirt, left sleeve,...) is known for each garment l. Note also that we do not match the skin boundaries; i.e. the sum over garments in Eq. (17) starts at $l = 1$. We want to emphasize that this term is generic enough to handle boundaries in skirts or even open jackets.

Coupling term: Similar to the coupling term in the single mesh alignment, this term couples the solution to be close to a posed multi-cloth model $C(\bar{\mathbf{L}}, \boldsymbol{\theta})$. Using a clothed scan in an "A-pose", we optimize the SMPL subject-specific shape parameters $\boldsymbol{\beta}^j$ and initialize $\bar{\mathbf{L}} = M(\boldsymbol{\beta}^j, \mathbf{0})$. To independently couple each garment, we use Eq. (5) to define the multi-cloth coupling term

$$E''_c(\mathcal{L}, \boldsymbol{\theta}; M(\boldsymbol{\beta}^j, \mathbf{0})) = \sum_{l=0}^{N_{\text{garm}}} E'_c(\mathcal{L}^l, \boldsymbol{\theta}; C^l(M(\boldsymbol{\beta}^j, \mathbf{0}), \boldsymbol{\theta})). \tag{18}$$

Cloth deformation terms: We include two additional prior terms for the garments: a Laplacian term and an additional smoothness term for the boundaries.

The Laplacian term: In order to keep the garment meshes well behaved we include a Laplacian term (Sorkine 2006; Sorkine et al. 2004). Given a mesh with adjacency matrix $\mathbf{Z}$, the graph Laplacian is obtained as $\mathbf{G}_{\text{lap}} = \mathbf{I} - \mathbf{H}^{-1}\mathbf{Z}$ where $\mathbf{H}$ is a diagonal matrix such that $\mathbf{H}_{ii}$ equals the number of neighbors of vertex $\mathbf{x}_i$. We minimize the squared norm of the mesh differential coordinates for the garments

$$E_{\text{lap}}(\mathcal{L}) = \sum_{l=1}^{N_{\text{garm}}} \|\mathbf{G}^l_{\text{lap}}\mathbf{L}^l\|_F^2, \tag{19}$$

where $\|\cdot\|_F$ denotes the matrix Frobenious norm. This term penalizes deviations of vertices from the center of mass of its neighbors. This avoids triangle flips and spikes, and results in more robust tracking.

Boundary smoothness term: Since the boundary detections from the segmentation algorithm are noisy, it is important to enforce that alignment boundaries are smooth. We enforce the smoothness by penalizing a second order derivative of the boundary rings. Given a set of ordered boundary vertices for every ring $\mathbf{x}_{r,n}$ (where r indexes rings and n indexes the points in the ring) the term is

$$E_{\text{smth}}(\mathcal{L}) = \sum_{r=0}^{R_l} \sum_{n} \|\mathbf{x}_{r,n-1} - 2\mathbf{x}_{r,n} + \mathbf{x}_{r,n+1}\|^2. \tag{20}$$

Fig. 9. Undressed shape estimation. Multi-cloth alignments (left images) capture only the visible skin parts and the body pose. The final result (right images) includes the undressed shape underneath the cloth.

We solve Eq. 15 in a first frame in "A" pose, and we unpose the result to obtain an updated subject specific multi-cloth template $\bar{\mathbf{L}}$. The unposing is done as before using Eq. (8). A computed multi-cloth template for one subject is illustrated in Fig. 8 c.

5.3.2 Stage 2: Subject-specific sequence registration. Given a subject specific multi-cloth template $\bar{\mathbf{L}}$ we optimize Eq. (15) for all the sequence frames. At every frame k the variables of optimization are $\Phi_k = [\boldsymbol{\theta}_k, \mathbf{L}_k]$ and the coupling term in Eq. (18) uses the updated template $\bar{\mathbf{L}}$

$$E'''_c(\mathcal{L}, \boldsymbol{\theta}; \bar{\mathbf{L}}) = \sum_{l=0}^{N_{\text{garm}}} E'_c(\mathcal{L}^l, \boldsymbol{\theta}; \bar{\mathbf{L}}). \tag{21}$$

5.3.3 Implementation details. We use the same set of weights for all the results in the experiments $w_g = 1000, w_b = 20, w_c = 1.5, w_s = 200, w_a = 20$. Weights were set empirically so that the terms are balanced (errors of roughly the same order of magnitude). The scan units are millimeters. When we compute the multi-cloth template (in the first frame in "A" pose) we set the weight higher $w_s = 2000$ to obtain a smooth template without pose-dependent wrinkles.

5.4 Retargeting

Our goal is to retarget the cloth from a subject, cap, to a new subject, target. To this end, two steps are required, a *preparation* step to compute displacements between the multi-cloth alignments and the underlying body shape and a *dressing* step to apply those displacements to a novel target shape. The multi-cloth alignment method only tracks the visible part of the skin, and thus the full underlying body shape is not computed, see Fig. 9 left images.

Preparation: we compute a minimally clothed shape (MCS) $\bar{\mathbf{N}}^{\text{cap}}$ of the subject. To do so, we scanned subjects wearing minimal clothes (tight fitting sports underwear) and fitted the body model to the scans. This provides us a good estimation of the static MCS of each subject. This step could be replaced by an automatic method to estimate shape under clothing (Zhang et al. 2017).

Ideally, one option to obtain the shape underneath the dynamic multi-cloth alignments would be to simply pose the MCS with the estimated alignment poses $\boldsymbol{\theta}_k$. However, the body shape stretches and compresses as we move and SMPL (while powerful) does not model such changes to the required accuracy.

Therefore, we estimate at every frame a time varying shape. We minimize an objective function that forces the shape to remain inside the cloth and tight to the visible parts of the multi-cloth alignment; for more details on the objective function we refer to (Zhang et al. 2017). Here, to make the problem better behaved we constrain the shape to remain close to the MCS. This results in time varying *undressed shapes* $\bar{\mathbf{N}}_k$ that lie inside the multi-cloth alignments $\bar{\mathbf{L}}_k$. All the processing is done in the unposed space (T pose), and therefore alignments and shapes have a superscript bar.

From the captured undressed subject shapes, $\bar{\mathbf{N}}_k^{\text{cap}}$, and multi-cloth alignments $\bar{\mathcal{L}}_k^{\text{cap}}$ we compute two sets of displacements: the dynamic shape displacements $\bar{\mathbf{D}}^{\text{dyna}}$ and the cloth displacements. The first ones encode time varying body shape deformations while the second ones encode cloth deformations. The dynamic shape displacements are computed as:

$$\bar{\mathbf{D}}^{\text{dyna}} = \bar{\mathbf{N}}_k^{\text{cap}} - \bar{\mathbf{N}}^{\text{cap}}, \tag{22}$$

and the cloth displacements are computed per garment as

$$\bar{\mathbf{D}}^{\text{cloth}} = \bar{\mathbf{L}}_k^{l,\text{cap}} - \mathbf{U}_{\text{cloth}}^{l} \bar{\mathbf{N}}_k^{\text{cap}}, \tag{23}$$

where $\mathbf{U}_{\text{cloth}}^{l}$ is a mask matrix that selects the undressed vertices corresponding to garment l.

Dressing: now given a novel shape we obtain the undressed shape as:

$$\bar{\mathbf{N}}_k^{\text{target}} = \bar{\mathbf{N}}^{\text{target}} + \bar{\mathbf{D}}^{\text{dyna}}. \tag{24}$$

The deformed garment on the new shape is obtained as

$$\bar{\mathbf{L}}_k^{l,\text{target}} = \mathbf{U}_{\text{cloth}}^{l} \bar{\mathbf{N}}_k^{\text{target}} + \bar{\mathbf{D}}^{\text{cloth}}. \tag{25}$$

For a given frame k the shape is then posed with the captured multi-cloth pose $\boldsymbol{\theta}_k$. The retarget result of a subject onto itself allows us to have an animation of the cloth and the full body underneath, see Fig. 9 right images. More results retargeting to new shapes are illustrated in the experiments. This illustrates the power of the proposed multi-cloth model.

6 EXPERIMENTS

6.1 Multi-cloth alignment results

In order to evaluate our method, we captured different subjects with different clothes performing a range of dynamic motions including running, jumping, posing or punching. We show results for a variety of garments including jeans, t-shirts, dres shirts, tops, shorts, jerseys and even skirts. In Fig. 10 we present results of the multi-cloth alignments. We find that the automatically computed segmentation greatly constrains the fitting process allowing different constraints for cloth and skin regions. The computed cloth boundaries help the method to accurately estimate and track the garment over time. In Fig. 11 we show the top part of the multi-cloth alignments. Notice

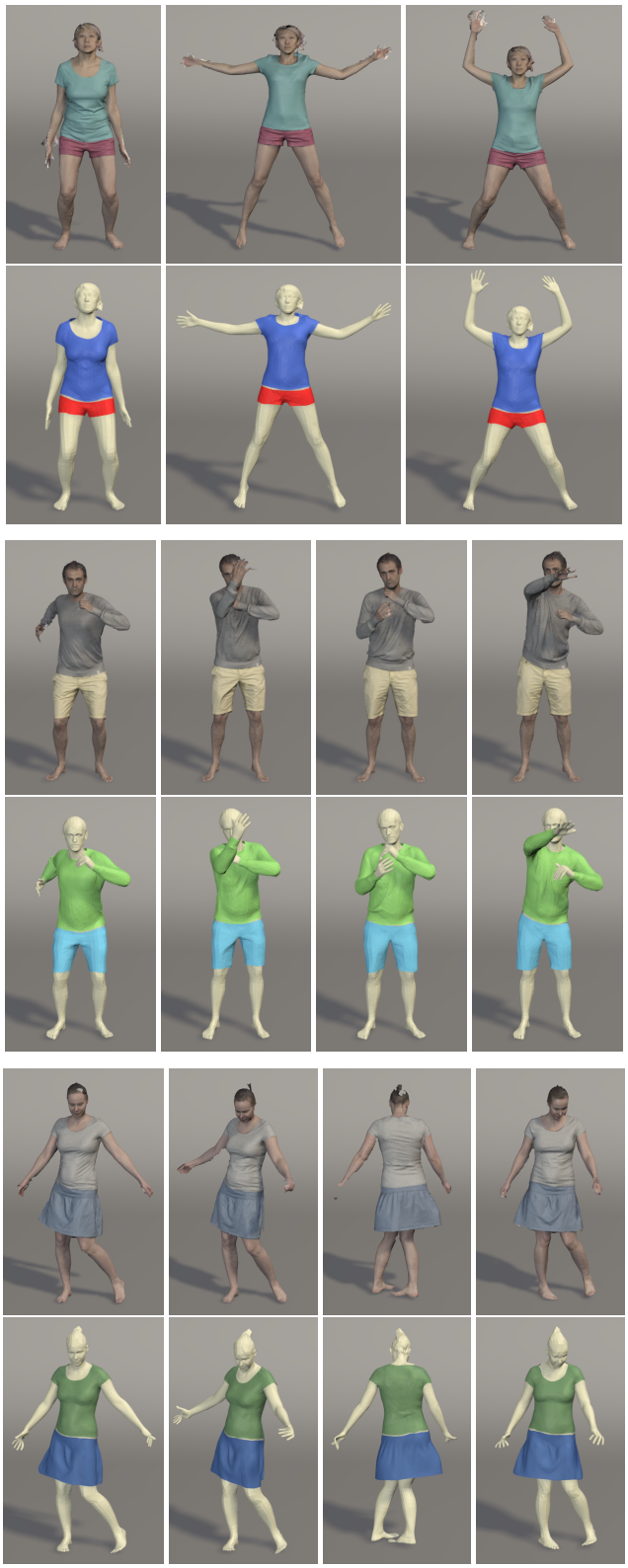

Fig. 10. Results of the multi-cloth alignments with shape underneath presented in pairs of (scan, result). From top to bottom we present different garments: t-shirt with shorts, jersey with shorts and t-shirt with skirt.

Fig. 11. Multi-cloth results. We present pairs of scans and multi-cloth results for two tops, a dress shirt a t-shirt. Notice how many of the wrinkles are captured by the aligned multi-cloth model. In order to capture the remaining high-frequency wrinkles, one could use a higher resolution segmented cloth template.

how the boundaries of the cloth are properly tracked, and the wrinkles in each frame captured. The sequences can be seen in the in the video.

Garments with topology different from the body. For garments with the same topology as the body, such as shirts and pants, our method for garment extraction is fully automatic (except for the crude segmentation prior explained in Section 5.2). For garments with topology different from the body such as skirts we need an extra manual step. To obtain the garment geometry we start with the automatic scan segmentation of the skirt. This provides a point cloud with noise and holes. Therefore, we cleaned the point cloud, and re-topologized it. If a garment template is available, this step could have been skipped. To associate the skirt with the model we associated all vertices with the root joint of SMPL. In this way, the skirt rotates and translates with the root motions; while this is a simplification it serves as a good template for tracking. After the template is available we use the exact same approach as before; results can be seen in Fig. 10. Notice how the skirt is nicely captured with sharp boundaries.

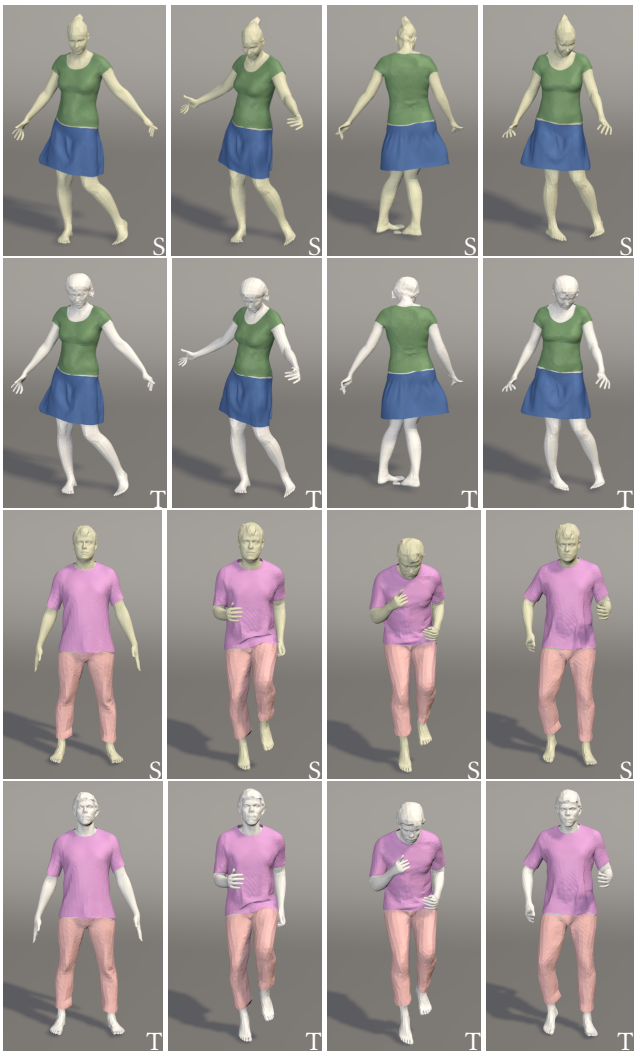

Fig. 12. Dynamic retargeting. We present pairs of: (S) source multi-cloth alignment, (T) retargeted clothes to new subject. The target body shape is rendered lighter for visual identification. Notice how wrinkles created by a pose are transferred from the source into the target.

6.2 Dynamic Retargeting

One of the powerful features of ClothCap is that the cloth is associated with a body model. This is not only useful for capture but also for retargeting to new bodies. Figure 12 shows results of retargeting of dynamic cloth sequences to different bodies. Our method produces compelling results even for highly dynamic motions such as running (please see the video). This provides a foundation for virtual try-on. A clothing retailer can scan a professional model once in a variety of clothing and then this real clothing can be retargeted to any body shape, which we animate with the original motion.

6.3 Changing the body shape

Animators might also want to edit the body shape without having to manually re-adjust all the clothing pieces. This can be easily achieved with ClothCap. We simply add weight to a subject using the SMPL body model and retarget the captured clothing sequence to this novel shape. See Fig. 13 and supplementary video.

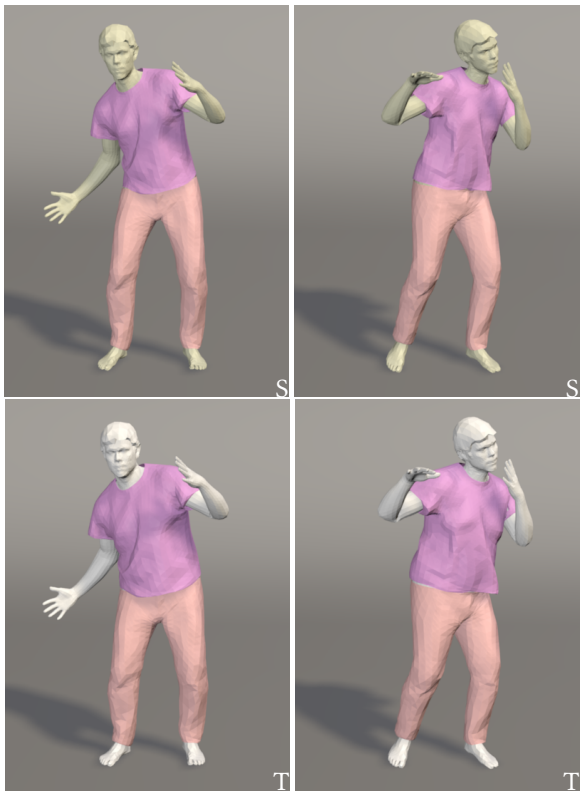

Fig. 13. Changing the body shape. The source (S) multi-cloth alignment allows retargeting captured clothing to a modified shape (T), producing realistic results.

6.4 Dressing the CAESAR dataset

To show the generalisation of our method to novel shapes we took the static shapes from the CAESAR dataset (Robinette et al. 2002) and we dressed them with the clothes of one of our captured subjects. This can be seen in Figure 15, where we show retargeting results for male subjects and for female subjects. Since the capture system also records RGB images we also add texture to the captured garments and visualize it on new bodies.

6.5 Virtual clothing from a single image

We further demonstrate the applicability of our method by retargeting clothing to virtual avatars reconstructed from an image in Fig. 14. Here we use the approach of Lassner et al. (2017) that estimates an average shape and a 3D pose from a single image. We first dress the bodies in a T-pose and then repose the multi-cloth meshes using the multi-cloth model function. This shows a potential virtual try-on application: a user takes a picture, chooses a clothing item that was already captured (*e.g*, by a clothing retailer) using ClothCap and obtains a clothed avatar. Note that a more detailed MCS could be obtained using shape word ratings (Streuber et al. 2016).

Fig. 14. Virtual try-on. We capture clothing from a scan (a). Then from a single image (b)[1], we use the method of Lassner et al. (2017) to reconstruct an avatar (c) and apply the captured clothing to dress the avatar (d).

7 DISCUSSION AND CONCLUSIONS

We have presented what we think is the most sophisticated method to date for capturing clothing in motion. We rely on a multi-cloth 3D model of the body and clothing. Given high resolution 4D scans, we segment the garments, estimate the undressed body shape, and track the clothing surface over time. We have demonstrated several applications relevant for virtual-try on among others. Clothing can be swapped, reshaped to new bodies, and reposed to some degree. We see this as a first step towards the statistical modeling of clothing from scans.

Limitations. There are several limitations to the method. The next step will be to model: 1) more complex garments like collars, cuffs, buttons, etc.; 2) open jackets, which are partially connected to the body and partially not; and 3) clothing items like ties, capes, and scarfs, which do not necessarily have a fixed relationship with the body.

Here we focused on a model with three parts (body, pants and shirt), and future work should look at more garments, including estimating the number of garments automatically (as done in layered optical flow estimation (Sun et al. 2010)). Our method estimates the undressed shape as a post-process step; future work will investigate the problem of capturing cloth and the MCS jointly in one step. This involves reasoning about occlusion and intersection between garments during optimization, which is challenging. We will also investigate the use of Inertial Measurement Units on the body (von Marcard et al. 2016, 2017) to facilitate the estimation of pose and the MCS with very baggy clothing.

We also need to deal with layers that never become fully disoccluded; e.g. can we infer what the top of pants look like even if we never see them? Currently we only track the visible parts of the garments and not their occluded parts. Here we have adopted a staged strategy for optimizing the segmentation, alignment, and undressed shape. While practical, better results could likely be obtained by a more integrated objective function and joint optimization strategy.

We showed that captured garments can be retargeted to new body shapes, are automatically rigged, and can be reposed. This retargetting is not physically realistic in terms of wrinkles. The reposing uses linear blend skinning with pose-dependent blend shapes copied from the SMPL body model. These blend shapes are learned to model changes in minimally cloth body shapes in different poses, but do not correctly model clothing deformations. While the reposed garments generally look realistic, and may be sufficient for some virtual try-on applications, a more sophisticated model is needed for animation that modifies the wrinkles with body pose; that is, we need a model of clothing blend shapes that capture wrinkles. Our ClothCap system provides the training data needed for researchers to learn such a model.

Future work. If we are able to scan many garments in motion then, given a single scan of a new garment, we should be able to find similar garments and animate the new garment realistically. This is a key technology that would enable an easy to use virtual try-on system. An on-line seller scans the garment once and then any shopper can visualize the garment on their body.

The garment prior from Fig. 4 must be manually specified. Although the task is not cumbersome, a per garment specific prior could be learned or automatically inferred from different scans of the garment. This way, more challenging garments could be segmented and tracked.

Future work should also explore the use of higher-resolution meshes for the clothing parts or use displacement maps to capture more wrinkle detail. While higher resolution meshes are interesting, applications for video games might benefit from even lower resolutions. With our existing technology, making simple clothing (e.g. sports outfits) that can be transferred to video game characters should be feasible.

Here we have not focused on capturing material appearance but, given our surface estimates, it should be possible to extract good reflectance information enabling relighting of garments. Also, given our multi-cloth segmentation and 4D data, we could also revisit the methods (Kim et al. 2017; Stoll et al. 2010) to estimate physical parameters for each garment and body separately, enabling a physics-based animation of the segmented garments. Moreover, adding appearance information to the tracking framework as in (Bogo et al. 2017) should result in even more stable tracking results.

Our segmentation method uses a fairly weak model of image appearance. Recent work in computer vision addresses the segmentation of clothing in images using neural networks. Networks trained for these problems may be readily adapted to our problem, which combines color and shape information. Thus, for very

[1]Credits: First and third row images b) reproduced from Dantone et al. (2014), second row image b) reproduced from Johnson and Everingham (2010).

Fig. 15. CAESAR bodies dressed. Given a single capture of a garment we can trivially retarget it to any novel body shape. We demonstrate this by dressing many subjects with a plausible result. Left column: MCS of the CAESAR subjects; middle column: dressed subjects with garment color coded. Right column: subjects dressed and texture applied on the cloth.

complex appearance patterns, more powerful unaries coming from neural networks could be easily integrated into our formulation.

Future work will investigate capture and retargeting of clothing items such as belts, ties, jackets, frills/ruffles, capes, etc. Accessories such as bags, which do not have a clear mapping to the body, can not be handled with our approach and remain an open and challenging problem. Finally, we should be able to dress cartoon characters as well as long as the body shape is in correspondence with our body mesh.

ACKNOWLEDGMENTS

We thank J. Romero for interesting discussions; A. Keller, T. Alexiadis, E. Holderness, S. Polikovsky, J. Marquez for help with data acquisition; N. Mahmood and T. Zaman for help with video editing and voice recording; A. Quiros Ramirez for help with the project website.

REFERENCES

A. Balan and M. J. Black. 2008. The naked truth: Estimating body shape under clothing,. In *European Conf. on Computer Vision, ECCV (LNCS)*, Vol. 5304. Springer-Verlag, Marseilles, France, 15–29.

Kiran S. Bhat, Christopher D. Twigg, Jessica K. Hodgins, Pradeep K. Khosla, Zoran Popović, and Steven M. Seitz. 2003. Estimating Cloth Simulation Parameters from Video. In *Proceedings of the 2003 ACM SIGGRAPH/Eurographics Symposium on Computer Animation (SCA '03)*. Eurographics Association, Aire-la-Ville, Switzerland, Switzerland, 37–51. http://dl.acm.org/citation.cfm?id=846276.846282

Federica Bogo, Javier Romero, Gerard Pons-Moll, and Michael J. Black. 2017. Dynamic FAUST: Registering Human Bodies in Motion. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

K.L. Bouman, B. Xiao, P. Bataglia, and W.T. Freeman. 2013. Estimating the Material Properties of Fabrics from Videos. In *International Conference in Computer Vision (ICCV)*. 1984–1991.

Derek Bradley, Tiberiu Popa, Alla Sheffer, Wolfgang Heidrich, and Tamy Boubekeur. 2008. Markerless Garment Capture. *ACM Trans. Graphics (Proc. SIGGRAPH)* 27, 3 (2008), 99.

Rémi Brouet, Alla Sheffer, Laurence Boissieux, and Marie-Paule Cani. 2012. Design Preserving Garment Transfer. *ACM Transactions on Graphics* (Aug. 2012). http://hal.inria.fr/hal-00695903

Dan Casas, Marco Volino, John Collomosse, and Adrian Hilton. 2014. 4d video textures for interactive character appearance. *Computer Graphics Forum* 33, 2 (2014), 371–380.

Xiaowu Chen, Bin Zhou, Feixiang Lu, Lin Wang, Lang Bi, and Ping Tan. 2015. Garment Modeling with a Depth Camera. *ACM Trans. Graph.* 34, 6, Article 203 (Oct. 2015), 12 pages. https://doi.org/10.1145/2816795.2818059

Alvaro Collet, Ming Chuang, Pat Sweeney, Don Gillett, Dennis Evseev, David Calabrese, Hugues Hoppe, Adam Kirk, and Steve Sullivan. 2015. High-Quality Streamable Free-Viewpoint Video. *ACM Transactions on Graphics (SIGGRAPH)* 34, 4 (2015).

Radek Danecek, Endri Dibra, A. Cengiz Öztireli, Remo Ziegler, and Markus Gross. 2017. DeepGarment : 3D Garment Shape Estimation from a Single Image. *Computer Graphics Forum 36(2), Proceedings of the 38th Annual Conference of the European Association for Computer Graphics (Eurographics)* (2017).

Matthias Dantone, Juergen Gall, Christian Leistner, and Luc Van Gool. 2014. Body parts dependent joint regressors for human pose estimation in still images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 36, 11 (2014), 2131–2143.

Edilson de Aguiar, Leonid Sigal, Adrien Treuille, and Jessica K. Hodgins. 2010. Stable Spaces for Real-time Clothing. *ACM Trans. Graph.* 29, 4, Article 106 (July 2010), 9 pages. https://doi.org/10.1145/1778765.1778843

Edilson de Aguiar, Carsten Stoll, Christian Theobalt, Naveed Ahmed, Hans-Peter Seidel, and Sebastian Thrun. 2008. Performance Capture from Sparse Multi-view Video. *ACM Trans. Graph.* 27, 3, Article 98 (Aug. 2008), 10 pages. https://doi.org/10.1145/1360612.1360697

Mingsong Dou, Sameh Khamis, Yury Degtyarev, Philip Davidson, Sean Ryan Fanello, Adarsh Kowdle, Sergio Orts Escolano, Christoph Rhemann, David Kim, Jonathan Taylor, and others. 2016. Fusion4d: Real-time performance capture of challenging scenes. *ACM Transactions on Graphics (TOG)* 35, 4 (2016), 114.

Juergen Gall, Carsten Stoll, Edilson De Aguiar, Christian Theobalt, Bodo Rosenhahn, and Hans-Peter Seidel. 2009. Motion capture using joint skeleton tracking and surface estimation. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. IEEE, 1746–1753.

Russell Gillette, Craig Peters, Nicholas Vining, Essex Edwards, and Alla Sheffer. 2015. Real-Time Dynamic Wrinkling of Coarse Animated Cloth. In *Proc. Symposium on Computer Animation*.

Rony Goldenthal, David Harmon, Raanan Fattal, Michel Bercovier, and Eitan Grinspun. 2007. Efficient Simulation of Inextensible Cloth. *ACM Transactions on Graphics (Proceedings of SIGGRAPH 2007)* 26, 3 (2007).

P. Guan, L. Reiss, D. Hirshberg, A. Weiss, and M. J. Black. 2012. DRAPE: DRessing Any PErson. *ACM Trans. on Graphics (Proc. SIGGRAPH)* 31, 4 (July 2012), 35:1–35:10.

Nils Hasler, Carsten Stoll, Bodo Rosenhahn, Thorsten ThormÃďhlen, and Hans-Peter Seidel. 2009. Estimating body shape of dressed humans. *Computers & Graphics* 33, 3 (2009), 211 – 216. https://doi.org/10.1016/j.cag.2009.03.026 {IEEE} International Conference on Shape Modelling and Applications 2009.

Anna Hilsmann and Peter Eisert. 2009. Tracking and Retexturing Cloth for Real-Time Virtual Clothing Applications. In *Proceedings of the 4th International Conference on*

Computer Vision/Computer Graphics CollaborationTechniques (MIRAGE '09). Springer-Verlag, Berlin, Heidelberg, 94–105. https://doi.org/10.1007/978-3-642-01811-4_9

Peng Huang, Margara Tejera, John Collomosse, and Adrian Hilton. 2015. Hybrid skeletal-surface motion graphs for character animation from 4d performance capture. *ACM Transactions on Graphics (TOG)* 34, 2 (2015), 17.

Matthias Innmann, Michael Zollhöfer, Matthias Nießner, Christian Theobalt, and Marc Stamminger. 2016. VolumeDeform: Real-time Volumetric Non-rigid Reconstruction. In *Proceedings of European Conference on Computer Vision (ECCV)*. 17.

Arjun Jain, Thorsten Thormählen, Hans-Peter Seidel, and Christian Theobalt. 2010. MovieReshape: Tracking and Reshaping of Humans in Videos. *ACM Trans. Graph.* 29, 6, Article 148 (Dec. 2010), 10 pages. https://doi.org/10.1145/1882261.1866174

Sam Johnson and Mark Everingham. 2010. Clustered Pose and Nonlinear Appearance Models for Human Pose Estimation. In *Proceedings of the British Machine Vision Conference*. doi:10.5244/C.24.12.

Ladislav Kavan, Dan Gerszewski, Adam W. Bargteil, and Peter-Pike Sloan. 2011. Physics-inspired Upsampling for Cloth Simulation in Games. *ACM Trans. Graph.* 30, 4, Article 93 (July 2011), 10 pages. https://doi.org/10.1145/2010324.1964988

Doyub Kim, Woojong Koh, Rahul Narain, Kayvon Fatahalian, Adrien Treuille, and James F. O'Brien. 2013. Near-exhaustive Precomputation of Secondary Cloth Effects. *ACM Transactions on Graphics* 32, 4 (July 2013), 87:1–7. http://graphics.berkeley.edu/papers/Kim-NEP-2013-07/ Proceedings of ACM SIGGRAPH 2013, Anaheim.

Meekyoung Kim, Gerard Pons-Moll, Sergi Pujades, Sungbae Bang, Jinwwok Kim, Michael Black, and Sung-Hee Lee. 2017. Data-Driven Physics for Human Soft Tissue Animation. *ACM Transactions on Graphics, (Proc. SIGGRAPH)* 36, 4 (2017). http://dx.doi.org/10.1145/3072959.3073685

Christoph Lassner, Javier Romero, Martin Kiefel, Federica Bogo, Michael J. Black, and Peter V. Gehler. 2017. Unite the People: Closing the Loop Between 3D and 2D Human Representations. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. http://files.is.tuebingen.mpg.de/classner/up

X. Liang, C. Xu, X. Shen, J. Yang, J. Tang, L. Lin, and S. Yan. 2015. Human Parsing with Contextualized Convolutional Neural Network. In *Int. Conf. Comp. Vis. (ICCV)*.

Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J. Black. 2015. SMPL: A Skinned Multi-Person Linear Model. *ACM Trans. Graphics (Proc. SIGGRAPH Asia)* 34, 6 (Oct. 2015), 248:1–248:16.

E. Miguel, D. Bradley, B. Thomaszewski, B. Bickel, W. Matusik, M. A. Otaduy, and S. Marschner. 2012. Data-Driven Estimation of Cloth Simulation Models. *Comput. Graph. Forum* 31, 2pt2 (May 2012), 519–528. https://doi.org/10.1111/j.1467-8659.2012.03031.x

Alexandros Neophytou and Adrian Hilton. 2014. A layered model of human body and garment deformation. In *2014 2nd International Conference on 3D Vision*, Vol. 1. IEEE, 171–178.

Richard A Newcombe, Dieter Fox, and Steven M Seitz. 2015. Dynamicfusion: Reconstruction and tracking of non-rigid scenes in real-time. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 343–352.

Gerard Pons-Moll, Javier Romero, Naureen Mahmood, and Michael J. Black. 2015a. Dyna: A Model of Dynamic Human Shape in Motion. *ACM Transactions on Graphics, (Proc. SIGGRAPH)* 34, 4 (July 2015), 120:1–120:14.

Gerard Pons-Moll, Jonathan Taylor, Jamie Shotton, Aaron Hertzmann, and Andrew Fitzgibbon. 2015b. Metric Regression Forests for Correspondence Estimation. *International Journal of Computer Vision* (2015), 1–13.

Tiberiu Popa, Qingnan Zhou, Derek Bradley, Vladislav Kraevoy, Hongbo Fu, Alla Sheffer, and Wolfgang Heidrich. 2009. Wrinkling Captured Garments Using Space-Time Data-Driven Deformation. *Computer Graphics Forum (Proc. Eurographics)* 28, 2 (2009), 427–435.

Nadia Robertini, Edilson De Aguiar, Thomas Helten, and Christian Theobalt. 2014. Efficient Multi-view Performance Capture of Fine-Scale Surface Detail. In *Proceedings of the 2014 2Nd International Conference on 3D Vision - Volume 01 (3DV '14)*. IEEE Computer Society, Washington, DC, USA, 5–12. https://doi.org/10.1109/3DV.2014.46

K. Robinette, S. Blackwell, H. Daanen, M. Boehmer, S. Fleming, T. Brill, D. Hoeferlin, and D. Burnsides. 2002. *Civilian American and European Surface Anthropometry Resource (CAESAR) Final Report*. Technical Report AFRL-HE-WP-TR-2002-0169. US Air Force Research Laboratory.

Lorenz Rogge, Felix Klose, Michael Stengel, Martin Eisemann, and Marcus Magnor. 2014. Garment Replacement in Monocular Video Sequences. *ACM Transactions on Graphics* 34, 1 (Nov. 2014), 6:1–6:10.

Bodo Rosenhahn, Uwe Kersting, Katie Powell, Reinhard Klette, Gisela Klette, and Hans-Peter Seidel. 2007. A system for articulated tracking incorporating a clothing model. *Machine Vision and Applications* 18, 1 (2007), 25–40.

M. Sekine, K. Sugita, F. Perbet, B. Stenger, and M. Nishiyama. 2014. Virtual Fitting by Single-Shot Body Shape Estimation. In *Int. Conf. on 3D Body Scanning Technologies*. 406–413.

Leonid Sigal, Moshe Mahler, Spencer Diaz, Kyna McIntosh, Elizabeth Carter, Timothy Richards, and Jessica Hodgins. 2015. A Perceptual Control Space for Garment Simulation. *ACM Trans. Graph.* 34, 4, Article 117 (July 2015), 10 pages. https://doi.org/10.1145/2766971

Olga Sorkine. 2006. Differential Representations for Mesh Processing. *Computer Graphics Forum* 25, 4 (2006), 789–807.

Olga Sorkine, Daniel Cohen-Or, Yaron Lipman, Marc Alexa, Christian Rössl, and Hans-Peter Seidel. 2004. Laplacian Surface Editing. In *Proceedings of the EUROGRAPHICS/ACM SIGGRAPH Symposium on Geometry Processing*. ACM Press, 179–188.

Carsten Stoll, Juergen Gall, Edilson de Aguiar, Sebastian Thrun, and Christian Theobalt. 2010. Video-based Reconstruction of Animatable Human Characters. *ACM Trans. Graph.* 29, 6, Article 139 (Dec. 2010), 10 pages. https://doi.org/10.1145/1882261.1866161

Stephan Streuber, M. Alejandra Quiros-Ramirez, Matthew Q. Hill, Carina A. Hahn, Silvia Zuffi, Alice OâĂŹToole, and Michael J. Black. 2016. Body Talk: Crowdshaping Realistic 3D Avatars with Words. *ACM Trans. Graph. (Proc. SIGGRAPH)* 35, 4 (July 2016), 54:1–54:14.

D. Sun, E. Sudderth, and M. J. Black. 2010. Layered image motion with explicit occlusions, temporal consistency, and depth ordering. In *Advances in Neural Information Processing Systems 23 (NIPS)*. MIT Press, 2226–2234.

M. Tejera, D. Casas, and A. Hilton. 2013. Animation Control of Surface Motion Capture. *Cybernetics, IEEE Transactions on* 43, 6 (Dec 2013), 1532–1545. https://doi.org/10.1109/TCYB.2013.2260328

Daniel Vlasic, Ilya Baran, Wojciech Matusik, and Jovan Popović. 2008. Articulated mesh animation from multi-view silhouettes. *ACM Transactions on Graphics (TOG)* 27, 3 (2008), 97.

Timo von Marcard, Gerard Pons-Moll, and Bodo Rosenhahn. 2016. Human Pose Estimation from Video and IMUs. *Transactions on Pattern Analysis and Machine Intelligence PAMI* (Jan. 2016).

Timo von Marcard, Bodo Rosenhahn, Michael Black, and Gerard Pons-Moll. 2017. Sparse Inertial Poser: Automatic 3D Human Pose Estimation from Sparse IMUs. *Computer Graphics Forum 36(2), Proceedings of the 38th Annual Conference of the European Association for Computer Graphics (Eurographics)* (2017).

Huamin Wang, Florian Hecht, Ravi Ramamoorthi, and James F. O'Brien. 2010. Example-Based Wrinkle Synthesis for Clothing Animation. *ACM Transactions on Graphics* 29, 4 (July 2010), 107:1–8. http://graphics.berkeley.edu/papers/Wang-EBW-2010-07/ Proceedings of ACM SIGGRAPH 2010, Los Angles, CA.

Huamin Wang, James F. O'Brien, and Ravi Ramamoorthi. 2011. Data-Driven Elastic Models for Cloth: Modeling and Measurement. *ACM Transactions on Graphics, Proc. SIGGRAPH* 30, 4 (July 2011), 71:1–11.

Ruizhe Wang, Lingyu Wei, Etienne Vouga, Qixing Huang, Duygu Ceylan, Gerard Medioni, and Hao Li. 2016. Capturing Dynamic Textured Surfaces of Moving Targets. In *Proceedings of the European Conference on Computer Vision (ECCV)*.

Ryan White, Keenan Crane, and D. A. Forsyth. 2007. Capturing and Animating Occluded Cloth. *ACM Trans. Graph.* 26, 3, Article 34 (July 2007). https://doi.org/10.1145/1276377.1276420

Chenglei Wu, Kiran Varanasi, and Christian Theobalt. 2012. Full Body Performance Capture Under Uncontrolled and Varying Illumination: A Shading-based Approach. In *Proceedings of the 12th European Conference on Computer Vision - Volume Part IV (ECCV'12)*. Springer-Verlag, Berlin, Heidelberg, 757–770. https://doi.org/10.1007/978-3-642-33765-9_54

Stefanie Wuhrer, Leonid Pishchulin, Alan Brunton, Chang Shu, and Jochen Lang. 2014. Estimation of human body shape and posture under clothing. *Computer Vision and Image Understanding* 127 (2014), 31–42.

Feng Xu, Yebin Liu, Carsten Stoll, James Tompkin, Gaurav Bharaj, Qionghai Dai, Hans-Peter Seidel, Jan Kautz, and Christian Theobalt. 2011. Video-based Characters: Creating New Human Performances from a Multi-view Video Database. *ACM Trans. Graph.* 30, 4, Article 32 (July 2011), 10 pages. https://doi.org/10.1145/2010324.1964927

Jinlong Yang, Jean-Sébastien Franco, Franck Hétroy-Wheeler, and Stefanie Wuhrer. 2016. Estimation of Human Body Shape in Motion with Wide Clothing. In *European Conference on Computer Vision 2016*. Amsterdam, Netherlands. https://hal.inria.fr/hal-01344795

Chao Zhang, Sergi Pujades, Michael Black, and Gerard Pons-Moll. 2017. Detailed, accurate, human shape estimation from clothed 3D scan sequences. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Bin Zhou, Xiaowu Chen, Qiang Fu, Kan Guo, and Ping Tan. 2013. Garment Modeling from a Single Image. *Computer Graphics Forum* (2013). https://doi.org/10.1111/cgf.12215