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]
cipolla@eng.cam.ac.uk

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

Background and contributions. 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. The prior systems can be mainly classified into two types: silhouette-based systems, e.g. [1, 3], and measurements-based systems, e.g. [5]. 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 (see Fig. 2(a) for example). With a small amount of interaction from users, our system can quickly generate accurate 3D body shape models. 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.

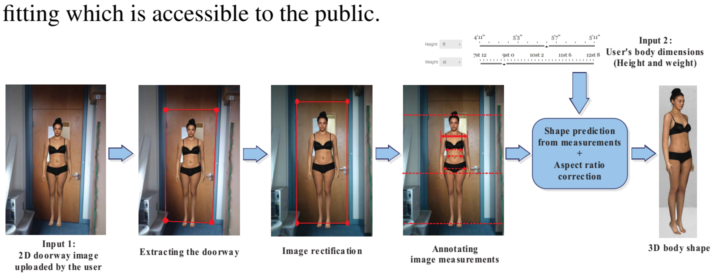

Figure 1: The flowchart of our 3D body shape modelling system.

Approach overview– As shown in Fig. 1, the operation of our 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 annotated and used to rectify the image into the ideal frontal view automatically. Then, a selected set of 2D anthropometric measurements are annotated on the rectified image interactively. Finally, the 3D body shape is predicted from the query input by a Gaussian Process (GP) [4] regressor. The aspect ratio ambiguity in the rectification stage can be also corrected simultaneously.

(1) Rectifying the doorway photo. Since input images are taken under uncontrolled conditions, they usually suffer from perspective distortion caused by arbitrary camera orientation and focal length (see Fig. 2(b)). These images have to be rectified to a frontal view so that more accurate image measurements can be extracted. From the projective geometry, we would like to compute the homography $\mathbf{H} = \mathbf{K}\mathbf{R}^{-1}\mathbf{K}^{-1}$, where $\mathbf{K}$ is the intrinsic camera matrix determined by the focal length f, and $\mathbf{R}$ is the camera rotation matrix. The rectangular doorframe uniquely determines the positions of vanishing points ($\tilde{\mathbf{v}}_1$ and $\tilde{\mathbf{v}}_2$ in Fig 2(b)), which can be used to compute the camera rotation $\mathbf{R}$. To determine f, earlier work, e.g. [2], exploited the orthogonal constraint of the rotational matrix 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 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 (height/width ratio) distortion in the rectification result. This problem is unexplored in the literature. In the paper, we show that the aspect ratio distortion factor α can be estimated by $\alpha = \sqrt{\frac{v_{1x}^2+f_t^2}{v_{1x}^2+f_e^2}}$ (when $\tilde{\mathbf{v}}_2$ is close to infinity) and $\alpha = \sqrt{\frac{v_{2y}^2+f_e^2}{v_{2y}^2+f_t^2}}$ (when $\tilde{\mathbf{v}}_1$ is close to infinity), where f_e and f_t are the estimated and the ground truth focal length, respectively.

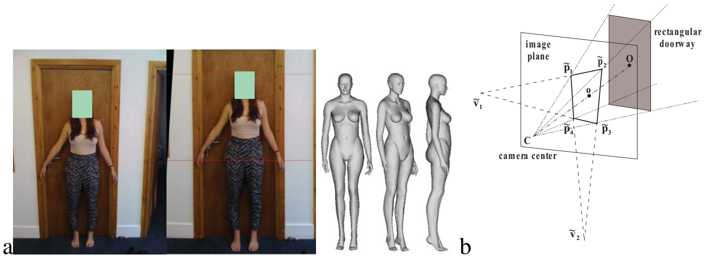

Figure 2: (a) Body shape from image measurements. left: an uploaded photo; center: the rectified image with annotated image measurements; right: the predicted 3D body shape (in 3 different views). (b) Image rectification by estimating vanishing points from the rectangular doorframe.

(2) Annotating image measurements. After the image rectification stage, users are asked to annotate 5 well-defined image measurements on the frontal-view image (see Fig. 2), including 2 vertical measurements: image body height and image crotch height; and 3 horizontal measurements: under-bust width, waist width, and hip width. We provide a novel interface such that these measurements can be easily annotated by users.

(3) Probabilistic estimation of body shapes. Learning a shape-from-measurements estimator can be formulated into a regression problem. We use the CAESAR dataset as the 3D training data. To concisely represent 3D body shapes, we register each instance in the dataset with a 3D morphable body model, and an arbitrary 3D body shape can be represented by a linear combination of body morphs. Given N pairs of known body morphing parameters and corresponding measurements, we learn a GP regressor $\mathscr{G}$ that gives a mapping from the normalised 2D measurement input $\mathbf{z}$ to those morphing parameters $\mathbf{y}$ that represent the body shape.

In the testing stage, the body morphing parameters $\hat{y}$ can be predicted from new measurement inputs using the obtained GP regressor $\mathscr{G}$. Considering the fact that the rectification algorithm may result in an distortion α in the image aspect ratio, which will affect all those horizontal image measurements, the complete set of user's testing measurements become $\hat{\mathbf{z}}(\alpha) = [\hat{\mathbf{z}}_\mathbf{V}, \hat{\mathbf{z}}_{\mathbf{I,v}}, \alpha\hat{\mathbf{z}}_{\mathbf{I,h}}]$, where $\hat{\mathbf{z}_\mathbf{V}}$ are the actual body dimensions; $\hat{\mathbf{z}}_{\mathbf{I,v}}$ and $\hat{\mathbf{z}}_{\mathbf{I,h}}$ represent vertical and horizontal image measurements, respectively. We look at the joint posterior of the GP estimator given an uncertain α.

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

where $\mu_\mathbf{y} = \mathbf{k}_\mathbf{Y}^\mathbf{T}(\hat{\mathbf{z}})\mathbf{K}_\mathbf{Y}^{-1}\mathbf{Y}$ and $\sigma_y^2 = k_Y(\hat{\mathbf{z}}, \hat{\mathbf{z}}) - \mathbf{k}_\mathbf{Y}^\mathbf{T}(\hat{\mathbf{z}})\mathbf{K}_\mathbf{Y}^{-1}\mathbf{k}_\mathbf{Y}(\hat{\mathbf{z}})$. Here, the prior $P(\alpha)$ of the aspect ratio α is modeled by a Gaussian distribution. Shape prediction and aspect ratio correction can thus be done by maximising the joint posterior with respect to $\mathbf{y}$ and α. Please refer to the full paper for more detailed explanations.

Extensive experiments have been carried out to evaluate our approach, including qualitative and quantitative tests on both synthetic doorway photos and real photos from volunteers. They support the efficacy of the system and also show that 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. The system presented in the paper is now being exploited commercially online.

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

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

[3] P. Guan, A. Weiss, A.O. Balan, and M.J. Black. Estimating human shape and pose from a single image. *ICCV*, 2009.

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

[5] 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.