# A Practical System for Modelling Body Shapes from Single View Measurements

Yu Chen[1]
yc301@cam.ac.uk
Duncan Robertson[2]
duncan@metail.co.uk
Roberto Cipolla[1]
rc10001@cam.ac.uk

[1] Department of Engineering, University of Cambridge
[2] Metail Inc. London

**Abstract**

This paper describes an interactive system for quickly modelling 3D body shapes from a single image. It provides the user with a convenient way to obtain their 3D body shapes so as to try on virtual garments online. For the ease of use, we first introduce a novel interface for users to conveniently extract anthropometric measurements from a single photo, while using readily available scene cues for automatic image rectification. Then, we propose a unified probabilistic framework using Gaussian processes, which predict the body parameters from input measurements while correcting the aspect ratio ambiguity resulting from photo rectification. Extensive experiments and user studies have supported the efficacy of our system. This system is now being exploited commercially online[1].

## 1 Introduction

Estimating or modelling 3D human body shape is of great importance in both computer vision and graphics, and it has significant commercial applications in entertainment and garment design. Achieving accurate and convincing body shape model quickly and easily for non-expert users is a major challenge for building a good body shape modelling system. In the past decade, many efforts have been done to pursue these tasks. These prior systems can be mainly classified into two types: silhouette-based systems [3, 5, 8, 13] and measurements-based systems [10, 17, 18]. Silhouette-based systems use one or more images as input and estimate the 3D body shape by fitting the silhouette in each view. They have achieved good accuracy in prediction of body shapes and poses but usually suffer from huge computational cost and require considerable manual effort for segmentation. Measurement-based systems, on the other hand, use anthropometric measurements (e.g. height, weight, chest circumference, etc.) or body parameters (e.g. BMI ratio) as input. Compared with silhouettes, anthropometric measurements are relatively invariant to articulation changes (e.g. arm pose changes), better reflect the physiological structure of humans, and give meaningful cues for both global and local body structures. Since a concise input is used, these systems usually run much faster than silhouette-based systems. However, they require anthropometrical



[1] The system is proprietary and is protected through several patent filings. Nothing in this paper constitutes a licence or permission to use any of the techniques described in it.

BMVC 2011 http://dx.doi.org/10.5244/C.25.82

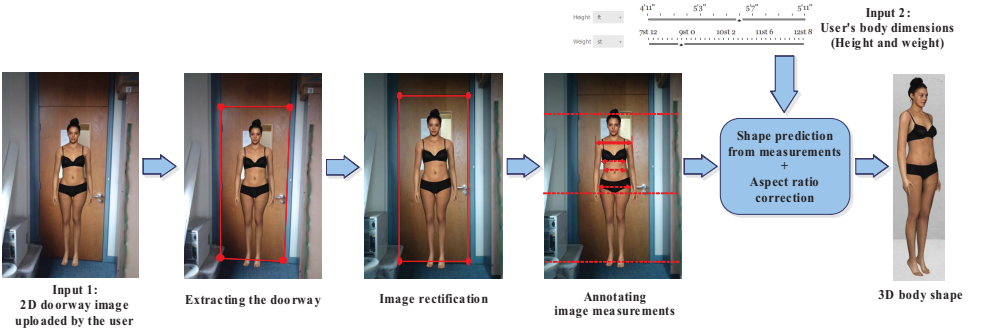


Figure 1: The flowchart of our 3D body shape modelling system.

knowledge to operate, and hence they are more suitable for professional graphics designers rather than non-expert external users.

To exploit the advantages of both types of systems, we present a novel interactive body-shape modelling system using 2D anthropometric measurements extracted from a single "doorway" image, i.e. a photo taken in front of an arbitrary doorway. With a small amount of interaction from users, our system can quickly generate accurate 3D body shape models. Extensive experiments and user studies have been conducted to verify the efficacy of our system. The main contributions of this paper includes: (1) a novel user interface for extracting anthropometric measurements from photos; (2) the automatic image rectification using vanishing points in degenerate cases; (3) a new probabilistic approach for simultaneously predicting the body parameters from measurements and correcting the aspect ratio of the doorway image; and finally (4) a working system for online 3D body shape modelling and garment fitting which is accessible to the public.

**Related Work:** We here give more details on previous studies of 3D human body shapes modelling Many approaches are based on a shape-from-silhouettes framework. Mündermann et al. [13] and Bălan et al. [3] used the SCAPE (Shape Completion and Animation for PEople) model [2], a parametric model for building body shapes, to fit multi-view image silhouettes obtained from 4 to 8 cameras. PCA coefficients are used to model individual body shape variations. Guan et al. [8] extended their work to solve the single-view input problem and improved the reconstruction by including the shading cues. Chen and Cipolla [5] also addressed the problem of recovering 3D body shapes from a single silhouette. They further used Gaussian Process latent variable models, a non-linear manifold approach, on top of the PCA coefficients for a more compact shape presentation and uncertainty measurements.

These shape-from-silhouettes methods have achieved fairly accurate 3D body shapes and good fitting to the input silhouettes. However, there are a couple of drawbacks that limit their applications. First, since silhouette matching involves optimising highly non-linear objective functions and extensive searching in the shape space, these approaches are usually very expensive in computation. To generate a single output, it may take hours, which makes them hard for online applications. Second, these prior-art systems usually require considerable amount of initialisation and interaction, e.g. registering the skeleton to the silhouette or using GrabCut [16] to crop out the silhouettes from images, which can be difficult for users without training. More problematically, the quality of the pose initialisation and segmentation can seriously affect final reconstructed 3D shape.

An alternative approach to body shape modelling is to use anthropometric measurements as input. In the field of graphics, Magnenat-Thalmann et al. [10, 17, 18] have proposed

systems to generate 3D body shapes from tape measurements, e.g. chest, waist, etc., or body parameters such as fat percentage and hip-to-waist ratio. To relate the input measurements with the skeleton joints and the PCA shape coefficients representing the 3D body shape, radial-basis interpolations and linear regressions have been used. The systems have achieved near real-time performance. However, these systems usually require a large number of input variables, and these professional anthropometric tape measurements or body parameters are usually not available for non-expert users.

In view of these problems, our system not only supports regular tape measurements and body dimensions, e.g. height and weight, but also introduces a novel mechanism that allows users to annotate 2D measurements on a single image and thereby avoid the need for a tape measure. This mechanism overcomes the problem that many non-expert users do not know their tape measurements and hence greatly improves the usability of the system.

**An Overview of the System:** As shown in Fig. 1, the operation of our body-shape-from-measurement system can be summarised into following stages. Firstly, the user is requested to provide their basic dimensions, e.g. height and weight, and upload a photo in which they stand against a doorway. Secondly, the doorframe is extracted and used to rectify the image into the ideal frontal view automatically (see Section 2.1). Then, a selected set of 2D anthropometric measurements are annotated on the rectified image interactively (see Section 2.2). Finally, the 3D body shape is predicted from query input measurements (both known body dimensions and image measurement extracted) by a Gaussian Process regressor, which is learned from a 3D human database and captures the relationship between these measurements and the 3D morphing parameters. The aspect ratio ambiguity resulted from the rectification stage is also corrected at this stage (see Section 3).

# 2 Extracting Body Cues by Single View Rectification

Our system provides a novel interface that allows the user to extract a few 2D measurements from their doorway photo in an interactive way. This interface is motivated by the concern that many customers cannot clearly remember their actual tape measurements. A simple on-site image-based measurement extraction interface can thus be a good replacement. The rest of this section will give the details of how the image viewpoint can be rectified using doorway information and how anthropometric measurements are then extracted from the corrected frontal-view image.

## 2.1 Rectifying the Doorway Images

Since input images are taken under uncontrolled conditions, they usually suffer from perspective distortion caused by arbitrary camera orientation and focal length, as shown in Fig. 2(a). These images have to be rectified to a frontal view so that more accurate image measurements can be extracted. The geometry of the rectangular doorframe and vanishing points provides cues for an automatic rectification. The problem of using vanishing point to calibrate intrinsic and extrinsic camera parameters has been addressed in existing literature [4, 6, 7, 9]. In this subsection, we further explore the problem in near-degenerate cases in which one of the vanishing points is close to infinity, and later present a novel algorithm to solve the problem in Section 3.

To formulate the problem, we denote the intrinsic matrix of the camera as $\mathbf{K} = diag(f, f, 1)$, where $f$ refers to the focal length of the camera, and extrinsic camera parameters as $\mathbf{R}$ and $\mathbf{t}$, representing the camera rotation and translation in the world coordinate system. Without loss of generality, we assume that the translation vector $\mathbf{t} = \mathbf{0}$ and the image center $\mathbf{u} = \mathbf{0}$.

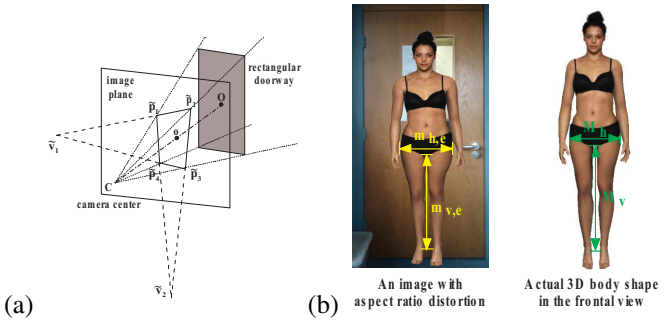


Figure 2: (a) Image rectification by estimating vanishing points from the rectangular doorframe; (b) Measurements for defining aspect ratio distortion. Here, the horizontal and the vertical measurements are hips width and crotch height, respectively.

The task of image rectification is then to estimate the unknown camera rotation matrix $\mathbf{R}$ and focal length $f$.

We denote the homogenous image coordinates of the two vanishing points in horizontal and vertical direction as $\tilde{\mathbf{v}}_1 \sim [v_{1x}, v_{1y}, 1]^T$ and $\tilde{\mathbf{v}}_2 \sim [v_{2x}, v_{2y}, 1]^T$, respectively. The positions of vanishing points are uniquely determined by the rectangular doorframe, as shown in Fig 2(a). We use the approach in [6] to find these vanishing points. According to projective geometry, we then have

$$\tilde{\mathbf{v}}_1 = s_1 \mathbf{K}[\mathbf{R}, \mathbf{t}][1,0,0,0]^T = s_1 \mathbf{K}\mathbf{r}_1, \quad \tilde{\mathbf{v}}_2 = s_2 \mathbf{K}[\mathbf{R}, \mathbf{t}][0,1,0,0]^T = s_2 \mathbf{K}\mathbf{r}_2, \tag{1}$$

where $\mathbf{r_j}$ $(j = 1,2,3)$ refers to the $j$-th column of matrix $\mathbf{R}$; $s_1$ and $s_2$ are scaling factors. From (1), the camera rotation matrix $\mathbf{R}$ can be solved as follows:

$$\mathbf{R} = [\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3] = \left[ \frac{\mathbf{K}^{-1}\tilde{\mathbf{v}}_1}{\|\mathbf{K}^{-1}\tilde{\mathbf{v}}_1\|}, \frac{\mathbf{K}^{-1}\tilde{\mathbf{v}}_2}{\|\mathbf{K}^{-1}\tilde{\mathbf{v}}_2\|}, \tau\left( \frac{\mathbf{K}^{-1}\tilde{\mathbf{v}}_1 \times \mathbf{K}^{-1}\tilde{\mathbf{v}}_2}{\|\mathbf{K}^{-1}\tilde{\mathbf{v}}_1 \times \mathbf{K}^{-1}\tilde{\mathbf{v}}_2\|} \right) \right]. \tag{2}$$

where $\tau \in \{1, -1\}$ is a sign coefficient which guarantees that $\det \mathbf{R} = 1$.

Finally, the frontal-view image is obtained by de-rotating the camera with $\mathbf{R}^{-1}$. For an arbitrary point $\tilde{\mathbf{p}}$ in the unrectified image, its corresponding homogenous coordinate $\tilde{\mathbf{p}}^{\mathbf{0}}$ in the frontal view image can thus be computed as $\tilde{\mathbf{p}}^{\mathbf{0}} = \mathbf{H}\tilde{\mathbf{p}} = \mathbf{K}\mathbf{R}^{-1}\mathbf{K}^{-1}\tilde{\mathbf{p}}$. Given an input image, the homography matrix $\mathbf{H}$ is a function of the uncalibrated camera focal length $f$. To determine $\mathbf{H}$, earlier work [4, 6] exploited the orthogonal constraint of the rotational matrix $\mathbf{r}_1^T\mathbf{r}_2 = 0$ and gave a closed-form estimate of the focal length $f = \sqrt{-(v_{1x}v_{2x} + v_{1y}v_{2y})}$. This works in the regular cases when $\tilde{\mathbf{v}}_1$ and $\tilde{\mathbf{v}}_2$ are at finite positions.

Unfortunately, we find that in practice most input images are near-degenerate, in which one of the vanishing points $\tilde{\mathbf{v}}_1$ or $\tilde{\mathbf{v}}_2$ is close to infinity. In those cases, $\mathbf{H}$ is ambiguous as the camera focal length $f$ cannot be accurately estimated, and this results in an aspect ratio distortion in the rectification result. This problem is unexplored in the literature. We here provide a novel quantitative uncertainty analysis of the distortion ratio $\alpha$, which is defined as: $\alpha = \frac{m_{v,e}M_h}{M_v m_{h,e}}$, where $M_v$ and $M_h$ are referred to two arbitrary vertical and horizontal measurements in the 3D space, and $m_{v,e}$ and $m_{h,e}$ are referred to the corresponding measurements taken on the rectified image with distortion (see Fig 2(b) for example). Without loss of generality, we assume that vanishing directions have been aligned with the axes of world coordinates. We first consider the case that the vertical vanishing point $\mathbf{v}_2$ is at infinity, i.e. $\tilde{\mathbf{v}}_2 \sim [0,1,0]^T$, while the horizontal vanishing point has a finite coordinate $\tilde{\mathbf{v}}_1 \sim [v_{1x}, v_{1y}, 1]^T$. Using (2) and the orthonormal constraint of the rotational matrix $\mathbf{r}_1^T\mathbf{r}_2 = 0$, we can infer that $v_{1y} = 0$, and hence the rotation matrix $\mathbf{R}$ and the homography matrix $\mathbf{H}$ for rectification

become

$$\mathbf{R} = \begin{bmatrix} \frac{v_{1x}}{\sqrt{v_{1x}^2+f^2}} & 0 & -\frac{f}{\sqrt{v_{1x}^2+f^2}} \\ 0 & 1 & 0 \\ \frac{f}{\sqrt{v_{1x}^2+f^2}} & 0 & \frac{v_{1x}}{\sqrt{v_{1x}^2+f^2}} \end{bmatrix}, \mathbf{H} = \mathbf{K}\mathbf{R}^{-1}\mathbf{K}^{-1} = \begin{bmatrix} \frac{v_{1x}}{\sqrt{v_{1x}^2+f^2}} & 0 & \frac{f^2}{\sqrt{v_{1x}^2+f^2}} \\ 0 & 1 & 0 \\ -\frac{1}{\sqrt{v_{1x}^2+f^2}} & 0 & \frac{v_{1x}}{\sqrt{v_{1x}^2+f^2}} \end{bmatrix}, \quad (3)$$

respectively. Both matrices are functions of the focal length $f$. Now we assume that we rectify the image based on an inaccurate estimate of the focal length $f_e$, while the actual focal length of the camera is $f_t$. Given an arbitrary 3D point $\mathbf{P} = [P_x, P_y, P_z]^T$ in the world coordinate system, its corresponding homogenous coordinate $\tilde{\mathbf{p}}_\mathbf{e} \sim [\tilde{p}_{e,1}, \tilde{p}_{e,2}, \tilde{p}_{e,3}]^T$ in the inaccurately rectified image will be:

$$\tilde{\mathbf{p}}_\mathbf{e} = \mathbf{H}_\mathbf{e}\mathbf{K}_\mathbf{t}\mathbf{R}_\mathbf{t}\mathbf{P} = \left[ fP_x\sqrt{\frac{v_{1x}^2+f_e^2}{v_{1x}^2+f_t^2}} + \frac{v_{1x}P_z(f_e^2-f_t^2)}{\sqrt{(v_{1x}^2+f_e^2)(v_{1x}^2+f_t^2)}}, fP_y, P_z\sqrt{\frac{v_{1x}^2+f_t^2}{v_{1x}^2+f_e^2}} \right]^T, \quad (4)$$

and the actual image coordinate $\mathbf{p}_\mathbf{e}$ can be written as

$$\mathbf{p}_\mathbf{e} = [p_{e,x}, p_{e,y}]^T = [\frac{\tilde{p}_{e,1}}{\tilde{p}_{e,3}}, \frac{\tilde{p}_{e,2}}{\tilde{p}_{e,3}}]^T = \left[ f\frac{P_x(v_{1x}^2+f_e^2)}{P_z(v_{1x}^2+f_t^2)} + \frac{v_{1x}(f_e^2-f_t^2)}{v_{1x}^2+f_e^2}, f\frac{P_y}{P_z}\sqrt{\frac{v_{1x}^2+f_e^2}{v_{1x}^2+f_t^2}} \right]^T. \quad (5)$$

Now we consider two measurements in the 3D world coordinate system: an arbitrary horizontal measurement $M_h$ defined by two 3D points $\mathbf{P_l} = [P_x^l, P_y, P_z]^T$ and $\mathbf{P_r} = [P_x^r, P_y, P_z]^T$, and an arbitrary vertical measurement $M_v$ defined by two 3D points $\mathbf{P_u} = [P_x, P_y^u, P_z]^T$ and $\mathbf{P_d} = [P_x, P_y^d, P_z]^T$ (see Fig 2(b)). It is obvious that the lengths of these two measurements are $M_h = |P_x^r - P_x^l|$ and $M_v = |P_y^u - P_y^d|$, respectively. Using (5), we can transform the end points of these two measurements into the rectified image and obtain the corresponding image measurements $m_{h,e}$ and $m_{v,e}$ as:

$$m_{h,e} = |p_{e,x}^r - p_{e,x}^l| = f\frac{|P_x^r - P_x^l|(v_{1x}^2+f_e^2)}{P_z(v_{1x}^2+f_t^2)}, \; m_{v,e} = |p_{e,y}^u - p_{e,y}^d| = f\frac{|P_y^u - P_y^d|}{P_z}\sqrt{\frac{v_{1x}^2+f_e^2}{v_{1x}^2+f_t^2}}, \quad (6)$$

By combining the equations above, we can finally obtain the distortion ratio $\alpha$ caused by the inaccurate focal length estimate as follows.

$$\alpha = \frac{m_{v,e}M_h}{M_v m_{h,e}} = \sqrt{\frac{v_{1x}^2+f_t^2}{v_{1x}^2+f_e^2}}. \quad (7)$$

Similarly, in the case that the horizontal vanishing point $\mathbf{v_1}$ is at infinity, i.e. $\tilde{\mathbf{v}}_\mathbf{1} \sim [1,0,0]^T$, we have $\alpha = \sqrt{\frac{v_{2y}^2+f_e^2}{v_{2y}^2+f_t^2}}$. The method for correcting this aspect ratio ambiguity in the body shape estimation will be further addressed in Section 3.

## 2.2 Selection of Image Measurements

After the image rectification stage in Section 2.1, users are asked to annotate a selected set of measurements on the frontal-view image. The selection of image measurements are based on following criteria. First, these 2D image measurements are well-defined by the anthropometric positions, e.g. top of head, crotch, heels, etc., which are easy to discern and unambiguous to users, so that they can be easily and accurately annotated on the photo. Second, they should have good correlations with the corresponding tape measurements and convey enough information for estimating the 3D body shape. Third, user's effort for annotation should be minimised. We finally adopt the following 5 image measurements according to the criteria above, including two vertical measurements: **image body height** (from the top of head to the heels) and **image crotch height** (from the crotch to the heels); and three horizon-

tal measurements: **under-bust width**, **waist width**, and **hip width**. These three horizontal measurements are taken at fixed vertical positions between the top of head and the crotch. These vertical positions have been optimised on a training set of annotated image (see Section 4) so that the resulting image measurements have high correlation factors ($\rho > 0.90$) with their corresponding tape measurements. Our system provides tools that allow users to mark up these 5 photo measurements quickly.

# 3 Probabilistic Estimation of Body Shapes

In our system, a body shape estimator is learned to predict the 3D body shape from user's input, including both image measurements (see Section 2.2) and actual measurements.

**Generating 3D Training Data:** The Civilian American and European Surface Anthropometry Resource (CAESAR) dataset [15], which contains over 2000 North American and European 3D human body instances, is used as the 3D training data. To concisely represent 3D body shapes, we register each 3D instance in the dataset with a 3D morphable human body model, which is similar to [1]. An arbitrary 3D body shape $\mathbf{V}$ is decomposed into a linear combination of body morphs: $\mathbf{V} = \mathbf{V_0} + \sum_{j=1}^{P} y_j \Delta \mathbf{V_j}$, where $\mathbf{V_0}$ refers to the zero body shape and $\Delta \mathbf{V_j}$ $(j = 1, 2, \cdots, P)$ are $P$ different modes of body morphs. In this way, any body shape can be defined by a $P$-dimensional vector of morphing parameters $\mathbf{y} = (y_1, y_2, \cdots, y_P)$.

**Learning a Body Shape Estimator:** Learning a shape-from-measurements estimator can be formulated into a regression problem. We denote $\mathbf{y}$ as the vector of morphing parameters, and $\mathbf{z} = \{\mathbf{z_V}, \mathbf{z_I}\}$ as the vector of input measurements, which is concatenated by actual measurements $\mathbf{z_V}$ and image measurements $\mathbf{z_I}$. In the following contexts, $\mathbf{z_V}$ and $\mathbf{z_I}$ are normalised by the actual and image body height, respectively.

Given $N$ pairs of known body morphing parameters $\mathbf{Y} = \{\mathbf{y_1}, \mathbf{y_2}, \cdots, \mathbf{y_N}\}$ and corresponding measurements provided by users: $\mathbf{Z} = \{\mathbf{z_1}, \mathbf{z_2}, \cdots, \mathbf{z_N}\}$, we learn a Gaussian process (GP) regressor $\mathscr{G}$ that gives a mapping from the measurement input $\mathbf{z}$ to those morphing parameters $\mathbf{y}$ that represent a 3D body shape. With the assumption of dimension independence, the likelihood of observations can be formulated as the following product of $m$ independent Gaussian processes [11]: $P(\mathbf{Y}|\mathbf{Z}, \theta_\mathbf{Y}) = \prod_{i=1}^{m} \mathscr{N}(\mathbf{Y_{:,i}}; 0, \mathbf{K_Y})$, where $\mathscr{N}(*;*,*)$ denotes a Gaussian distribution; $\mathbf{Y_{:,i}}$ denotes the $N \times 1$ column vectors constructed from the $i$-th dimension of $\mathbf{Y}$; $\mathbf{K_Y} = [K_Y^{(i,j)}]_{1 \le i \le N, 1 \le j \le N}$ is the kernel matrix and it is defined as "RBF+linear" kernels [14] in this paper. To train the model $\mathscr{G}$, we minimise the negative log likelihood $L$ with respect to the kernel hyper-parameters $\theta_\mathbf{Y}$ as

$$L = -\log P(\mathbf{Y}|\mathbf{Z}, \theta_\mathbf{Y}) = \frac{1}{2} tr(\mathbf{K_Y^{-1}} \mathbf{Y} \mathbf{Y}^T) + \frac{m}{2} \log |\mathbf{K_Y}| + const. \tag{8}$$

This optimisation can be done using the scaled conjugate gradient (SCG) approach [12].

**Morphing Parameters Prediction with Aspect Ratio Correction (ARC):** In the testing stage, the morphing parameters $\hat{y}$ can be predicted from measurement inputs using the GP regressor $\mathscr{G}$ obtained in the previous section. However, we have mentioned that the rectification algorithm proposed in Section 2.1 may result in an ambiguity in the image aspect ratio (height/width ratio), which will affect all those horizontal image measurements. In this subsection, we propose a probabilistic approach to infer the morphing parameters and correct this aspect ratio simultaneously. Assume that a testing user provides his/her actual body measurements $\hat{\mathbf{z_V}}$ as well as a set of image measurements $\hat{\mathbf{z}}_\mathbf{I} = \{\hat{\mathbf{z}}_{\mathbf{I,v}}, \hat{\mathbf{z}}_{\mathbf{I,h}}\}$ annotated on the rectified doorway image, where $\hat{\mathbf{z}}_{\mathbf{I,v}}$ and $\hat{\mathbf{z}}_{\mathbf{I,h}}$ represent vertical and horizontal image measurements, respectively. The complete set of testing measurements can thus be written as $\hat{\mathbf{z}}(\alpha) = [\hat{\mathbf{z}}_\mathbf{V}, \hat{\mathbf{z}}_{\mathbf{I,v}}, \alpha \hat{\mathbf{z}}_{\mathbf{I,h}}]$, where $\alpha$ represents the ambiguous image aspect ratio resulted from

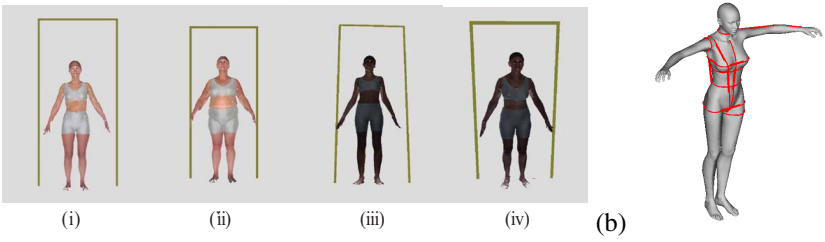


Figure 3: (a) Examples of synthetic doorway images from CAESAR laser scans; (b) Exemplar tape measurements on the 3D mesh for the purpose of evaluation.

the photo rectification procedure (see Section 2.1). Then the joint posterior of the Gaussian process estimator given the uncertainty of $\alpha$ can be formulated as:

$$J = P(\hat{\mathbf{y}}, \alpha | \hat{\mathbf{z}}_\mathbf{V}, \hat{\mathbf{z}}_\mathbf{I}, \mathscr{G}) = P(\hat{\mathbf{y}} | \alpha, \hat{\mathbf{z}}_\mathbf{V}, \hat{\mathbf{z}}_\mathbf{I}, \mathscr{G}) P(\alpha) = \mathscr{N}(\hat{\mathbf{y}} | \mu_\mathbf{y}, \sigma_y^2 \mathbf{I}) \mathscr{N}(\alpha | \mu_\alpha, \sigma_\alpha^2). \quad (9)$$

where $\mu_\mathbf{y} = \mathbf{k}_\mathbf{Y}^\mathbf{T}(\hat{\mathbf{z}}) \mathbf{K}_\mathbf{Y}^{-\mathbf{1}} \mathbf{Y}$ and $\sigma_y^2 = k_Y(\hat{\mathbf{z}}, \hat{\mathbf{z}}) - \mathbf{k}_\mathbf{Y}^\mathbf{T}(\hat{\mathbf{z}}) \mathbf{K}_\mathbf{Y}^{-\mathbf{1}} \mathbf{k}_\mathbf{Y}(\hat{\mathbf{z}})$. Here, the prior $P(\alpha)$ of the aspect ratio $\alpha$ is modeled by a Gaussian distribution, and its parameters $\mu_\alpha$ and $\sigma_\alpha^2$ are estimated from a number of real images. In our system, we set the mean to be $\mu_\alpha = 0.9935$ and the standard deviation to be $\sigma_\alpha = 0.0506$, respectively, which are obtained from our statistics. Shape prediction and aspect ratio correction can thus be done by minimising the negative log joint posterior $-\log J$ with respect to $\mathbf{y}$ and $\alpha$:

$$\begin{aligned}(\hat{\mathbf{y}}_\mathbf{MAP}, \alpha_{MAP}) =& \arg\min_{(\hat{\mathbf{y}}, \alpha)} -\log J \\ =& \arg\min_{(\hat{\mathbf{y}}, \alpha)} \left( \frac{d_y}{2} \log(\sigma_y^2) + \frac{\|\hat{\mathbf{y}} - \mu_\mathbf{y}\|^2}{2\sigma_y^2} + \frac{1}{2} \log(\sigma_\alpha^2) + \frac{(\alpha - \mu_\alpha)^2}{2\sigma_\alpha^2} \right). \quad (10)\end{aligned}$$

By taking derivatives $\frac{\partial F}{\partial \hat{\mathbf{y}}} = \mathbf{0}$ and $\frac{\partial F}{\partial \alpha} = 0$, we have $\hat{\mathbf{y}}_\mathbf{MAP} = \mu_\mathbf{y}$ and $\frac{d_y}{2\sigma_y^2} \frac{\partial \sigma_y^2}{\partial \alpha} + \frac{\alpha - \mu_\alpha}{\sigma_\alpha^2} = 0$, in which the optimal aspect ratio $\alpha_{opt}$ can be solved efficiently using the fixed point equation $\alpha^{(t+1)} = \mu_\alpha + \frac{d_y \sigma_\alpha^2}{2\sigma_y^2(\alpha^{(t)})} \frac{\partial \sigma_y^2}{\partial \alpha} \Big|_{\alpha = \alpha^{(t)}}$ within a few iterations. Multiple random initialisations are used to avoid local minima.

# 4 Experimental Results

In this section, we evaluate the performance our body-shape-from-measurements system on both synthetic data and real data from users. For the synthetic data, we render a doorway image dataset with 3D laser scans of 1027 European female instances in the CAESAR dataset. For each instance, three parallel doorway images under different settings are generated. This give rises to the following 3 image sets. **Set 1: Standard** (perfect frontal view, fixed camera distance and focal length, no body rotation; see Fig. 3(a,i) for example); **Set 2: Frontal View** (perfect frontal view, varied camera distance and focal length, slight body rotation; see Fig. 3(a,ii) for example); **Set 3: Unrectified** (random view, varied camera distance and focal length, slight body rotation; see Fig. 3(a,iii) and (a,iv) for examples). Set 1 images are annotated and used for training our system, while Set 2 and 3 images are for testing purpose.

**Tests on Frontal View Images:** We first examine the accuracy of the regression approach on perfectly frontal-view images (Set 2 images). In this experiment, a smaller subset of measurements are used as input to recover the 3D body shapes, and the quantitative evaluation is done by taking 15 standard measurements on the resulting 3D mesh (see Fig. 3(b) for examples) and comparing them with the ground truths. We perform cross validations on the dataset to reduce the danger of over-fitting. In each experiment, 75% of instances are used

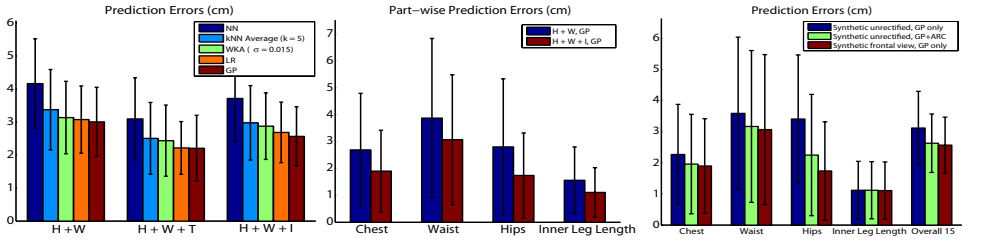


Figure 4: (left) Overall measurement prediction errors achieved by different regressors; (center) Prediction errors on perfectly rectified images at specific body parts; (right) Overall and part-wise prediction errors on synthetic unrectified images.

as the training set and the rest are used for testing. We investigated the following 3 different combinations of input measurements:

- H+W: height and weight only;
- H+W+T: height, weight, and 3 tape measurements: under-chest circumference, waist circumference, and hip circumference;
- H+W+I: height, weight, and the 5 selected image measurements (see Section 2.2) which are annotated on frontal view doorway images (in Set 2).

We compared the GP regressor we used against other 4 different regressors: (1) **nearest-neighbor regressor (NN)**; (2) **average of k-nearest neighbor regressor (kNN-average)**; (3) **linear regressor (LR)**; and (4) **weighted-kernel average regressor (WKA)**: a non-parametric regressor based on the normalised Parzen window. It estimates the body morph parameter $\hat{\mathbf{y}}$ based on all the nearby training data as $\hat{\mathbf{y}} = \frac{\sum_{i=1}^{N} \Phi(\hat{\mathbf{z}},\mathbf{z_i})\mathbf{y_i}}{\sum_{i=1}^{N} \Phi(\hat{\mathbf{z}},\mathbf{z_i})}$,where the radial basis function (RBF) kernel function $\Phi(\hat{\mathbf{z}},\mathbf{z_i}) = \frac{\exp\left(-\|\hat{\mathbf{z}}-\mathbf{z_i}\|^2/2\sigma^2\right)}{\sum_{j=1}^{N}\exp\left(-\|\mathbf{z_j}-\mathbf{z_i}\|^2/2\sigma^2\right)}$ $(i = 1,2,\cdots,N)$ is used and $\sigma$ is the radius of the smoothing kernel. In the experiments, the kernel widths are adjusted to $\sigma = 0.015$ in WKA such that the regressor can provide a near-optimal performance.

Fig. 4(left) shows the performances of different regressors on all 4 different combinations of measurement inputs. Here, average absolute error (AAE) on the whole dataset is used as the criterion for all comparisons. Concerning the performance of each individual regressor, the results show that GP constantly performs equally or better than the other 4 regressors in all 3 settings. Concerning the effectiveness of input measurement combinations, we observe that predictive accuracy of each regressor are considerably improved in the settings H+W+T and H+W+I compared with the baseline setting H+W. As expected, the results for the H+W+I setting lies between those for the H+W setting and the H+W+T setting whichever regressor is used. It shows that tape measurements are more effective to define a body shape than the corresponding image measurements. The major cause could be that image measurements suffer from annotation errors and the loss of depth information. However, image measurements do provide enough extra information to constrain the body shape and they can be used as a compromise when tape measurements are unavailable.

To further evaluate the performance of the H+W+I setting, we compute the prediction errors of 4 specific measurements: chest, waist, hips, and inner leg length, which define the dimensions of specific body parts and are crucial for garment design. We compare the results with those from the H+W input setting, as shown in Fig. 4(center). By introducing image measurements, the average prediction errors on actual tape measurements are reduced by $15-45\%$ from the results based on height and weight only. Among all 4 measurements,

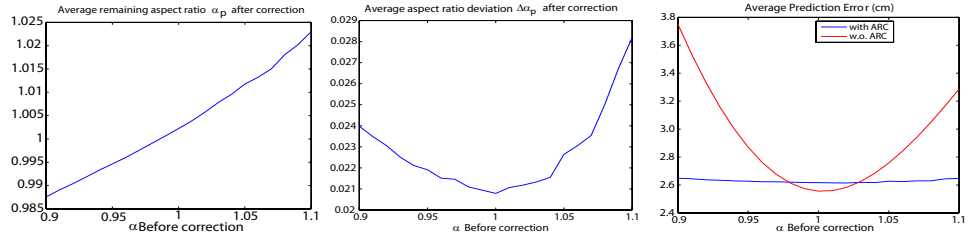


Figure 5: Simulating aspect ratio correction on distorted Set 2 images. (left/center): Average remaining aspect ratio $\alpha_p$ and deviation $\Delta\alpha_p$ after correction are provided; (right): Average absolute prediction errors under different aspect ratios $\alpha$.

the prediction on waist circumference is less accurate compared with the others, due to its ambiguous definition in the CAESAR dataset.

**Tests on Unrectified Images and Aspect Ratio Correction:** To evaluate the efficacy of the aspect ratio correction (ARC) scheme we proposed in Section 3, we synthetically stretch each perfectly frontal view testing image (in Set 2) into different aspect ratios $\alpha$, ranging from 0.9 to 1.1, and try correcting them with our algorithm. Two error measurements are used: (1) remaining aspect ratio $\alpha_p$ after correction; (2) aspect ratio deviation $\Delta\alpha_p$ after correction, which is defined as $\Delta\alpha_p = |\alpha_p - 1|$. We compute the average $\alpha_p$ and $\Delta\alpha_p$ over all the images in Set 2, and summarise the results in Fig. 5(left/center). It can be observed that our algorithm neutralises about 60% of the aspect ratio distortion in the input on average. We also perform body measurements prediction on the images with different distortion levels, and plot the prediction errors against $\alpha$ in Fig. 5(right), which shows that the ARC scheme can considerably improve the accuracy when a large aspect ratio distortion is present.

We then test the performance of the system on unrectified doorway images (in Set 3) by going through the complete procedures of image rectification and morphing parameter prediction, and checking the prediction accuracy of the same 15 mesh measurements. Contrast experiments have been conducted by performing shape parameter prediction from image measurements either with or without ARC, as shown in Fig. 4(right). For the purpose of comparison, we also provide results on corresponding ideal frontal view images (Set 2) as the baselines. These results indicate that using the ARC scheme can considerably increase the system robustness against the aspect ratio distortion during the rectification stage.

**Tests on Photos from Real Users:** We finally evaluate the usability and accuracy of the proposed system on photos of real users. Volunteers were invited to test our virtual fitting-room software online and then fill in questionnaires afterwards. They were asked to take photos at home in an uncontrolled environment, and then use our system to extract image measurements and create their own body shape models. Fig. 6 gives some qualitative results for different body shapes, showing that our system can provide meaningful body shape predictions based on the given input information, although detailed shapes around waist areas may sometimes be less accurate (e.g. instance 2 in Fig. 6). For a quantitative evaluation, we also collect 4 frequently used tape measurements: chest, waist, hips, and inner leg length, from these volunteers, and compare them with the corresponding measurements taken from the predicted body shapes. The statistics given in Table 1 show the usability of our system.

Concerning efficiency, feedback shows that most users take about 1.5 - 3 minutes to finish marking the doorway and 5 image measurements (see Section 2.2) on the photo using our current interface. This includes the time for the automatic image rectification. Then, it takes less than one second for a PC with a 2.4GHz CPU to correct the aspect ratio, estimate

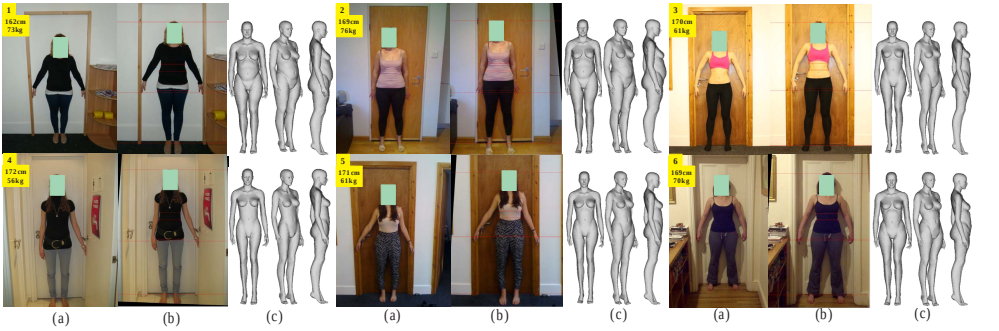


Figure 6: Qualitative results generated by real users: (a) uploaded photos; (b) rectified images; (c) 3D body shapes predicted by our system (in three different views).

| Body part | Chest | Waist | Hips | Inner leg length |
|---|---|---|---|---|
| Error(cm) | $1.52 \pm 1.36$ | $1.88 \pm 1.06$ | $3.10 \pm 1.86$ | $0.79 \pm 0.90$ |

Table 1: Prediction errors on real user images in Fig. 6.

the morphing parameters, and generate the 3D body shape from input measurements.

# 5 Conclusion

In this paper, we present a novel system for reconstructing reliable 3D human body shapes from anthropometric measurements, and introduce an interface for user to conveniently extract these measurements on doorway photos. Qualitative and quantitative experiments have supported the efficacy and usability of our shape-from-measurements system. Our ongoing and future work includes: (1) building up a public database for shape-from-measurement research, which includes anthropometric measurements, 3D body scans, as well as doorways images from volunteers; (2) carrying out a more thorough user study; and (3) extending the algorithms and converting them into a commercial online fitting-room system.

# References

[1] B. Allen, B. Curless, and Z. Popović. The space of human body shapes: reconstruction and parametrization from range scans. *ACM Transactions on Graphics*, 22(3):587–594, 2003.

[2] D. Anguelov, P. Srinivasan, D. Koller, S. Thrun, and J. Rodgers. SCAPE: Shape completion and animation of people. *SIGGRAPH*, 24:408–416, 2005.

[3] A.O. Balan, L. Sigal, M.J. Black, J.E. Davis, and H.W. Haussecker. Detailed human shape and pose from images. *IEEE Conference on Computer Vision and Pattern Recognition*, 2007.

[4] B. Caprile and V. Torre. Using vanishing points for camera calibration. *International Journal of Computer Vision*, 4:127–140, 1990.

[5] Y. Chen and R. Cipolla. Single and sparse view 3d reconstruction by learning shape priors. *Computer Vision and Image Understanding*, 115(5):586–602, 2011.

[6] R. Cipolla, T. Drummond, and D. Robertson. Camera calibration from vanishing points in images of architectural scenes. *British Machine Vision Conference*, 2:382–391, 1999.

[7] A. Criminisi, I. Reid, and A. Zisserman. Single view metrology. *International Journal of Computer Vision*, 40(2):123–148, 2000.

[8] P. Guan, A. Weiss, A.O. Balan, and M.J. Black. Estimating human shape and pose from a single image. *IEEE International Conference on Computer Vision*, 2009.

[9] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2000.

[10] M. Kasap and N. Magnenat-Thalmann. Parameterized human body model for real-time applications. *Proceedings of 2007 International Conference on Cyberworlds*, pages 160–167, 2007.

[11] D.J.C. Mackay. Introduction to gaussian processes. *Neural Networks and Machine Learning*, pages 133–166, 1998.

[12] M.F. Moller. A scaled conjugate gradient algorithm for fast supervised learning. *Neural Networks*, 6:525–533, 1993.

[13] L. Mundermann, S. Corazza, and T. Andriacchi. Accurately measuring human movement using articulated ICP with soft-joint constraints and a repository of articulated model. *IEEE Conference on Computer Vision and Pattern Recognition*, 2007.

[14] R. Navaratnam, A. Fitzgibbon, and R. Cipolla. Semi-supervised joint manifold learning for multi-valued regression. *IEEE International Conference on Computer Vision*, 2007.

[15] K.M. Robinette, H.A.M. Daanen, and E. Paquet. The CAESAR project: a 3-D surface anthropometry survey. *International Conference on 3-D Digital Imaging and Modeling*, pages 380–386, 1999.

[16] C. Rother, V. Kolmogorov, and A. Blake. "grabcut" – interactive foreground extraction using iterated graph cuts. *Proc. of SIGGRAPH*, pages 309–314, 2004.

[17] H. Seo and N. Magnenat-Thalmann. An automatic modeling of human bodies from sizing parameters. *ACM SIGGRAPH Symposium on Interactive 3D Graphics Proc.*, pages 19–26, 2003.

[18] H. Seo, F. Cordier, and N. Magnenat-Thalmann. Synthesizing animatable body models with parameterized shape modifications. *Eurographics/SIGGRAPH Symposium on Computer Animation*, pages 120–125, 2003.