Parametric Reshaping of Human Bodies in Images

Shizhe Zhou[†]   Hongbo Fu[‡]   Ligang Liu[†*]   Daniel Cohen-Or[§]   Xiaoguang Han[†]

[†]Zhejiang University   [‡]City University of Hong Kong   [§]Tel Aviv University

Figure 1: *Our parametric reshaping technique allows users to easily reshape a human body in a single image (Leftmost) by simply manipulating a small set of sliders corresponding to semantic attributes such as height, weight and waist girth.*

Abstract

We present an easy-to-use image retouching technique for realistic reshaping of human bodies in a single image. A *model-based* approach is taken by integrating a 3D whole-body morphable model into the reshaping process to achieve globally consistent editing effects. A novel *body-aware image warping* approach is introduced to reliably transfer the reshaping effects from the model to the image, even under moderate fitting errors. Thanks to the parametric nature of the model, our technique parameterizes the degree of reshaping by a small set of semantic attributes, such as weight and height. It allows easy creation of desired reshaping effects by changing the full-body attributes, while producing visually pleasing results even for loosely-dressed humans in casual photographs with a variety of poses and shapes.

Keywords: Image Manipulation, Portrait Retouching, Warping

1   Introduction

With professional image editing packages such as Adobe Photoshop[©], users enjoy the art of portrait retouching and enhancing. However, the existing retouching tools are mainly designed for low-level editing tasks, and thus are often limited to local modification of skins such as blemish and wrinkle removal. Retouching tasks such as subtle reshaping of human bodies require *global consistency* of editing, demanding professional skills, and are still a tedious task even for skillful users. In fact, it usually takes a professional artist considerable effort by employing hundreds of local editing operations in order to achieve a visually pleasing global change [1]. Moreover, since this time-consuming editing process is not parametric, it cannot be reused for alternative retouching tasks.

A possible solution for high-level editing of human bodies is to operate at the level of body parts, and change each part with respect to its underlying skeleton, which is, however, challenging due to the following reasons. First, a realistic reshaping effect demands a spatially-varying deformation *within* individual body parts (Figure 1), preventing a simple scaling of body parts along the skeletal bone axes or their perpendicular directions. Second, it is unclear how to make the changes introduced to individual parts globally coherent, especially for regions with occlusion.

To address these problems, we take a *model-based* approach for reshaping human bodies in a single image. We rely on a 3D morphable model of human shapes to achieve globally consistent editing and desired spatially-varying deformation within and across individual body parts. Our approach is based on the following two key observations. First, while existing works for estimating 3D faces or human shapes from images aim for a faithful 3D shape reconstruction, our target is image-based, which permits merely a *view-dependent matching* between the 3D model and the image. Second, the retouching process by professional artists indicates that changes of human bodies in images are *largely governed by changes of 2D body contours with respect to skeletons*. These two

---

[1]Please visit our project pages for the video links to interactive editing sessions by professional artists:

At CityU: http://sweb.cityu.edu.hk/hongbofu/projects/ParametricBodyReshaping/
At ZJU: http://www.math.zju.edu.cn/ligangliu/CAGD/Projects/HumanReshaping/

observations allow formulation of the reshaping problem in the image domain.

We pose the body reshaping problem as the design of a 2D warping of a human body image, and introduce a novel *body-aware* image warping approach incorporating changes of the morphable model, while being tolerant to moderate errors in view-dependent matching. Our system follows the steps as shown in Figure 2. We first roughly match the morphable model to the image with a moderate amount of user interaction (Section 4). The user can then resculpt the matched 3D human model by tuning a small set of semantic attributes such as weight and height. We finally deform the image human shape in a body-aware manner so that its changes with respect to its skeleton, especially on the body contours, are as similar as possible to the corresponding changes at the 3D human model (Section 5). Working in image space avoids the difficult problem of accurate reconstruction of the 3D model and its texture, making our technique applicable to the reshaping of humans wearing loose garments under arbitrary poses without any additional difficulty.

The parametric nature of our reshaping technique enables users to easily achieve various reshaping results by simply manipulating a small set of sliders corresponding to the set of semantic attributes (Figure 1). Our fast warping algorithm provides real-time manipulation feedback, allowing users to obtain desired reshaping effects with ease (see interactive editing sessions in the accompanying video). It is demonstrated that our technique is effective for human bodies with a variety of poses, shapes and garments in both casual and professional photographs.

2 Related Work

Image Retouching. Retouching images often involves a set of image editing tools, e.g., tone adjustment, recoloring, image composition, image repairing, image denoising, image warping etc. Each of these editing tools has been widely studied and reviewing them is beyond our scope. Most existing retouching tools operate at the pixel level and are competent for low-level editing tasks for instance blemish and wrinkle removal [Eismann and Palmer 2006]. However, they become very awkward for high-level editing tasks such as face and body reshaping, because they have to rely on users to maintain the coherence of editing operations. High-level retouching has seldom been studied due to its difficult nature: it needs to deal with the semantics of objects being edited [Barrett and Cheney 2002].

Image Warping and Resizing. There exist many image warping methods allowing users to manipulate an image through a set of control handles specified on it. Deformation propagation from control handles to the rest of the image can be achieved by either interpolation methods (e.g., Radial Basis Functions (RBF) [Arad and Reisfeld 1995] or Moving Least Squares (MLS) [Schaefer et al. 2006]) or optimization methods (e.g., as-rigid-as-possible shape manipulation [Igarashi et al. 2005]). Recently, some image warping methods have been successfully extended to address the problem of content-aware resizing of images or videos (see an insightful survey in [Shamir and Sorkine 2009]). To some extent, our proposed warping approach can be considered as another resizing method which, however, "resizes" a human body figure along its skeleton axes with spatially-varying scaling factors.

Data-Driven Face Retouching. Data-driven methods have been proved effective for easy reshaping of human faces in images. Blanz and Vetter [1999] present a 3D morphable face model, which enables a few photo-realistic image manipulations, including the reshaping of faces in single images (e.g., simulation of weight gain or loss). It is achieved by first reconstructing a textured 3D face from a face image using the morphable model and then directly rendering modified versions of the 3D face back into the image. As we shall demonstrate later, a naïve extension of this idea to reshaping human bodies in images requires *accurate* recovery of both 3D human shape and texture from a single image. Even small errors can easily lead to noticeable editing artifacts. Since human faces have no articulated structure, most recent face retouching works simply resort to a training set of 2D facial images instead of 3D face models. For example, Leyvand et al. [2008] present a technique to increase the attractiveness of faces by deforming the facial shape of an input image using a 2D warping function learned from a training set of facial images.

3D Whole-Body Morphable Models. Recent advances on the analysis of 3D human shape space have paved the way for our research. Some early researches [Allen et al. 2003; Seo and Magnenat-Thalmann 2004] extend the morphable face model to capture full-body shape variations across different people, but only in similar poses. These methods provide a direct way to explore the human shape space with intuitive controls. Several recent morphable models [Anguelov et al. 2005; Allen et al. 2006; Hasler et al. 2009] encode the variability of body shape in terms of both pose and identity, enabling the generation of a 3D human body shape of any identity in any pose. Among them, the SCAPE model proposed by Anguelov et al. [2005] supports decoupled parameters for pose- and identity-dependent deformations, allowing efficient optimization for them.

Human Pose from Images. 3D pose estimation from a single image has been widely studied in the field of computer vision. Many automatic pose estimation methods have been proposed (see [Hua et al. 2009] and references therein). However, the ill-posed nature of this problem often requires a certain amount of user assistance (e.g., manually specifying joint correspondences between a predefined skeletal structure and the image) to obtain more reasonable recovered poses [Taylor 2000; Davis et al. 2003; Parameswaran and Chellappa 2004; Hornung et al. 2007].

Body Shape from Images. Only a few techniques for estimating detailed human shape from images exist due to the challenges posed by the articulated and non-rigid nature of the human form. Most existing techniques rely on multi-view video recordings to capture dynamic human shapes. They match either a pre-scanned template model of a known subject [de Aguiar et al. 2008; Vlasic et al. 2008] or the SCAPE model [Balan et al. 2007; Balan and Black 2008] to every video frame. However, it is generally insufficient and difficult to constrain a 3D body shape from a single image. Recently, Guan et al. [2009] were the first to study the problem of estimating detailed body shape from a single image. Given an initial guess of the 3D pose and shape with user assistance, a solution for the parameters of the SCAPE model is found using a complex optimization which encodes a variety of cues including silhouette overlap, height constraint and smooth shading. Their method is not meant to deal with clothing, but mainly naked or minimally dressed humans.

Our work is also related to view-dependent modeling of general 3D shapes, more specifically, shape modeling from single images with the assistance of a single 3D template model [Kraevoy et al. 2009; Tan et al. 2010]. Like [de Aguiar et al. 2008; Vlasic et al. 2008], these methods deform the template model to match target silhouettes or contours. We use the idea of Kraevoy et al. [2009] to fit a 3D morphable model to a 2D body contour in the shape fitting step. While they aim to reshape the model with respect to the image, our ultimate goal is to reshape the figure with respect to the 3D morphable model.

Deformation transfer is often used to retarget detailed animation from one model to another. Transfer can take place between 3D
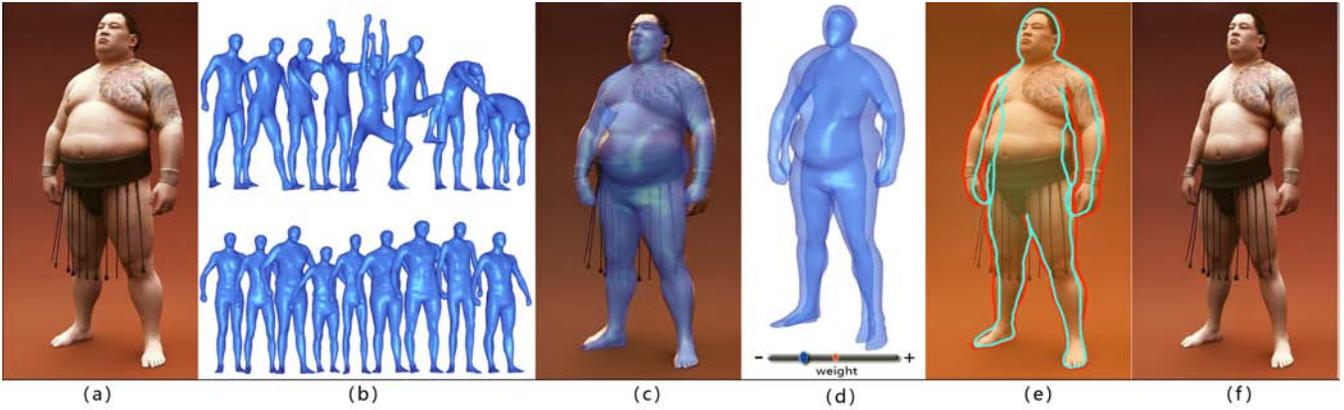
Figure 2: *System overview. Our system first learns a 3D whole-body morphable model from a database of 3D full-body scans (b), including pose (Top) and shape (Bottom) data sets. Users then help to achieve view-dependent fitting between the morphable model and input human image (a), resulting in (c). The fitted model is reshaped by manipulating semantic attributes like weight (d). The reshaping result (f) is finally obtained by the body-aware image warping driven by the changes introduced to the model, especially at its projected contours (e).*

meshes [Sumner and Popović 2004], from cartoon animations to 3D meshes [Zhou et al. 2005], or from videos to videos [Vlasic et al. 2005].Aiming for a pose-driven animation with shape preservation, Hornung et al. [2007] present a character animation system by transferring 3D mocap data to a 2D character in a single image. However, to the best of our knowledge, no work has been done on transferring detailed shape deformation from 3D models to human subjects in images.

3 Overview

A reshaping of a human body in an image requires a content-aware image warping. Different parts of the human body deserve different degrees of modification to yield a globally consistent and visually pleasing result. Digital artists interactively employ a series of local editing operations to achieve this goal. Instead, we present a *fit-and-adjust* method that utilizes a 3D whole-body morphable model to guide a parametric reshaping of human bodies in single images. See Figure 2 for an overview of our method.

We adopt SCAPE [Anguelov et al. 2005] as our morphable model and fit it to an image of a human body (Section 4). This requires solving for the pose and shape parameters of the morphable model subject to image data. We rely on user assistance to solve this ill-posed problem. We design a user interface that requires only a rather small amount of initial user input, and allows interactive refinement if needed, demonstrated in the accompanying video. To achieve interactive rate, we trade fitting accuracy for speed by simplifying the complex fitting optimization of [Guan et al. 2009].

Next, the user specifies the desired attribute values for warping the human body in the image (Section 5). In response, the system first deforms the fitted 3D model according to the semantic attribute changes and then uses the changes of the model with respect to its 3D skeleton, especially at its projected contour (Figure 2(e)), to guide the image warping.

Specifically, we introduce a body-aware image warping approach which coherently resizes image body parts along directions parallel and orthogonal to their bone axes subject to *length changes* along the corresponding directions as dictated by the morphable model. This warping algorithm runs efficiently and supports real-time manipulation feedback (see the accompanying video).

4 View-Dependent Model Fitting

In this section we present the technical details of matching the morphable model with a human shape in a single image and begin by briefly reviewing the morphable model method.

4.1 3D Whole-Body Morphable Model

The SCAPE model [Anguelov et al. 2005] is adopted as our morphable model since it allows efficient solving for pose and shape from images. SCAPE is a low-dimensional, parametric model, which has separate sets of parameters for pose- and shape-dependent deformations. More specifically, pose-driven deformations are parameterized by a set of 3D rotations of skeletal bones, denoted as θ. Shape variations among individuals are characterized by shape parameters β, a coefficient vector corresponding to a point in a low-dimensional shape space obtained by Principal Component Analysis (PCA). The SCAPE model, denoted as $M(\theta, \beta)$, leads to a 3D human body shape of any identity in any pose by specifying proper values for θ and β. Refer to [Anguelov et al. 2005] for more details.

We follow the same approach of [Anguelov et al. 2005] to learn the parametric function $M(\theta, \beta)$ from a database of registered dense 3D full-body scans. Since Anguelov et al. [2005] do not make their full data set available to the public, we use a publicly available database of full body scans [Hasler et al. 2009] (see a small portion of the database in Figure 2(b)). Moreover, this database provides per-scan semantic attributes like height, weight, waist girth, leg lengths etc., which benefits the manipulation of the learning model using the semantic attributes, as will be introduced shortly. We rig a standard 3D skeleton template with 18 joints and 17 bones into each scanned model [Baran and Popović 2007] to learn $M(\theta, \beta)$ in terms of θ. As shape variations for female and male subjects are different, we learn an individual model for each gender.

Although PCA provides a compact human shape space, it does not offer a direct way to explore the space with intuitive attributes like weight and height. To address this problem, we use a linear regression method by Allen et al. [2003] to learn a linear mapping: $\beta = f(\gamma)$, where $\gamma = [\gamma_1, \ldots, \gamma_l, 1]^\top$ with γ_i representing semantic attribute values of an individual. As the original attribute values associated with a subject in the image are generally unknown, for reshaping we let the user specify desired *attribute*
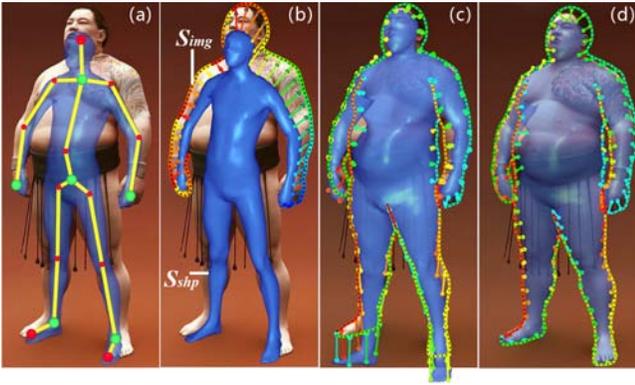
Figure 3: *View-dependent model fitting. (a) A 3D pose recovered with user assistance. (b) Optimal correspondence (shown in arrows) established between the image contour S_{img} and 3D body projected contour S_{shp}. The 3D model here is a mean body shape associated with the recovered pose. (c) Fitted shape using the contours of upper body in (b). (d) Updated fitted shape by using the additional contours of lower body in (c).*

changes instead of absolute target values for the attributes. Let $\Delta\gamma = [\Delta\gamma_1, \ldots, \Delta\gamma_l, 0]^\top$ denote an attribute offset vector, where each $\Delta\gamma_i$ is the change of value introduced to attribute γ_i. We can then manipulate an initial human shape, represented by $M = M(\theta, \beta)$, to generate a new shape, $M^{\Delta\gamma} = M(\theta, \beta + f(\Delta\gamma))$, which reflects the changes of the semantic attributes. For example, if γ_1 corresponds to the weight attribute, we can make the subject gain or lose weight by $20kg$ by setting $\Delta\gamma = [\pm 20, 0, \ldots, 0, 0]^\top$ (Figure 2(d)).

Our first key task is to fit the learned female/male morphable model $M(\theta, \beta)$ to a female/male human body in a single image. Here we use $M(\theta, \beta)$ instead of $M(\theta, f(\gamma))$, since the former has more modes of shape variations (e.g., $\beta \in \mathbf{R}^{20}$ but $\gamma \in \mathbf{R}^{10}$ in our case) and thus can better characterize image data. We employ a similar method to that of Guan et al. [2009], but with two major differences. First, as the shading cues are generally not very reliable and have little influence on the quality of *view-dependent matching*[2] [Guan et al. 2009], we focus on contour-related cues, which are easily accessible with user assistance, even for subjects with loose garments. Second, instead of solving for the pose and shape parameters simultaneously using a slow randomized search, we trade accuracy for speed and solve for θ and β successively.

4.2 Pose Fitting

We first solve for the pose parameters θ. For simplicity, we follow the semi-automatic method of [Taylor 2000] to find a best-fit 3D pose under the assumption of weak-perspective camera projection: given the 3D skeleton template, 3D orientations of skeletal bones are recovered by matching joint correspondences between markups in the skeleton template and the image (Figure 3(a)). For many images, we can rely on exposed skin regions to reduce the number of manually specified joint correspondences [Hua et al. 2009]. Otherwise, the user has to input more joints. The user is also allowed to interactively adjust the camera scale factor [Taylor 2000] and the estimated 3D pose with a user-friendly interface like [Davis et al. 2003]. Given our target of view-dependent matching, it is unneces-

---

[2]From Figure 8 of [Guan et al. 2009] (the only comparison figure for shading cues), it is seen that the visual difference between results with and without shading cues is not very obvious from the *camera view*.

sary to explicitly address the pose ambiguity problem [Guan et al. 2009], since ambiguous poses are noticeable only from viewing angles that are different from the camera view (Figure 4).
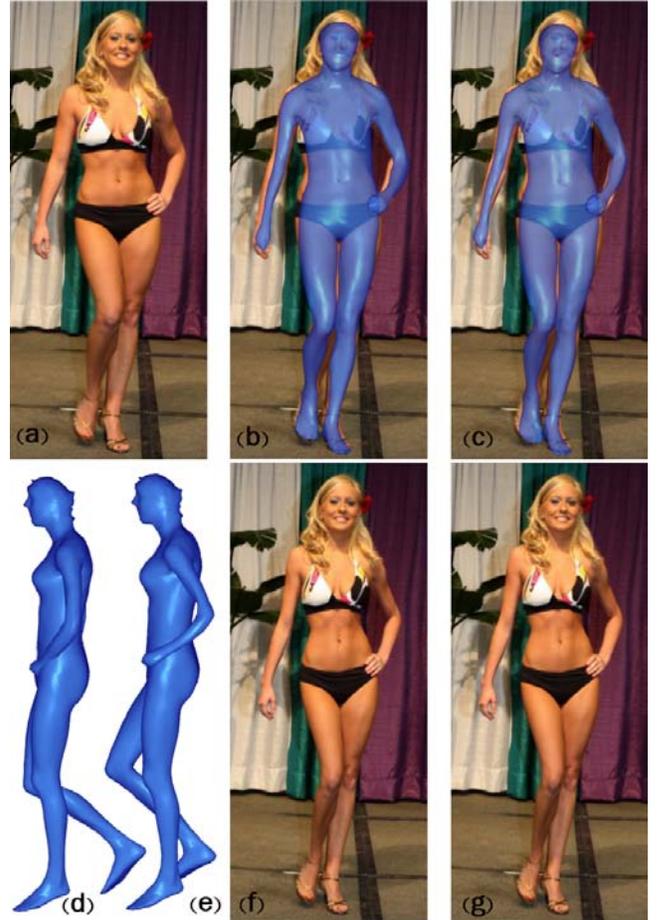
Figure 4: *Our approach is insensitive to pose ambiguity. Two shapes (b and c) fit almost equally well with an input image (a), though they exhibit different poses (d and e). Our technique produces visually the same reshaping results (f and g).*

4.3 Shape Fitting

In this step, we solve for the shape parameters β given the pose parameters θ^* obtained at the previous step. We seek for a 3D human shape $M(\theta^*, \beta)$ which best matches image observations. Since body contour information can be robustly measured in images and largely depicts the global body shape, we define the matching errors as the difference between the image subject contour S_{img} and the projected contour S_{shp} of $M(\theta^*, \beta)$ (Figure 3(b)). The contour S_{img} generally includes silhouettes, self-occlusion edges, and invisible occluded contour edges. S_{img} is obtained by standard interactive segmentation tools like lazy snapping [Li et al. 2004] and intelligent scissors [Mortensen and Barrett 1995]. Note that due to the low-dimensional nature of $M(\theta^*, \beta)$, users need not specify all contour edges at once for optimizing β and can incrementally provide more interaction to improve fitting accuracy (see Figure 3(c,d)). This also implies that our fitting approach works well for images with moderate occlusions, e.g., caused by various clothes.

Given the fixed θ^*, the morphable model $M(\theta^*, \beta)$ reduces to

a linear function of β. A variant of the *match-and-deform* method [Kraevoy et al. 2009] is used to find an optimal value for β such that its corresponding projected contour best matches S_{img}. The process starts with a 3D human body shape $M^0 = M(\theta^*, \beta^0)$, where β_0 represents the mean body shape parameters (Figure 3(a)). Let S_{shp}^0 be the projected contour of M^0. Like [Kraevoy et al. 2009], we first establish optimal correspondences between S_{shp}^0 and S_{img} (Figure 3(b)), and then move each vertex on S_{shp}^0 to its corresponding position on S_{img} with an adaptive step size. Let S_{shp}^1 denote the modified version of S_{shp}^0. Since under weak-perspective camera projection, there is a linear relationship between the vertices of S_{shp} and β, given S_{shp}^1, we can easily update β^0 as β^1. The above process is repeated to get the optimal shape parameters, denoted as β^*. The linearity of $M(\theta^*, \beta)$ in β makes the above optimization run at interactive rate. In some cases it was found that where the above optimization cannot produce desirable shape fitting results, it was very helpful to refine the fitted shape by slightly tuning some semantic attributes.

Since pose and shape are generally interdependent [Hasler et al. 2009], ideally the above pose and shape fitting steps should be repeated in an alternating manner to obtain a globally more optimal solution. However, as will be shown shortly, our warping method can tolerate fitting errors and does not require an accurate fit. In fact, our experiments show that $M(\theta^*, \beta^*)$ is usually already sufficient for our editing purpose in many cases. In few cases where the pose and the shape match poorly, the user needs to manually adjust the joint locations to initiate a new iteration for computing θ and β.

5 Body-Aware Image Warping

Let $M^* = M(\theta^*, \beta^*)$ be the fitted 3D human shape obtained in the previous step (Figure 2(c)) and $M^{\Delta\gamma} = M(\theta^*, \beta^* + f(\Delta\gamma))$ the reshaped model of M^* under the changes of semantic parameters $\Delta\gamma$ (Figure 2(d)). In this section we introduce our novel body-aware image warping approach which reliably transfers the reshaping effects from the model to the image. It is important to first clarify why neither straightforward model-based solutions nor conventional image-warping methods will yield satisfactory results.

5.1 Straightforward Retouching Solutions

Following the retouching idea based on the morphable face model [Blanz and Vetter 1999], a straightforward solution to produce a retouched image corresponding to $M^{\Delta\gamma}$ is to use a *texture-mapping based method*. That is, first texture M^* using user-guided texture mapping methods like FlexiStickers [Tzur and Tal 2009] (Figure 5(d)) and then *directly project* the textured model $M^{\Delta\gamma}$ back onto the image (Figure 5(e)). However, such methods apparently require visually unnoticeable texture distortion during texture mapping, since the textured model needs directly rendered as final results. Any incompatibility either among the user-guided correspondence set or between the 3D model and the underlying shape in the image would cause texture distortion. Small distortion can already lead to visual artifact even *before* reshaping (see a direct rendering of the textured model M^* in Figure 5(d)). In addition, it is difficult for such solutions to handle off-body parts like loosely-dressing garments and hair.

Another possible solution is to use the *changes* between M^* and $M^{\Delta\gamma}$ in the image space to drive a standard image warping (e.g., RBF method [Arad and Reisfeld 1995] or MLS method [Schaefer et al. 2006]). Such warping-based methods allow the propagation of retouching effects from the human body to parts which are not modeled, including clothes and hair. A straightforward way to define the changes of M^* and $M^{\Delta\gamma}$ projected in the image space is to
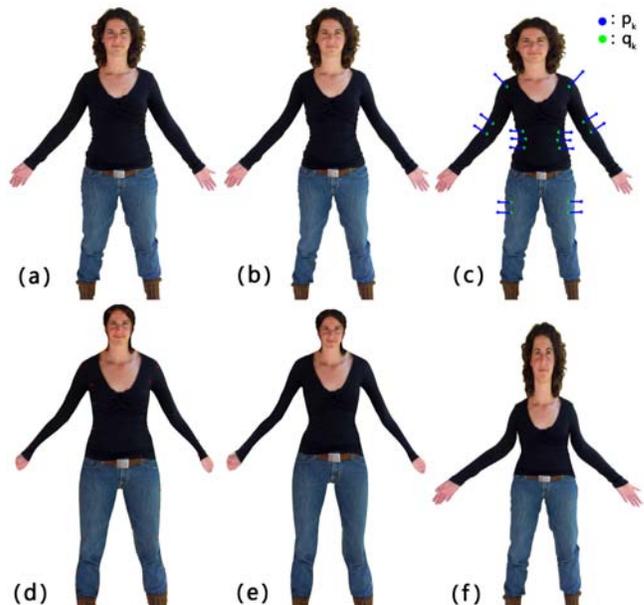
Figure 5: *Comparisons with straightforward solutions. (a) Input image. (b) Body-aware warping. (d, e) Texture mapping with Flexi-iStickers and its reshaping result. Overlaid red points in (d) are correspondence markups for texture mapping. (c, f) Body-oblivious warping with RBF and MLS, respectively. The same set of positional constraints (i.e., $\{\mathbf{p}_k \to \mathbf{q}_k\}$) are used for both (c) and (f) but shown only in (c) for brevity.*

use the *absolute* position offsets, i.e., $\mathbf{q}_k - \mathbf{p}_k$, where $\mathbf{p}_k \in \mathbf{R}^2$ and $\mathbf{q}_k \in \mathbf{R}^2$ are the projected positions of vertices of M^* and their corresponding vertices of $M^{\Delta\gamma}$, respectively (Figure 5(c)). These offsets can then be applied to corresponding positional constraints defined at $\mathbf{p}_k$ to induce image warping. However, since such use of the changes of M^* and $M^{\Delta\gamma}$ relies on the absolute positions $\mathbf{p}_k$, the retouching quality would still heavily depend on the fitting quality. In addition, since conventional image warping methods are *body-oblivious*, they may fail to preserve human shape features under dense sets of positional constraints possibly due to the over-fitting problem (Figure 5(c,f)).

5.2 Body-Aware Image Warping

We observed that many reshaping effects, like body height changes, directly exhibit length changes. In fact, reshaping of a human body shape largely means resizing of body parts either along their corresponding skeletal bone axes, denoted as $\mathbf{d}_{ske}$, or along their orthogonal directions, $\mathbf{d}_{ske\perp}$. For a simple example, resizing all body parts along $\mathbf{d}_{ske\perp}$ can simulate weight gain or loss. Similarly, resizing along $\mathbf{d}_{ske}$ can simulate increase or decrease of height. This motivated us to design a skeleton-aware image resizing approach and to use *changes of length* along directions $\mathbf{d}_{ske}$ and $\mathbf{d}_{ske\perp}$ on M^* to induce resizing of image body parts. Note that unlike traditional image resizing, which always has a single scale change for x-axis or y-axis, body reshaping exhibits spatially varying length changes with respect to skeletal bones even within a single body part (e.g., a calf in Figure 6(left)).

To guarantee coherent resizing effects across body parts, we embed the image into a 2D triangular mesh (Figure 6(right)), denoted as $\mathcal{G}$, and use this mesh to drive image warping. The mesh $\mathcal{G}$ is constructed by performing a constrained Delaunay triangulation

over a set of sampled points on the contours and image boundaries. Steiner vertices are automatically introduced to guarantee the high-quality of the mesh. Subtle changes of subject contour S_{img}, especially along $\mathbf{d}_{ske\perp}$, should be faithfully represented in the deformation of $\mathcal{G}$, demanding sufficient mesh vertices to reside along S_{img}. Hence, we sample hundreds of pairs of points on S_{img}, denoted as S_{pair}. Specifically, such pairs are obtained by first uniformly sampling pairs of vertices on the shape contour S_{shp}, which are roughly perpendicular to associated 3D bone axes[3] (Figure 6(left)) and then finding the corresponding positions on S_{img} (Figure 6(right)) through the contour mapping between S_{shp} and S_{img} (Section 4.3). A 2D affine transformation $\mathbf{A}_i$ is used to model the deformation of each planar triangle of $\mathcal{G}$, which is linearly represented in terms of the unknown vertex positions of $\mathcal{G}$ after warping, denoted as $\mathbf{V}$.
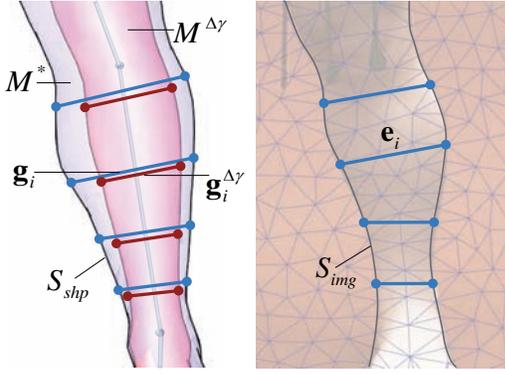
Figure 6: *An illustration for modeling length changes along directions perpendicular to bone axes (i.e., $E_{ske\perp}$). For brevity, the changed version of $\mathbf{e}_i$ (i.e., $\mathbf{e}_i^{\Delta\gamma}$) is not visualized here.*

We formulate an optimization to warp $\mathcal{G}$, whose objective function consists of five energy terms. Among them, three energy terms (i.e., $E_{ske\perp}$, E_{ske} and E_{sil}) are tailored specifically to our problem, which intend to minimize relative length changes along corresponding directions between 2D and 3D during reshaping. Those three terms have a common form as follows:

$$E_{\Delta len}(P) = \sum_i \left\| \mathbf{e}_i^{\Delta\gamma}(\mathbf{V}) - \frac{\|\mathbf{g}_i^{\Delta\gamma}\|}{\|\mathbf{g}_i\|}\mathbf{e}_i \right\|^2, \quad (1)$$

where $P = \{(\mathbf{e}_i, \mathbf{g}_i)\}$ is a set of edge pairs with $\mathbf{e}_i \in \mathbf{R}^2$ and $\mathbf{g}_i \in \mathbf{R}^3$ denoting an edge vector in the original image space and the corresponding edge from M^*, respectively (Figures 6, 7 and 8). $\mathbf{e}_i^{\Delta\gamma}$ and $\mathbf{g}_i^{\Delta\gamma}$ are the changed versions of $\mathbf{e}_i$ and $\mathbf{g}_i$ during reshaping, respectively. Therefore, $\mathbf{g}_i^{\Delta\gamma}$ can be directly obtained from $M^{\Delta\gamma}$, but $\mathbf{e}_i^{\Delta\gamma}$ is an unknown vector, represented in terms of $\mathbf{V}$.

Energy Term $E_{ske\perp}$. To map the changes of body parts *along* $\mathbf{d}_{ske\perp}$ from the 3D model to the image subject, we design the first energy term $E_{ske\perp} = E_{\Delta len}(\{(\mathbf{e}_i, \mathbf{g}_i)\})$, where $\mathbf{e}_i$ are edge vectors defined by pairs of points in S_{pair} used when generating $\mathcal{G}$ (Figure 6 right), and $\mathbf{g}_i$ are the corresponding edge vectors on S_{shp} (Figure 6 left). Since every point involved in S_{pair} has a corresponding vertex in $\mathcal{G}$, $\mathbf{e}_i^{\Delta\gamma}(\mathbf{V})$ can be simply represented as the difference between two unknown vertex positions.

Energy Term E_{sil}. To avoid drastic deviation of the new subject contour from the original contour, another energy term is introduced which considers length changes *along the contours*: $E_{sil} =$

---

[3]We do not sample points at the regions near to skeletal nodes with a degree larger than two, since the perpendicularity there is not well defined.
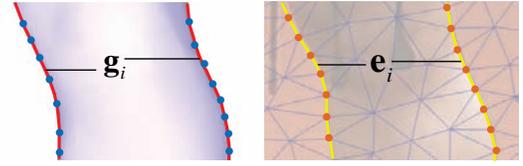
Figure 7: *An illustration for modeling length changes along the contours (i.e., E_{sil}).*

$E_{\Delta len}(\{(\mathbf{e}_i, \mathbf{g}_i)\})$, where $\mathbf{e}_i$ are defined by adjacent vertices of $\mathcal{G}$ along S_{img} and $\mathbf{g}_i$ are the corresponding vectors defined on S_{shp} (Figure 7).
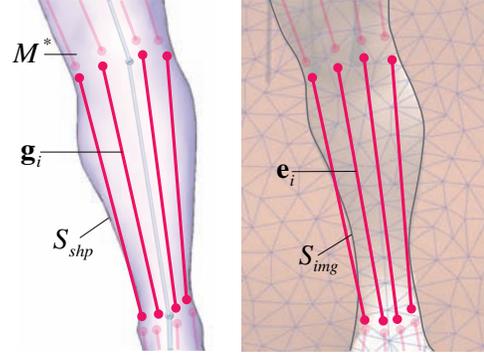
Figure 8: *An illustration for modeling length changes along bone axes (i.e., E_{ske}).*

Energy Term E_{ske}. To model resizing effects *along the bone axis of each body part*, we need to find a set of point pairs which define directions roughly along the bone axis. This is done by sampling pairs of points along end boundaries of pre-segmented 3D body parts such that each sampled pair is roughly parallel to its corresponding 3D bone axis (Figure 8(left)). To address the visibility problem, we discard pairs if both the associated endpoints are hidden in the image view. Denote $\mathbf{g}_i$ the edge vectors defined by the resulting pairs of points, and $\mathbf{e}_i$ the projected edges of $\mathbf{g}_i$ onto image space (Figure 8(right)). We then define the third energy term as $E_{ske} = E_{\Delta len}(\{(\mathbf{e}_i, \mathbf{g}_i)\})$. Since the endpoints of $\mathbf{e}_i$ might be located within triangles of $\mathcal{G}$, we define $\mathbf{e}_i^{\Delta\gamma}(\mathbf{V})$ as a linear interpolation of $\mathbf{V}$ based on barycentric coordinates.

Optimization. Apart from the energy terms E_{ske}, $E_{ske\perp}$ and E_{sil}, we use two relatively common energy terms E_{reg} and E_{dis} and solve for $\mathbf{V}$ by minimizing the following objective function:

$$\omega_{ske}E_{ske} + \omega_{ske\perp}E_{ske\perp} + \omega_{sil}E_{sil} + \omega_{reg}E_{reg} + \omega_{dis}E_{dis},$$

subject to positional constraints on the image borders, as common in warping-based image resizing [Shamir and Sorkine 2009]. Here $E_{reg} = \sum_i \sum_{j \in \mathcal{N}(i)} \|\mathbf{A}_i - \mathbf{A}_j\|_F^2$ is a regularization term which minimizes the deformation difference between every pair of adjacent triangles. $E_{dis} = \sum_i u_i \|\mathbf{A}_i - \mathbf{I}\|_F^2$ penalizes the deviation of deformation from the identity transformation $\mathbf{I}$. The degree of penalization is controlled by weights u_i, which are computed as content saliency: regions with high saliency (e.g., face and hair) deserve high weights to prevent serious distortion. Similar to media resizing works [Shamir and Sorkine 2009], our system supports both automatic and interactive generation of saliency maps. The resulting optimization is a quadratic function of $\mathbf{V}$, leading to a sparse linear system. Since the changes of the semantic attributes (i.e., corresponding to length changes) only influence the right hand side of
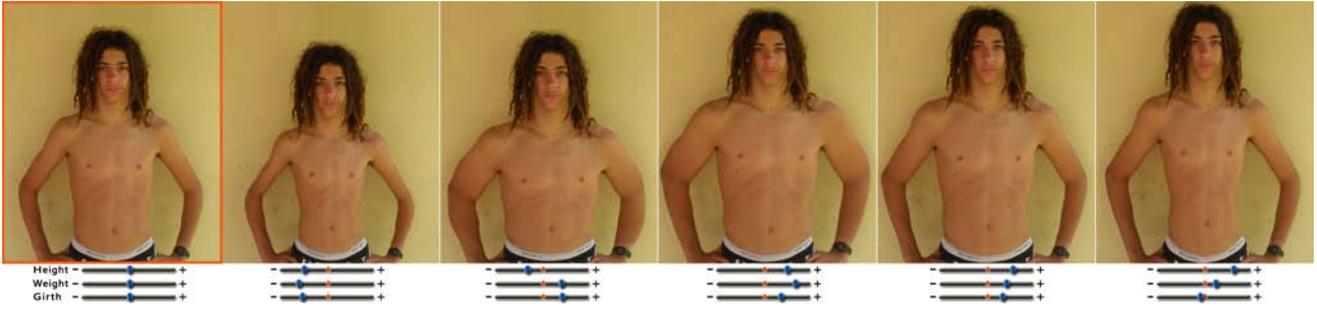
Figure 9: *Various reshaping results of a given image (Leftmost) by manipulating multiple semantic attributes.*

system, we pre-compute the system matrix factorization to enable real-time warping at runtime.

Weights. Due to possible fitting errors, preserving the length changes from the morphable model is formulated as soft constraints only. The weights ω_{ske}, $\omega_{ske\perp}$, ω_{sil}, ω_{reg}, and ω_{dis} are used to balance their corresponding energy terms. Our system uses the following default weights: $\omega_{reg} = 1$, $\omega_{dis} = 1$, $\omega_{ske} = 8$, $\omega_{ske\perp} = 10$, and $\omega_{sil} = (\omega_{ske} + \omega_{ske\perp})/4$. These weights were found to be effective for most cases. In our experiments, we always fix ω_{reg} and ω_{dis}. Users can increase (decrease) weight or length related reshaping effects by intuitively increasing (decreasing) values for $\omega_{ske\perp}$ or ω_{ske}, respectively. Note that we always use a relative small weight for E_{sil} to ensure that smooth but complex deformations can be represented in the deformed subject contour.

6 Results and Discussions

We have evaluated the performance of our reshaping system on a large set of images which contain human subjects with various poses, shapes and (self-)occlusion. Since we aim for subtle reshaping effects, we predefine a maximum range of changes for each semantic attribute as well as its corresponding slider (e.g., [-30kg, +30kg] for weight adjustment). The range settings are also motivated by the training body scans and the derived human shape space: resulting 3D body shapes should be always visually plausible; otherwise it is meaningless to transfer the reshaping effect from the model to the image. After roughly fitting the morphable model to the image, users are allowed to manipulate either individual sliders or multiple sliders simultaneously to form a feature offset $\Delta\gamma$, leading to a reshaped effect.

Figure 13 presents a few representative results for various adjustments of height and weight, roughly organized in order of the complexities of pose and occlusion. Figure 10 shows some reshaping results by manipulating other individual semantic attributes such as waist girth, breast girth and leg length. The parametric power of our system allows users to quickly achieve a variety of reshaping effects by arbitrarily adjusting the attribute parameters, as shown in Figures 1 and 9. The reader is referred to the project pages[4] for more examples.

Our pose and shape fitting needs user assistance, which takes a few minutes for simple cases like examples in Figure 11(top) and Figure 10(right), and up to a dozen of minutes for complex scenarios where most of the interaction is spent on specifying subject contours (e.g., Figure 11(bottom) and Figure 13(d3)). Since our system allows partial contour as input, users can incrementally provide more input in response to the system feedback. For example, users

---

Figure 10: *Parametric reshaping with the manipulation of individual semantic attributes.*

can first focus on easily identifiable contour edges (e.g., boundaries of exposed skin regions) and add more at uncertain regions (e.g., heavily covered by opaque clothes like in Figure 13(a3)) only if the current fitting result is still not satisfactory. The performance bottleneck lies in the shape fitting step, whose computational complexity depends on the length of newly added contour edges. For example, it takes a few seconds to fit the model to the contour shown in Figure 3(b). We argue that the types of interaction involved in the fitting step are intuitive to control, since they are always done with respect to *existing* image features (e.g., silhouettes, joints). In contrast, body reshaping is creative endeavor in which desired visual results cannot always be anticipated, thus having it parametric is particularly useful.

Our warping algorithm deals well with off-body parts which are tightly coupled with the body, like cloth or hair. This avoids the common but more tedious process of first segmenting the human region (including hair, clothes and ornaments etc.), then reshaping the segmented subject, and finally pasting the reshaped subject back to the image by completing in-between holes. However, the image warping might introduce artifacts to the background, which is especially noticeable at regions with structural features (e.g.,

Figure 11: *Our reshaping method is robust for various occlusions and poses. Left: Input images. Middle: Fitting results. Right: Reshaping results (weight increasing for both examples).*

Figure 12: *Reshaping of multiple subjects in a single image: the weights of both subjects are decreased.*

ground and wall in Figure 13(a3)). Letting users manually adjust the saliency map, as done in the literature of content-aware image resizing [Shamir and Sorkine 2009] or warping [Carroll et al. 2009], could alleviate this artifact. Although our warping approach can be applied to either the entire image or the human body region only, for simplicity we warp the entire images for all the examples shown in the paper.

In addition, ideally, a multi-layer representation should be adopted for warping regions with self-occlusion. However, we found that our simple single-layer representation already works well in practice, because the major warping energy terms (i.e., E_{ske}, $E_{ske\perp}$, and E_{sil}) are *occlusion-aware* and have much larger weights than the *occlusion-oblivious* energy terms (i.e., E_{reg} and E_{dis}). Let us elaborate further using the example in Figure 10(left). In this example, the relative changes introduced by the 3D model make the waist shrink horizontally. However, the arms and hands placed on the belly will not be shrunk, because the relative changes for these parts should remain unchanged according to the 3D model, enforced by the occlusion-aware constraints.

Our system is flexible and can reshape either the entire human body or just part of it. It is also applicable to the reshaping of multiple subjects in a single image: first, the fitting is applied one by one, and then the length change constraints are formulated while considering all the fitted models into a single image warping optimization (see an example in Figure 12).

We have compared our parametric reshaping results with those created by professional artists using existing image editing packages like Photoshop$^\copyright$. Please visit our project pages for comparison results. Through the manipulation of two semantic attributes instead of many local operations, we achieve comparable editing results. Since our technique already produces desired global shape changes of the human body, other retouching tools can then be easily used for local adjustment of fine details to our results (e.g., wrinkle removal or facial retouching).

Limitations. Since our morphable model is learned from a limited number of human body scans, it apparently cannot span the entire human shape space. The sparsity of the morphable model may lead to large fitting errors for human images with extreme poses and shapes. For example, our method generally does not work well for infants and children, since the training set only contains adult subjects. Clearly, being limited by the range of the morphable model, we also cannot represent all possible reshaping effects. Finally, since our method achieves reshaping via 2D warping, it cannot completely mimic 3D reshaping effects, and it is especially weak to capture shape changes of the 3D model along the camera projection direction, possibly making the shading cues belie the edits. For example, it can be seen from Figure 2 that the belly looks a bit too drooping with respect to the decreased weight.

7 Conclusion and Future Work

This paper has presented an easy-to-use image retouching system for reshaping a human body in a single image. We rely on a 3D whole-body morphable model to parameterize the reshaping process and globally consistent and visually pleasing editing effects are easily achieved. We introduce a novel body-aware image warping approach which successfully addresses the problems caused by imperfect fitting and occlusions due to various clothes or poses. After roughly fitting the model into the image with user assistance, in appreciation of the parametric nature of the morphable model, our system allows real-time reshaping of an image subject by simply and intuitively manipulating the set of semantic attributes.

As a future work, we are interested in extending our approach for reshaping subjects in videos or multiple views. Although such input data might provide more information for pose and shape fitting, keeping coherent reshaping results across images or frames is challenging. It is also exciting to explore whether the parametric power of our model-based editing framework can be similarly applied to other types of creatures or objects for image manipulation.

References

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

ALLEN, B., CURLESS, B., POPOVIĆ, Z., AND HERTZMANN, A. 2006. Learning a correlated model of identity and pose-dependent body shape variation for real-time synthesis. In *SCA '06*, 147–156.

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

ARAD, N., AND REISFELD, D. 1995. Image warping using few anchor points and radial functions. *Computer Graphics Forum 14*, 1, 35–46.

BALAN, A., AND BLACK, M. 2008. The naked truth: Estimating body shape under clothing. In *ECCV '08*.

BALAN, A. O., SIGAL, L., BLACK, M. J., DAVIS, J. E., AND HAUSSECKER, H. W. 2007. Detailed human shape and pose from images. In *CVPR '07*, 1–8.

BARAN, I., AND POPOVIĆ, J. 2007. Automatic rigging and animation of 3D characters. *ACM Trans. Graph. 26*, 3, 72.

BARRETT, W. A., AND CHENEY, A. S. 2002. Object-based image editing. *ACM Trans. Graph. 21*, 3, 777–784.

BLANZ, V., AND VETTER, T. 1999. A morphable model for the synthesis of 3d faces. In *SIGGRAPH '99*, 187–194.

CARROLL, R., AGRAWALA, M., AND AGARWALA, A. 2009. Optimizing content-preserving projections for wide-angle images. *ACM Trans. Graph. 28*, 3.

DAVIS, J., AGRAWALA, M., CHUANG, E., POPOVIĆ, Z., AND SALESIN, D. 2003. A sketching interface for articulated figure animation. In *SCA '03*, 320–328.

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

EISMANN, K., AND PALMER, W. 2006. *Adobe Photoshop Restoration & Retouching*. New Riders Press; 3 edition.

GUAN, P., WEISS, A., BALAN, A., AND BLACK, M. J. 2009. Estimating human shape and pose from a single image. In *ICCV '09*.

HASLER, N., STOLL, C., SUNKEL, M., ROSENHAHN, B., AND SEIDEL, H.-P. 2009. A statistical model of human pose and body shape. *Computer Graphics Forum 28*, 2, 337–346.

HORNUNG, A., DEKKERS, E., AND KOBBELT, L. 2007. Character animation from 2D pictures and 3D motion data. *ACM Trans. Graph. 26*, 1, 1.

HUA, Z., WANG, G., LIN, X., AND YAN, H. 2009. Recovery of upper body poses in static images based on joints detection. *Pattern Recognition Letters 30*, 5, 503–512.

IGARASHI, T., MOSCOVICH, T., AND HUGHES, J. F. 2005. As-rigid-as-possible shape manipulation. *ACM Trans. Graph. 24*, 3, 1134–1141.

KRAEVOY, V., SHEFFER, A., AND VAN DE PANNE, M. 2009. Modeling from contour drawings. In *SBIM '09*, 37–44.

LEYVAND, T., COHEN-OR, D., DROR, G., AND LISCHINSKI, D. 2008. Data-driven enhancement of facial attractiveness. *ACM Trans. Graph. 27*, 3, 38.

LI, Y., SUN, J., TANG, C.-K., AND SHUM, H.-Y. 2004. Lazy snapping. *ACM Trans. Graph. 23*, 3, 303–308.

MORTENSEN, E. N., AND BARRETT, W. A. 1995. Intelligent scissors for image composition. In *SIGGRAPH '95*, 191–198.

PARAMESWARAN, V., AND CHELLAPPA, R. 2004. View independent human body pose estimation from a single perspective image. In *CVPR '04*, vol. 2.

SCHAEFER, S., MCPHAIL, T., AND WARREN, J. 2006. Image deformation using moving least squares. *ACM Trans. Graph. 25*, 3, 533–540.

SEO, H., AND MAGNENAT-THALMANN, N. 2004. An example-based approach to human body manipulation. *Graphical Models 66*, 1, 1–23.

SHAMIR, A., AND SORKINE, O. 2009. Visual media retargeting. In *SIGGRAPH Asia Course 2009*.

SUMNER, R. W., AND POPOVIĆ, J. 2004. Deformation transfer for triangle meshes. *ACM Trans. Graph. 23*, 3, 399–405.

TAN, G., CHEN, W., AND LIU, L. 2010. Image driven shape deformation using styles. *Journal of Zhejiang University (SCIENCE C) 11*, 1, 27–35.

TAYLOR, C. 2000. Reconstruction of articulated objects from point correspondences in a single uncalibrated image. *Computer Vision and Image Understanding 80*, 3, 349–363.

TZUR, Y., AND TAL, A. 2009. FlexiStickers - photogrammetric texture mapping using casual images. *ACM Trans. Graph. 28*, 3.

VLASIC, D., BRAND, M., PFISTER, H., AND POPOVIĆ, J. 2005. Face transfer with multilinear models. *ACM Trans. Graph. 24*, 3, 426–433.

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

ZHOU, K., HUANG, J., SNYDER, J., LIU, X., BAO, H., GUO, B., AND SHUM, H.-Y. 2005. Large mesh deformation using the volumetric graph Laplacian. *ACM Trans. Graph. 24*, 3, 496–503.
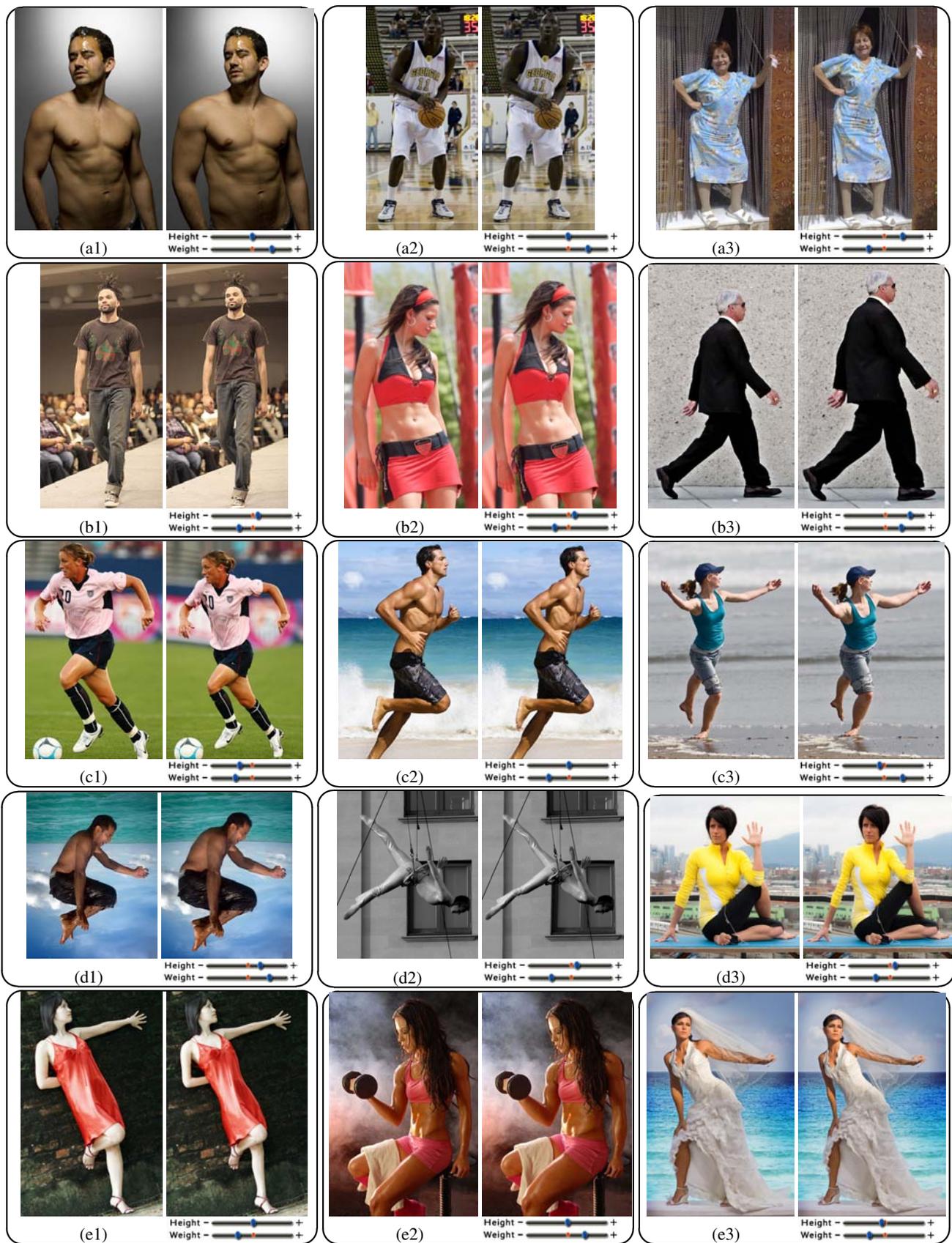
Figure 13: *A series of representative reshaping results for various adjustments of height and weight. We organize the results roughly in order of the complexities of pose (increasing from top to bottom) and occlusion (increasing from left to right).*