Title
Predicting 3D body shape and body composition from conventional 2D photography

Permalink
https://escholarship.org/uc/item/8186v18t

Authors
Tian, Isaac Y
Ng, Bennett K
Wong, Michael C
et al.

Peer reviewed

Predicting 3D Body Shape and Body Composition from Conventional 2D Photography

Isaac Y. Tian,
Paul G. Allen School of Computer Science and Engineering, University of Washington, Seattle, WA, 98195, USA

Bennett K. Ng,
Intel Corporation, Santa Clara, CA, 95052, USA

Michael C. Wong,
University of Hawaii Cancer Center, University of Hawaii - Manoa, Honolulu, HI, 96813, USA

Samantha Kennedy,
Pennington Biomedical Research Center, Louisiana State University, Baton Rouge, LA, 70808, USA

Phoenix Hwaung,
Pennington Biomedical Research Center, Louisiana State University, Baton Rouge, LA, 70808, USA

Nisa Kelly,
University of Hawaii Cancer Center, University of Hawaii - Manoa, Honolulu, HI, 96813, USA

En Liu,
University of Hawaii Cancer Center, University of Hawaii - Manoa, Honolulu, HI, 96813, USA

Andrea K. Garber,
UCSF School of Medicine, University of California – San Francisco, San Francisco, CA, 94118, USA

Brian Curless,
Paul G. Allen School of Computer Science and Engineering, University of Washington, Seattle, WA, 98195, USA

Steven B. Heymsfield,
Pennington Biomedical Research Center, Louisiana State University, Baton Rouge, LA, 70808, USA

Corresponding Author: Isaac Yuheng Tian, MS, Paul G. Allen School of Computer Science & Engineering, University of Washington, Box 352350, Seattle, WA 98195-2350, Tel: (858) 729-8538, iytian@cs.washington.edu.

Conflict of Interest Disclosure

John Shepherd has research grants from Hologic, Inc and GE Healthcare.

Steven Heymsfield is on the Tanita Medical Advisory Board.

John A. Shepherd

University of Hawaii Cancer Center, University of Hawaii - Manoa, Honolulu, HI, 96813, USA

Abstract

Purpose: Total and regional body composition are important indicators of health and mortality risk, but their measurement is usually restricted to controlled environments in clinical settings with expensive and specialized equipment. A method that approaches the accuracy of the current gold standard method, dual-energy X-ray absorptiometry (DXA), while only requiring input from widely available consumer grade equipment, would enable the measurement of these important biometrics in the wild, enabling data collection at a scale that would have previously been prohibitive in time and expense. We describe an algorithm for predicting 3-dimensional body shape and composition from a single frontal 2-dimensional image acquired with a digital consumer camera.

Methods: Duplicate 3D optical scans, 2D optical images, and DXA whole body scans were available for 183 men and 233 women from the Shape Up! Adults Study. A principal component analysis vector basis was fit to 3D point clouds of a training subset of 152 men and 194 women. The relationship between this vector space and DXA-derived body composition was modeled with linear regression. The principal component 3D shape was then fitted to match a silhouette extracted from a 2D photograph of a novel body. Body composition was predicted from the resulting 3D shape match using the linear mapping between the principal component parameters and the DXA metrics. Accuracy of body composition estimates from the silhouette method was evaluated against a simple model using height and weight as a baseline, and against DXA measurements as ground truth. Test-retest precision of the silhouette method was evaluated using the duplicate 2D optical images and compared against precision of the duplicate DXA scans. Paired *t*-tests were performed to detect significant differences between the sets.

Results: Results were reported on a held-out set. Body composition prediction achieved R^2s of 0.81 and 0.74 for percent fat prediction of males and females, respectively, on a held-out test set consisting of 31 males and 39 females. Precision estimates for fat mass were 2.31% and 2.06% for males and females, respectively, compared to 1.26% and 0.68% for DXA scans. The *t*-tests revealed no statistically significant differences between the silhouette method measurements and DXA measurements, or between retests.

Conclusion: Total and regional body composition measures can be estimated from a single frontal photograph of a human body. Body composition prediction using consumer level photography can enable early screening and monitoring of possible physiological indicators of metabolic disease in regions where medical imagery or clinical assessment is inaccessible.

Keywords

Body Composition; Dual-energy X-ray absorptiometry; Principal Components Analysis; Silhouette; Obesity; Nutritional Assessment

I. INTRODUCTION

Predicting body composition has many useful clinical and research applications. Obesity is considered a primary risk factor for the development of type 2 diabetes, cardiovascular disease, and multiple forms of cancer. [1][2][3] Regional composition of selected body regions has been shown to be even more specific for prediction of the aforementioned health risks than whole body measures such as total body fat. Anthropometric surrogate measures of these regional tissue compartments such as waist circumference (WC), waist to hip ratio (WHR), surface markers of visceral adipose tissue (VAT) and related depots, have been shown to be stronger indicators of metabolic disease and mortality risk than total body fat. [4][5] Mid-upper-arm circumference (MUAC) is recognized by the World Health Organization as a marker of nutritional status, particularly in populations at risk for malnutrition.[6] Appendicular lean mass index is a marker for limb strength and can be used to diagnose muscle wasting disorders such as sarcopenia.[7] A criterion method for body composition assessment is Dual-Energy X-ray absorptiometry (DXA), an imaging technique that is currently considered the gold standard for measurement of total and regional body composition in clinical trials and research studies because of its precision and accuracy.[8] However, DXA is only available in specialized clinics and its use of ionizing radiation limits its frequent repetitive use.

The importance of body composition monitoring coupled with its high cost and low accessibility suggest a need for methods that can easily be used without access to a controlled clinical environment with cost prohibitive equipment and expertise to monitor the status of and changes in total and regional body composition compartments. Ideally, this technology would be affordable to middle- and low-income individuals, who are the populations most likely to be adversely affected by high costs and low access due to the increased risk of metabolic disease among lower socioeconomic brackets, and accessible through hardware that is widely distributed and commonly available outside of specialized clinics. Such a method would allow for measurement of body composition "in the wild" and would enable the outsourcing of body composition tracking from the professional clinic to the domestic household. This large-scale broadening of accessibility to clinically important body metrics can enable participation in self-monitoring and population health data analysis at previously infeasible scales. Commercial candidate solutions exist that are minimally invasive and relatively inexpensive by clinical standards. These include bioimpedance scales in both the bathroom scale format and in the tetrapolar configuration (BF-680W and MC-980U, Tanita Corporation, Arlington Heights, IL, USA). Although tetrapolar scales are more accurate and can provide more regional composition information, they cost between $12,000 and $20,000 and are generally only purchased by commercial gyms. Another candidate technology is air-displacement plethysmography (ADP) such as the BodPod (Cosmed, Rome, Italy). This device has been shown to be similarly accurate as DXA but does not provide regional measures and is laboratory based. 3D optical scanners have recently been shown to be able to accurately measure body circumferences and estimate body composition in both adults and children. [9][10] However, they too are not available for home use and can be expensive for individuals.

We propose a method for estimating fat and lean masses from a single front-facing 2D RGB photo taken from a consumer camera. Digital home photography is now easier and more

accessible than ever with the mass popularity of mobile devices in the last decade. Cameras, whether standalone or integrated into a phone, are general purpose-devices that are not purchased solely for the purpose of body composition evaluation. The hardware is already widely accessible to people even in the lowest income brackets, requiring no additional cost to obtain composition metrics: 95% of Americans making less than $30,000 a year own some kind of cell phone, and 71% own some kind of smart phone [11]. Such a method could remove the barrier to preventative care and diagnostic evaluations that tend to disproportionately impact communities underserved by the medical profession by outsourcing the data collection method to household devices that are readily available.

The objective of this study was to show that DXA body composition measurements could be reliably estimated using a photograph of a human body. We first created a model to estimate DXA body composition from 3D optical scans. We then synthesized a 3D body shape that best matched the binary silhouette of the human body in a 2D image taken in front of a green background and predicted the expected body composition from the parameters of the fitted 3D shape. The model for predicting DXA body composition from a 3D optical scan was thus extended to support a 2D optical image. We described the accuracy and precision of the 3D and 2D composition estimation models relative to DXA in a population of healthy adults.

II.   Materials and Methods

We performed a prospectively acquired cross-sectional study on adults with a wide variety of age, Body Mass Index (BMI), and ethnicities for both sexes. All participants received duplicate whole body DXA scans, 3D optical scans, and 2D color photos. Advanced statistical methods were used to relate 2D and 3D body shapes to DXA body composition. The accuracy of the optical methods to DXA as well as their test retest precision are described and reported below.

A.   Study Population and Procedures

Participants were recruited in the Honolulu, HI area at the University of Hawaii at Manoa, in the San Francisco, CA area at the University of California, San Francisco, and in the Baton Rouge, LA area at Pennington Biomedical Research Center as part of the Shape Up! Adults Study (NIH R01 DK109008). Recruitment was stratified by age (18–40, 40–60, > 60 years), ethnicity (non-Hispanic white, non-Hispanic black, Hispanic, Asian, and Native Hawaiian or Pacific Islander (NHOPI)), gender, and BMI (< 18, 18–25, 25–30, > 30 kg/m^2). Participants wore skintight underwear consisting of grey or black bike shorts and either a grey or black untextured and unstructured sports bra (women) or were shirtless (men). For optical scans, participants hid their hair in a swim cap. Following the Shape Up protocol, each participant underwent duplicate whole-body DXA and 3D Optical (3DO) scans, blood tests for diabetes and lipid biomarkers, as well as handgrip and thigh strength tests. Handgrip strength was measured as the average of three squeezes on a handgrip dynamometer (JAMAR 5030J1, Sammons Preston Rolyan, Nottinghamshire, UK) on each hand. Leg strength was measured as isokinetic and isometric knee extension and flexion on a HUMAC NORM (Computer Sports Medicine Inc., Stoughton, MA, USA) or Biodex Systems (Biodex Medical System

Inc., Shirley, NY, USA) dynamometer. Participants were excluded if they could not stand without aid for two minutes or lie flat for ten minutes without movement, had metal objects in their body, or previously had major body-shape-altering procedures (e.g., liposuction, amputations, etc.). Female participants were also excluded if pregnant or breast feeding. Written informed consent was obtained from each participant upon arrival and all procedures were approved by the Pennington Biomedical Research Center Institutional Review Board (IRB# 2016–053-PBRC), the UH Office Of Research Compliance (CHS# 2017–01018), and the Human Research Protection Program Institutional Review Board at the University of California, San Francisco (IRB# 15–18066). The study is publicly listed on ClinicalTrials.gov as ID NCT03637855.

B. DXA Scanning

As part of the data acquisition procedure for Shape Up, we captured two whole-body DXA scans, with body repositioning between scans, on either a Hologic Horizon/A system (UCSF) or a Discovery/A system (PBRC and UHCC) (Hologic Inc., Marlborough, MA, USA) for each participant. Participants were positioned and scanned according to each manufacturer's guidelines. All DXA scans were analyzed at UHCC by a single certified technologist using Hologic Apex version 5.6 with the National Health and Nutrition Examination Survey (NHANES) Body Composition Analysis calibration option disabled. DXA systems quality control was performed by monitoring the weekly values of the Hologic Whole Body Phantom. Cross calibration was checked between sites using a whole-body phantom scanned at each site. No cross-calibration adjustments were needed. [9] Body composition measurements from DXA included total and regional (trunk, arms, legs) measures of total fat mass and fat free (lean) mass (FFM). Percent fat (% fat) is represented as fat mass divided by total mass.

C. 3D Optical Scanning

For each participant, we also captured two 3DO whole-body surface scans on a Fit3D ProScanner (Fit3D, Inc., Redwood City, CA, USA). Subjects were repositioned between scans. Participants followed a manufacturer specified positioning protocol. The ProScanner captures 3D shape by rotating a stationary subject 360 degrees in front of one or more light-coding depth sensors. Scanning takes approximately 40 seconds to complete. The Iterative Closest Point (ICP) algorithm is used to align unorganized point clouds captured by the sensor as the subject rotates. [9] The final body-shape-approximating point cloud is converted to a triangle mesh with approximately 350,000 vertices and 700,000 faces. All 3DO scan data were transferred from the measurement sites and stored securely at UHCC prior to statistical analysis.

D. 2D Optical Scanning

Each participant was photographed twice in front of a green screen using a digital single-lens reflex (DSLR) camera and repositioned between the two photos. Participants stood in a neutral A-pose facing the camera with feet placed at fixed, marked locations on the floor 11 inches apart. This pose was chosen to best mimic the 3D optical pose. Each subject held a positioning bar that fixed the position of their arms such that their hands were 34.75 inches apart with straight elbows. Photos were de-identified by superimposing a black oval on the

face without obscuring the outline of the head. Images were captured in RAW format and converted into 16-bit linear TIFF files using an open-source software routine *dcraw*.

E. Constructing 3D-to-composition model

Our training procedure is described below; separate models were created for each gender:

1. Prepare inputs: ground truth 3D scans, DXA-derived body composition measures, 2D photographs.

2. Construct 3D shape space using Principal Component Analysis (PCA) from mesh templates fitted to ground truth 3D optical scans.[12]

3. Determine the best fit of a projection of the 3D model to the silhouette extracted from the 2D image.

4. Derive the body composition estimates from the PCA weight coefficients of the best fit 3D shape.

F. Applying 3D model to 2D images

The study procedure is then as follows for any new subject with input comprised of their height, weight, an RGB photo of the subject against a green screen, and the camera parameters:

1. Automatically detect 2D joint locations and segment subject from background. Manually correct any errors in the segmentation.

2. Initialize 3D shape with input height, weight. Initialize rigid transformation to align initial shape to detected joints on image. Fit 3D PCA shape to silhouette minimizing energy function E (described below).

3. Map optimized 3D PCA coefficients to body composition using the mapping learned in the training phase.

G. Training Procedure

Our pipeline mapped a 2D image to a 3D statistical shape, and then mapped the parameters of that shape to body composition statistics. The 3D statistical shape was represented by a PCA basis consisting of d column vectors of size $n = 180,003$. This PCA basis was constructed from eigen decomposition of a zero-mean-centered set of N body meshes represented as 1D column vectors of length 180,003, representing 60,001 3D points in XYZ interleaved format. Meshes were created by deforming a watertight template to fit ground truth 3D optical scans of each subject in the manner described by Allen *et al.* [12] (Fig. 1). Template fitting was required to maintain topological consistency and to give consistent positioning of vertex locations across subjects.

We can then describe any new body shape parameterized by this PCA basis as:

$$\mathbf{s}_{PCA} = \boldsymbol{\mu} + \sum w_i \mathbf{a}_i = \boldsymbol{\mu} + \mathbf{A}\mathbf{w} \qquad (1)$$

Where $\boldsymbol{\mu}$ is the mean of all training meshes, $\mathbf{A} = [\mathbf{a}_1 \ldots \mathbf{a}_d]$ is the PCA basis matrix, and $\mathbf{w} = [w_1 \ldots w_d]^{\mathrm{T}}$ is a length d vector of PCA coefficients that parameterize a given shape as an offset from the mean. The first 80 vectors of the PCA matrix sorted by descending eigenvalue represented just over 99% of the shape variance in the training meshes for both males and females. In Ng *et al.*[9], we used the first 15 vectors which only explained 95% of the shape variance. However, as more data became available, we found that 95% representation resulted in overly smoothed shape reconstructions that insufficiently captured details such as fatty skin folds. We defined dimensionality d as 80 for the rest of this work. We also recorded the corresponding standard deviations σ_i of each principal component defined as the square root of the explained variance. The standard deviations are useful for regularizing the space of anatomically plausible human body shapes, as we will explain later.

A key contribution of this work is the ability to map between a 3D shape and its associated body composition metrics. Ng *et al.*[9] defined a stepwise regression method mapping the first 15 PCA components to composition. We performed a simpler mapping using least squares and demonstrated that even such a naive method is quite effective despite using over five times the number of parameters.

For N training participants with M target features, we defined feature matrix $\mathbf{F}$ as:

$$\boldsymbol{F} = \begin{bmatrix} f_{1,1} & \cdots & f_{1,j} & \cdots & f_{1,N} \\ \vdots & & \vdots & & \vdots \\ f_{i,1} & \cdots & f_{i,j} & \cdots & f_{i,N} \\ \vdots & & \vdots & & \vdots \\ f_{M,1} & \cdots & f_{M,j} & \cdots & f_{M,N} \end{bmatrix} \tag{2}$$

Where the jth column in $\mathbf{F}$ represents the feature vector (for example, [height, weight, % fat]$^{\mathrm{T}}$) for subject j.

For the same N training participants, we define PCA weight matrix $\mathbf{W}$ as:

$$\boldsymbol{W} = \begin{bmatrix} w_{1,1} & \cdots & w_{1,j} & \cdots & w_{1,N} \\ \vdots & & \vdots & & \vdots \\ w_{i,1} & \cdots & w_{i,j} & \cdots & w_{i,N} \\ \vdots & & \vdots & & \vdots \\ w_{d,1} & \cdots & w_{d,j} & \cdots & w_{d,N} \end{bmatrix} \tag{3}$$

Where the jth column in $\mathbf{W}$ is the PCA basis projection of the body shape mesh of subject j in d reduced dimensions.

We defined augmented matrix $\widetilde{\mathbf{W}} = \begin{bmatrix} \mathbf{W} \\ 1 \end{bmatrix}$, and the following linear relationship:

$$\mathbf{F} = \mathbf{M}_{\boldsymbol{w} \rightarrow \mathbf{f}} \widetilde{\boldsymbol{W}} \tag{4}$$

The augmented row of ones is necessary to allow for a non-zero intercept for the linear relationship. Matrix $\mathbf{M_{w \to f}}$ now represents a linear transformation between a PCA coefficient vector $\mathbf{w}$ and the predicted features $\mathbf{f}$. We can solve for the least squares optimal solution for $\mathbf{M_{w \to f}}$ using the pseudoinverse $\widetilde{\mathbf{W}}^{+}$:

$$\mathbf{M_{w \to f}} = \mathbf{F} \widetilde{\mathbf{W}}^{+} \tag{5}$$

Conversely, we define augmented matrix $\widetilde{\mathbf{F}} = \begin{bmatrix} \mathbf{F} \\ \mathbf{1} \end{bmatrix}$ and:

$$\mathbf{W} = \mathbf{M_{f \to w}} \widetilde{\mathbf{F}} \tag{6}$$

$\mathbf{M_{f \to w}}$ maps a vector of feature priors to a predicted shape $\mathbf{w}$. This is useful for initializing our shape parameter vector, e.g., given easily measured features like height and weight, to increase the convergence speed and accuracy of our optimization as we describe in the next section. We solve for the least squares optimal matrix using the pseudoinverse again as above:

$$\mathbf{M_{f \to w}} = \mathbf{W} \widetilde{\mathbf{F}}^{+} \tag{7}$$

H.   Testing Procedure

The input to our algorithm was an RGB front-facing photo of a subject in a neutral pose in front of a green background, height of the subject in meters, weight of the subject in kilograms, camera intrinsic parameters comprised of focal length and sensor dimensions, and an estimate of the distance between the camera and the subject.

As a pre-process, we extracted the approximate joint locations and the detailed silhouette of the subject. Given the input photo (Fig. 2a), we performed CNN-based automatic joint detection on the RGB image (Fig. 2b) using DeepCut. [13] The joints were used to initialize a skeleton foreground label (Fig. 2c) for automatic segmentation using GrabCut. [14] It is important to get as close to pixel accuracy as possible for the silhouette of the subject; therefore, it is sometimes necessary to manually patch holes or erase background in the automatic result. We used this mask to extract the silhouette pixels $\{B_j\}$, defined as the set of all foreground pixels that neighbor a background pixel (Fig. 2d). In addition, corresponding 3D joint locations were picked manually on the average template mesh once, and the vertex indices were saved for all further joint location references on the 3D mesh.

Because each subject did not stand in precisely the same location relative to the camera, it was necessary to allow for a rigid transformation, $\mathbf{T}$, of the PCA space to maximize the alignment with the detected silhouette both before and during the fitting procedure. Our goal was to solve for the 3D body shape $\mathbf{s}_{\text{PCA}}(\mathbf{w})$ and camera transform $\mathbf{T}$ that best fits the subject seen in the 2D image. To achieve this fitting, we defined an objective comprised of multiple energy terms to be minimized together.

The first term $E_{\text{sil}}(\mathbf{w}, \mathbf{T})$ minimized the distance between the silhouette of the perspective projection of the 3D PCA shape and the silhouette of the 2D input image:

$$E_{\text{sil}}(\boldsymbol{w}, \boldsymbol{T}) = \sum_j \tau_j \text{dist}^2\big(B_j, \pi(\mathbf{T} \cdot \mathbf{s}_{\text{PCA}}(\mathbf{w}))\big) \tag{8}$$

where dist() measures the distance between image silhouette point B_j and the nearest compatible silhouette point of the PCA mesh $\mathbf{s}_{\text{PCA}}(\mathbf{w})$ transformed by $\mathbf{T}$ under camera projection π. Distances are weighted by τ_j depending on body part as described below. $E_{\text{sil}}(\mathbf{w}, \mathbf{T})$ is the sum of pairwise 2D distances between the image silhouette points $\{B_j\}$ and matched PCA silhouette vertices defined as $\pi(\mathbf{T} \cdot \mathbf{s}_{\text{PCA}}(\mathbf{w}))$. For every point on the image silhouette B_j, its nearest compatible PCA silhouette vertex was defined as the nearest transformed and projected neighbor that is a PCA silhouette vertex and shares a similar orientation.

A PCA silhouette vertex is a vertex whose normal is nearly orthogonal to the viewing ray, defined by the condition $\mathbf{e}_i \cdot \mathbf{n}_i < 0.05$ for vertex normal $\mathbf{n}_i$ and viewing direction $\mathbf{e}_i$ taken from the camera center of projection to the current vertex, both transformed by rigid transformation $\mathbf{T}$. We matched each image silhouette pixel to a PCA vertex by performing a nearest neighbor search across the set of candidate PCA silhouette vertices. The search was performed after the 3D PCA vertices were transformed by $\mathbf{T}$ and projected under perspective projection π to the same image coordinates as the image silhouette. We tracked the surface orientation of both the PCA boundary points and the image silhouette points. We rejected matches that did not have similar surface orientations to prevent incorrect registrations between different body surfaces due to poor alignment or initialization. Since deforming the PCA shape during fitting changes the candidate silhouette vertex coordinates, we repeat this registration in each iteration of the algorithm for intermediate shapes.

Additionally, limb misalignments were inevitable in our model as the 3D model our PCA space was trained on has no pose parameters. When participants were 3D scanned for the training set, everyone stood on the same footprints and grasped the same stationary handlebars, but differences in body proportions caused slight variations in limb angles and posture. The only way to attempt to match a discrepancy in limb alignment was to deform the entire body shape in the objective function. This deformation creates undesirable penalties in optimization energy when pose is slightly mismatched. Misaligned hands or feet contribute to large amounts of error in the energy function even if the rest of the body largely aligns. We introduced a term τ_j to give greater weight to the torso and hip silhouette points (6.0) relative to the limbs (1.0). We segmented the 3D average template mesh $\boldsymbol{\mu}$ in advance to identify points on the torso and hips.

The second term $E_{\text{joints}}(\mathbf{w}, \mathbf{T})$ is the sum of squared distances between the CNN-detected joints and the transformed and projected joint vertices on the 3D PCA model.

$$E_{\text{joints}}(\mathbf{w}, \mathbf{T}) = \sum_k \text{dist}^2\big(J_k, \pi\big(\mathbf{T} \cdot \mathbf{J}_k^{\text{PCA}}(\mathbf{w})\big)\big) \tag{9}$$

where J_k is the kth detected joint on 2D image and $\mathbf{J}_k^{\mathrm{PCA}}(\mathbf{w})$ is the kth joint vertex on the 3D PCA mesh.

Joint vertices were picked once on the average template shape $\boldsymbol{\mu}$. Because topological consistency was guaranteed when the average shape was deformed to some new shape $\mathbf{w}$, the labeled joints had the same joint indices and were in approximately the same anatomical location. We used 10 joints representing shoulders, hips, knees, and ankles, plus a vertex for the crown of the head and a vertex for the base of the neck defined as the midpoint of the clavicles. This term provided a loose constraint on anatomical consistency for the fitting and favors a shape that has similar limb proportions under camera projection. Note that the detected elbows and wrists were not used in this term; arm position was highly variable and would have introduced noise to the fit.

The next two terms E_{height} and E_{mass} are regularizers based on the known prior height and mass of the subject to improve the anatomical accuracy of the shape fit:

$$E_{\mathrm{mass}}(\mathbf{w}) = (\mathbf{M}_{\mathbf{w} \to \mathbf{f}} \mathbf{w}[i_m] - m_0)^2 \tag{10}$$

$$E_{\mathrm{height}}(\mathbf{w}) = (||\mathbf{v}_{\mathrm{crown}}(\mathbf{w}) - \mathbf{v}_{\mathrm{heel}}(\mathbf{w})|| - h_0)^2 \tag{11}$$

$E_{\mathrm{mass}}(\mathbf{w})$ is the squared difference between the input known body mass m_0 and the predicted body weight using mapping matrix $\mathbf{M}_{\mathbf{w} \to \mathbf{f}}$ and the PCA shape vector of the estimated $\mathbf{w}$. $\mathbf{M}_{\mathbf{w} \to \mathbf{f}}$ in general produces a vector of body features: i_m gives the index of the total body mass feature in this vector. The predicted height was calculated simply as the squared 3D distance between a vertex at the crown of the head and a vertex at the base of the heel of the PCA model. The position of these vertices are functions of $\mathbf{w}$. $E_{\mathrm{height}}(\mathbf{w})$ is defined as the squared difference between this predicted height and the input height h_0.

The last term $E_\sigma(\mathbf{w})$ penalizes for large magnitudes of PCA shape vector $\mathbf{w}$, biasing the solution towards the mean. It is a weighted L2 regularization:

$$E_\sigma(\mathbf{w}) = \sum_{i=1}^{d} \left(\frac{w_i}{\sigma_i} \right)^2 \tag{12}$$

where w_i is the ith element of vector $\mathbf{w}$ and σ_i is the standard deviation of ith PCA vector. This regularizer prevents overfitting to the silhouette at the expense of producing unrealistic and unlikely body shapes. Shapes that are multiple standard deviations away from the mean (defined as $w_i = 0$ for all i) receive a larger penalty than shapes that deformed minimally from the origin (the mean).

We can now define the full energy function E as:

$$E(\mathbf{w},\mathbf{T}) = E_{\mathrm{sil}}(\mathbf{w},\mathbf{T}) + \alpha E_{\mathrm{joints}}(\mathbf{w},\mathbf{T}) + \beta E_{\mathrm{height}}(\mathbf{w}) + \gamma E_{\mathrm{mass}}(\mathbf{w}) + \lambda E_{\sigma}(\mathbf{w}) \tag{13}$$

where α, β, γ, and λ are hyperparameters that determine the relative influence of each term in the energy function.

Due to the mesh projection step and the association of nearest compatible points, this is a non-linear objective. We iteratively optimized for $\mathbf{w}$ and $\mathbf{T}$ by minimizing $E(\mathbf{w},\mathbf{T})$ using the Ceres[15] implementation of the Levenberg-Marquardt algorithm until the change in parameters $\mathbf{w}$ from the previous iteration was less than some cutoff ε. This difference was defined as the root sum of squared difference between the two vectors. Hyperparameters for (13) are listed in Table I.

Using mapping matrix $\mathbf{M_{f \to w}}$, with $\mathbf{f}$ containing height and weight, we initialized shape parameters $\mathbf{w}$ as the PCA shape $\mathbf{w}_0 = \mathbf{M_{f \to w}}\mathbf{f}_0$ where $\mathbf{f}_0 = [\text{height, weight, 1}]^{\mathrm{T}}$. This step initialized the PCA coefficients to an average person with the given height and weight, which increases the initial alignment with the target silhouette.

We initialized rigid transformation $\mathbf{T}$ by solving for the minimization of $E_{\mathrm{joints}}(\mathbf{w}_0, \mathbf{T})$ with $\mathbf{w}_0$ fixed. A summary of our optimization loop is given in Algorithm 1. A visualization of the shape terms E_{sil} and E_{joints} is shown in Fig. 3.

Algorithm 1: 3D PCA to 2D Silhouette Alignment

Initialize $\mathbf{w}$ as $\mathbf{w}_0 = \mathbf{M_{f \to w}}\mathbf{f_0}$, where $\mathbf{f_0}$ is $[\text{height, weight, 1}]^{\mathrm{T}}$
Initialize $\mathbf{T}$ as $\min E_{\mathrm{joints}}(\mathbf{w}_0, \mathbf{T})$
while $\|\Delta\mathbf{w}\| < \varepsilon$ **do**
    Recompute normals of intermediate shape $\mathbf{s}_{\mathrm{PCA}}(\mathbf{w})$;
    for each image silhouette point B_j,
        find closest compatible point from $\pi(\mathbf{T}(\mathbf{s}_{\mathrm{PCA}}(\mathbf{w})))$;
    Minimize $E(\mathbf{w},\mathbf{T})$;
    Update $\mathbf{w}$ and $\mathbf{T}$ with converged result from this alignment
end
$\tilde{\mathbf{w}} = [\mathbf{w}, 1]^{\mathrm{T}}$
Compute body features $\mathbf{f} = \mathbf{M_{w \to f}}\tilde{\mathbf{w}}$;
return f

I. Statistical Evaluation

We tested our method on a randomly selected held-out test set of 31 males and 39 females. Hyperparameters for reported results were chosen as indicated in Table I based on performance on a single male subject. Test set participants were not included in the PCA space construction, nor were they included in computing the mapping from PCA to body features. We performed 5-fold cross validation on this construction to verify the consistency of the PCA to composition regression. This was done by making $k = 5$ random folds of all

subjects and creating 5 PCA spaces using each combination of k-1 folds. For each PCA space, we performed linear regression between its fold members and their associated body statistics and reported validation results on the held-out fold representing 20% of total subjects. The experimental fold that we reported in the results section was a separate random fold and was not any of the above folds. Cross validation was necessary to demonstrate that our results are repeatable on arbitrary principal component spaces provided there is sufficient representation of body shapes and not just on a particularly favorable training – test split selected for this experiment.

We reported root-mean-square-error (RMSE) and the coefficient of determination (R^2) of our regression results from our predicted shapes using DXA measurements as the ground truth. We compared our predictions to a few different diagnostic scenarios to demonstrate the predictive quality of our silhouette fitting method. The lower bound scenario was demonstrated by predicting all body composition metrics on a simple linear regression from the known input scalars, height and weight, without any body geometry fitting. The upper bound scenario was demonstrated by taking the ground truth 3D scans of the test set and projecting them into principal component space by performing the inverse operation of (1); that is, subtracting out the mean shape and multiplying by the transpose of the PCA matrix. This produced a PCA coordinate vector that represented the projection of the 3D scan onto the principal component basis to give a prediction using the best possible geometric fit. We also reported the RMSE and R^2 of our 5-fold cross validation, using the sum total of prediction to ground-truth pairs across all 5 folds to compute these metrics. This demonstrated the robustness of the method against overfitting.

To ensure that our method is robust to natural variability in body pose and positioning we performed a test-retest precision evaluation on the experimental fold. Specifically, we evaluated a second set of images of the same test participants and compared predicted measurements against those from the first set of images. Participants were repositioned between the two images, and thus stood in slightly different poses and positions. Precision of the 2D estimates was compared to the precision estimates from duplicate DXA scans. Coefficient of variation (%CV) results, defined as Glüer *et al* [16] as the ratio of the standard deviation of repeat measurements to the mean of repeat measurements averaged across all test subjects, are shown in Table II and an example 3D to 2D fit in Fig. 4.

We performed a paired *t*-test on the test-retest trials for our method, the test-retest scans of DXA, and on the difference between our method and the DXA measurements. Since there were 12 different body composition measurements evaluated, a Bonferroni-corrected critical *P*-value of 0.05 / 12 = 0.004 was considered significant.

III. Results

Repeatability comparison to the DXA gold standard of measuring % fat is shown in Table II and represented as the coefficient of variation (CV). RMSE and R^2 values between the test and retest trials are also shown. %CV and RMSE values for our method were around 2–3 times larger than those from DXA. R^2 are all greater than 0.90 and are comparable to the DXA equivalents with the exception of female visceral fat and leg fat, at $R^2 = 0.60$ and 0.85

respectively. While reduced precision in limb compartment estimates may be explained by the lack of consistent pose alignment between photos of the same subject and the inability of our shape model to account for pose differences independent of body shape, the visceral fat imprecision suggests that particular measurement is not well modeled in females by our method.

The R^2 and RMSE values of every predicted body composition metric are shown in Table III and Table IV. In Table III we compared our results to 1) the 5-fold cross validation performance of each feature representing an estimate of the expected performance of the regression method on scans with known shape and PCA vectors, 2) the prediction produced only by a linear regression of the known BMI of the subject, 3) the prediction produced only by a linear regression of the known initialization variables [height, weight] to each of the desired features, and 4) the prediction using the projection of the 3D scan of each subject to PCA basis space. The 5-fold cross validation comparison was necessary to demonstrate that our held-out test set was fairly representative of the predictive capabilities of the PCA method sampled across multiple training – test splits, rather than being an overperforming outlier set picked for the purposes of this publication. Comparison to linear regression using only BMI demonstrates the predictive power of this method relative to a common scalar analogue for % fat. Comparison to linear regression with the variables [height, weight] may seem redundant, but it is necessary to demonstrate that the silhouette fitting method adds predictive accuracy to the baseline input information of height and weight and represents a lower bound for performance. As this method is intended to be accessible to a nonprofessional audience, height and weight were chosen to be the initializer variables rather than BMI. We show that in every predicted variable, the silhouette fitting method improves upon the lower bound predictions that would have been available from using the initialization variables alone for both BMI and height + weight. Females were more accurately predicted by the initialization variables alone, showing 20% decreases in RMSE from the initialization result to the shape fitted result in fat and lean mass, as opposed to males which exhibited almost a 40% decrease.

The prediction using the projected PCA coordinates of the 3D scan represented a rough upper bound of the prediction capability of the method. It is the approximate best-case scenario of the regression function assuming shape prediction was perfect. This allowed us to evaluate how effective the shape fitting was at improving composition prediction independent of the noise inherent in the regression functions. However, this was not an exact upper bound because subjects were not photographed and scanned in the exact same motionless position. This introduced some variance to the shape caused by slight differences in limb pose and posture, which our shape model is currently not capable of separating from body shape. Some metrics in females, such as lean mass, showed higher R^2 and lower RMSE in our test prediction from 2D data than from the best case 3D shape projection as a result.

Fat mass and fat free mass (FFM) estimates for females showed an RMSE of almost 40% lower than those for males. For trunk fat mass and fat free mass, females were 16% and 27% lower, respectively. Percent fat (% fat) was calculated in two ways: first by dividing the predicted fat mass by the known input body mass, and then by directly predicting percent fat

as a feature in the linear regression described by (4). The first method achieved 15% lower RMSE on females, which is consistent with their lower fat mass error. However, linear regression of the percent fat variable produced the opposite effect, with males having 15% lower RMSE than females. We treat the first method as the standard method in future references to percent fat to be consistent with previous work. Every limb compartment fat and fat free mass estimate had lower RMSE for females, there was an accepted amount of limb misalignment for both genders due to pose variations in the dataset. Visceral fat was the only measurement for which the model for males notably outperformed the model for females (R^2 of 0.66 and 0.36, respectively).

Table IV compares our results, which starts from a 2D input (camera photo), to Ng *et al.*, [9] which starts from a 3D scan. We show that our method is comparable to this related method that also used PCA to predict body composition variables despite an additional step that requires predicting the 3D body shape from the silhouette, rather than having the ground truth 3D shape as input. RMSE in our method was 7% higher in fat and lean mass for males, but 23% lower in females.

Table V shows *p*-values for a paired *t*-test performed on three pairs of body composition measurement sets: DXA retrials, test-retest of our method, and our method against DXA. T1 vs DXA1 tested the accuracy of our method (T1) against the accepted ground truth (DXA1). Although a few tests produced *p*-values below a single-test critical value of 0.05, none were below the Bonferroni corrected critical *p*-value of 0.004. Importantly, total body fat and lean mass along with percent fat all greatly exceeded the individual significance level of 0.05. Thus, the mean differences between retrials and between our method and the DXA measured composition variables were not statistically significantly different from zero.

We show some examples of our method on individual subjects from the test set in Table VI. From left to right, we show the input 2D photo, the initial shape as predicted by input height and weight, the extracted silhouette from the 2D photo aligned with the initial shape, the optimal converged shape aligned with the same silhouette, and the 3D scan. The 3D scan cannot be regarded as explicitly ground truth because subjects were not scanned in the exact same pose or location as the 2D photo, but it shows the level of detail that can be expected of an actual optical scanner compared to our prediction method. On individual examples, percent fat prediction accuracy ranged from <1% to as high as 6%. Because our method was not able to factor in depth cues such as the shading of the torso region, indicating either a convex abdomen or a lean figure with defined musculature, many of the higher error examples tended to have proportions that were not well predicted by the silhouette alone. Subjects that had average waist breadth but were deep in the sagittal plane tended to be underpredicted in fat mass and percent fat, while subjects that were wide shouldered and muscular while being somewhat lean tended to be overpredicted.

IV. Discussion

In the current study we demonstrated that composition of a human body can be inferred from a 2D silhouette taken from an RGB image given known height and weight. Previous publications have presented work in both computer vision and medical research that parallel

parts of our project, but to the best of our knowledge, no other publication has gone from a single 2D image to body composition estimates using 3D shape prediction as an intermediate. Guan *et al.* [17] presented an early method of mapping a 3D human shape space to a single monocular RGB image. This method has the advantage of modeling pose variation and shading, which ours does not, but there is no subsequent mapping to clinical metrics. Bogo *et al.* [18] used a more advanced posable shape model, the skinned multi-person linear model (SMPL), to estimate a 3D shape from arbitrary poses, but the actual 2D to 3D mapping was based solely on joint projections without silhouette fitting, resulting in very coarse fits. Using Shape Up! 3D optical depth scans, we had previously derived a PCA model of body shape and related those PCA vectors to criterion body composition measures from DXA. Here we extend that work using only the 2D photograph, the camera focal length, and the subject's height and weight to predict the PCA parameterized body shape in cases where 3D depth scans are not available. We estimated the composition of these predicted body shapes using linear regression from PCA parameters to criterion measures derived from DXA. Affuso *et al.* [19] presented a method that uses both front and side images to generate features for a support vector regression that achieved an R^2 of 0.78 for percent fat across all adults in 3-fold cross validation. Our method achieved R^2 of 0.73 and 0.74 on randomly held-out sets of males and females respectively using only a single frontal image, with 5-fold cross validation results showing 0.68 and 0.77 respectively. Unlike this work, we separated our experiment by males and females and did not include children. Farina *et al.* [20] presented a method that predicts fat mass from a single side-profile photograph. We believe our method is more robust due to the larger sample size (152 males, 194 females compared to 54 males, 63 females) and verification on a separate held-out set. The R^2 values greater than 0.95 in Farina *et al.* appear to be reported on the training set, leaving the generalizability of this method uncertain. Furthermore, the methods are not reproducible because they depend on an undisclosed, proprietary body segmentation algorithm as part of their training procedure. More recently, Lu *et al.* [21] predicted body fat directly from a 3D body mesh with machine learning methods. This method was trained on a limited sample of 50 adult males and makes the prediction on a 3D scan with a minimum RMSE on percent fat of 3.17. This result was reported using the leave-one-out method, where training was performed on n-1 samples and testing done on just one. Our method achieved comparable RMSE of 3.9 and 3.3 on males and females respectively, using one consistent model on a randomly selected held-out test set and only requiring a 2D photo, height, and weight as input.

Although effective, our method could be improved by going beyond silhouettes and including shading information in the input images. Guan *et al.* [17] demonstrated a method that optimizes geometry to explain the observed shading over the surface of the subject with a single light source. Although the shading model was not based on human skin reflectance models, it was shown to improve the fit to the silhouette and pose of images that feature human participants in differing poses. Including a shading term in our optimization could produce more accurate 3D reconstructions, as we currently only use the silhouette pixels and ignore the interior pixel information. While Guan *et al.* only used the shading term to enhance the geometric similarity between predicted shapes and ground truth geometry, this additional detail may enhance the accuracy of our body composition prediction.

Our shape models in this work were not constructed to explicitly handle pose-dependent shape variation. A posable model with joint angle parameters would allow pose to be optimized separately from "intrinsic" body shape, as in Guan *et al.* and *Bogo et al.* [17][18] Although our pose space is constrained to only frontal images of participants standing on footprints with handlebars, the amount of variation between people of different sizes fixing their extremities to static points in space is substantial enough to affect the PCA formulation. Differences in the lean, leg spread, and arm spread were misconstrued as fundamental body shape variations by our PCA model. This pose variation causes fitting issues when differences in leg position cannot be isolated from height or girth, or conversely when limbs cannot be matched without compromising the accuracy of the torso alignment. Building our PCA model on top of a posable model such as SMPL will allow us to isolate pose from shape and theoretically produce better reconstructions and results.

In the absence of a posable model that can account for variations in arm and leg angles, we created a demo of a smartphone app that facilitates the collection of 2D image data in the wild for non-professionals. Our app projected a stick figure to the camera screen of the phone, indicating to the photographer how the subject should be aligned in frame to best fit the expected pose of the PCA space. Silhouette accuracy is extremely important and requires near pixel accurate segmentation of the human body, ideally clothed with no more than a skintight bathing suit equivalent. While this is easy to accomplish with standard methods against a green screen background, reliable automatic segmentation against arbitrary real-world backgrounds such as the one shown in Fig. 5 requires more advanced computer vision methods that are beyond the scope of this work.

Our mapping function $\mathbf{M}$ was assumed to be linear and derived from a simple least-squares regression. It is possible that a more ideal function can be more complex, such as a polynomial kernel or a neural network function, an area for future work. Our initial experiments using fully connected networks were unsuccessful as the predictions were very quickly overfitted.

As with all machine learning based methods, our predictive power is strongly based on the quality and variety of training data. Additional training data should add to the robustness and consistency of the model.

Finally, hyperparameters from Table I were tuned by trial and error on a single randomly chosen individual. Ideally, we would tune our hyperparameters on a third, held-out set that is not part of either the training or test set to tune our hyperparameters on (the validation set). Due to the low subject count, we did not further fragment our subject set to robustly optimize the many hyperparameters.

V. Conclusion

Frontal body silhouette provides substantial information on the body composition of a subject in the absence of other views or additional imaging information such as depth. This method requires minimal data inputs and can be employed in a much wider scope of practice than traditional medical imaging methods. Given the clinical significance of both total and

regional body adiposity for predicting metabolic disease and mortality risk, our method may be an impactful first step in propagating low-cost early screenings that can be performed outside of medical clinics by non-professionals for patients that may not warrant or cannot afford a clinical evaluation and gold-standard medical imaging. Future implementations of this project can deploy this algorithm to mobile devices, making it an attractive low-cost approximation of advanced imaging in more remote areas with lower rates of medical access.

Acknowledgment

Abbreviations:

3D	three-dimensional
BMI	body mass index
CAESAR	Civilian American and European Surface Anthropometry Resource Project
DXA	dual-energy X-ray absorptiometry
DSLR	digital single-lens reflex camera
FFM	Fat free mass
ICP	Iterative Closest Point algorithm
PCA	principal component analysis
RMSE	root-mean-square error

References

[1]. Zhang C, Rexrode KM, Dam RMV, Li TY, Hu FB. Abdominal Obesity and the Risk of All-Cause, Cardiovascular, and Cancer Mortality. Circulation. 2008;117(13):1658–1667. doi:10.1161/circulationaha.107.739714 [PubMed: 18362231]

[2]. Eckel RH, Kahn SE, Ferrannini E, et al. Obesity and Type 2 Diabetes: What Can Be Unified and What Needs to Be Individualized? The Journal of Clinical Endocrinology & Metabolism. 2011;96(6):1654–1663. doi:10.1210/jc.2011-0585 [PubMed: 21602457]

[3]. Calle EE, Kaaks R. Overweight, obesity and cancer: epidemiological evidence and proposed mechanisms. Nature Reviews Cancer. 2004;4(8):579–591. doi:10.1038/nrc1408 [PubMed: 15286738]

[4]. Price GM, Uauy R, Breeze E, Bulpitt CJ, Fletcher AE. Weight, shape, and mortality risk in older persons: elevated waist-hip ratio, not high body mass index, is associated with a greater risk of death. The American Journal of Clinical Nutrition. 2006;84(2):449–460. doi:10.1093/ajcn/84.2.449 [PubMed: 16895897]

[5]. Kuk JL, Katzmarzyk PT, Nichaman MZ, Church TS, Blair SN, Ross R. Visceral Fat Is an Independent Predictor of All-cause Mortality in Men*. Obesity. 2006;14(2):336–341. doi:10.1038/oby.2006.43 [PubMed: 16571861]

[6]. Mramba L, Ngari M, Mwangome M, et al. A growth reference for mid upper arm circumference for age among school age children and adolescents, and validation for mortality: growth curve construction and longitudinal cohort study. Bmj. March 2017. doi:10.1136/bmj.j3423

[7]. Imboden MT, Swartz AM, Finch HW, Harber MP, Kaminsky LA. Reference standards for lean mass measures using GE dual energy x-ray absorptiometry in Caucasian adults. Plos One. 2017;12(4). doi:10.1371/journal.pone.0176161

[8]. Lu Y, Mathur AK, Blunt BA, et al. Dual X-ray absorptiometry quality control: Comparison of visual examination and process-control charts. Journal of Bone and Mineral Research. 2009;11(5):626–637. doi:10.1002/jbmr.5650110510.

[9]. Ng BK, Sommer MJ, Wong MC, et al. Detailed 3-dimensional body shape features predict body composition, blood metabolites, and functional strength: the Shape Up! studies. The American Journal of Clinical Nutrition. 2019;110(6):1316–1326. doi:10.1093/ajcn/nqz218 [PubMed: 31553429]

[10]. Wong MC, Ng BK, Kennedy SF, et al. Children and Adolescents' Anthropometrics Body Composition from 3-D Optical Surface Scans. Obesity. 2019;27(11):1738–1749. doi:10.1002/oby.22637 [PubMed: 31689009]

[11]. Demographics of Mobile Device Ownership and Adoption in the United States. Pew Research Center: Internet, Science & Tech https://www.pewresearch.org/internet/fact-sheet/mobile/. Accessed March 18, 2020.

[12]. Allen B, Curless B, Popovi Z. The space of human body shapes. ACM SIGGRAPH 2003 Papers on - SIGGRAPH 03. 7 2003. doi:10.1145/1201775.882311

[13]. Pishchulin L, Insafutdinov E, Tang S, et al. DeepCut: Joint Subset Partition and Labeling for Multi Person Pose Estimation. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2016. doi:10.1109/cvpr.2016.533

[14]. Rother C, Kolmogorov V, Blake A. "Grabcut": Interactive foreground extraction using iterated graph cuts. ACM Trans. Graph, 23(3):309–314. 8 2004. doi: 10.1145/1186562.1015720

[15]. Agarwal S, Mierle K, et al. Ceres Solver. http://ceres-solver.org

[16]. Glüer C, Blake G, Lu Y, Blunt B, Jergas M, Genant H. Accurate assessment of precision errors: How to measure the reproducibility of bone densitometry techniques. Osteoporosis International, 1995;5(4), pp.262–270. [PubMed: 7492865]

[17]. Guan P, Weiss A, Balan AO, Black MJ. Estimating human shape and pose from a single image. 2009 IEEE 12th International Conference on Computer Vision 2009. doi:10.1109/iccv.2009.5459300

[18]. Bogo F, Kanazawa A, Lassner C, Gehler P, Romero J, Black MJ. Keep It SMPL: Automatic Estimation of 3D Human Pose and Shape from a Single Image. Computer Vision – ECCV 2016 Lecture Notes in Computer Science. 2016:561–578. doi:10.1007/978-3-319-46454-1_34

[19]. Affuso O, Pradhan L, Zhang C, et al. A method for measuring human body composition using digital images. Plos One. 2018;13(11). doi:10.1371/journal.pone.0206430

[20]. Farina G, Spataro F, Lorenzo AD, Lukaski H. A Smartphone Application for Personal Assessments of Body Composition and Phenotyping. Sensors. 2016;16(12):2163. doi:10.3390/s16122163

[21]. Lu Y, Mcquade S, Hahn JK. 3D Shape-based Body Composition Prediction Model Using Machine Learning. 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) 2018. doi:10.1109/embc.2018.8513261
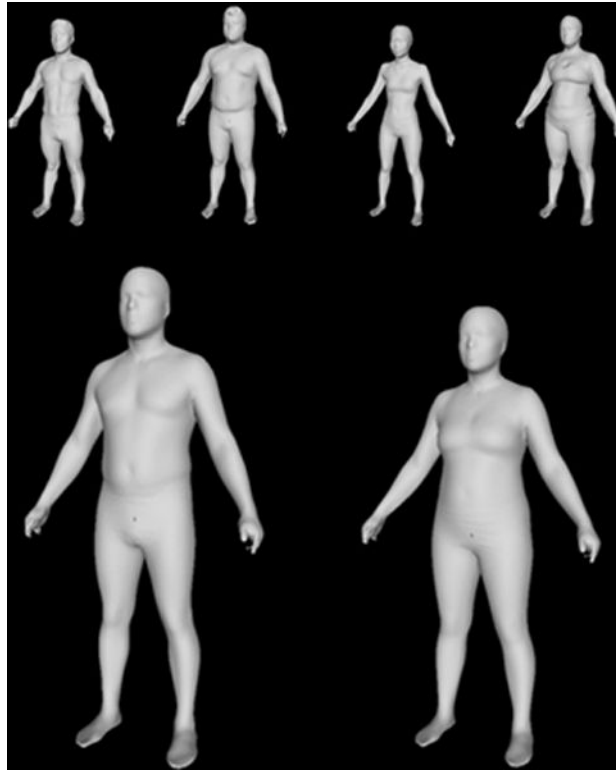
Fig. 1.
Top: Examples of template fitted 3D scans used to construct PCA space. Bottom: The mean male and female shape μ that are the starting points for all shape deformations.
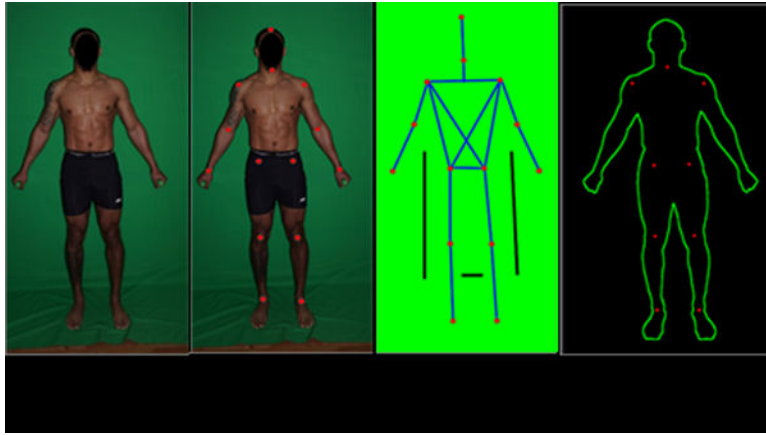
Fig. 2.
Example of preprocessing an input image. a: the input RGB image. b: CNN detected joints. c: skeleton foreground seed label (blue) created by connecting detected joints. Background initialized as black marked lines. Initializations are done automatically. d: extracted silhouette (green) and joints used for fitting (red).
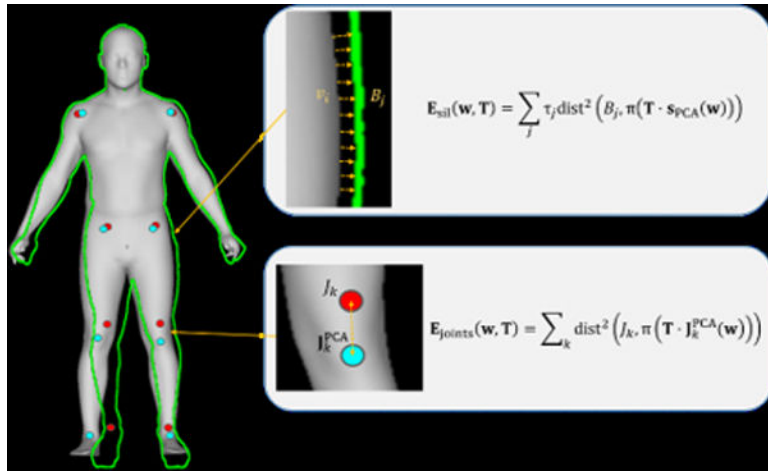
Fig. 3.
Visualization of the initial projected shape $\mathbf{w}_0$ overlaid onto the target silhouette (green). This projected 3D shape is fit by minimizing the closest pairwise distances between a boundary vertex and its closest silhouette point (top box) and by minimizing distances between detected joints on the silhouette (red) and the projected mesh joint vertices (blue) (bottom box).

Fig. 4.
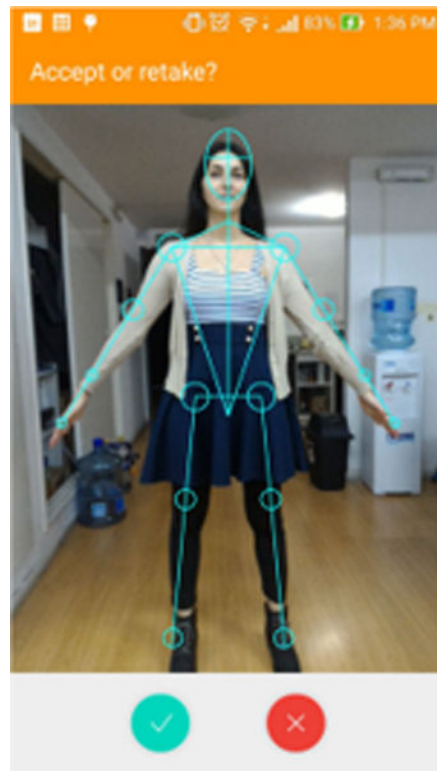An example of a final aligned shape projected onto the target silhouette.

Fig. 5.
Smartphone app screenshot indicating pose alignment landmarks.

TABLE I

<small>HYPERPARAMETER OPTIMAL VALUES</small>

Parameter	Description	Value
τ	Silhouette match weight	6.0 torso, 1.0 else
d	# of PCA components	80
α	Joint alignment weight	3.0
β	Height (m) alignment weight	5.0
γ	Mass (kg) alignment weight	1.0
λ	PCA std. dev. weight	0.001
ε	Convergence condition	0.3

TABLE II

TEST-RETEST PRECISION

	This Work						DXA					
	Male (n=31)			Female (n=39)			Male (n=31)			Female (n=39)		
	%CV	R²	RMSE	%CV	R²	RMSE	%CV	R²	RMSE	%CV	R²	RMSE
FMI [kg/m²]	2.40	**0.99**	**0.161**	2.19	**0.99**	**0.210**	1.27	**1.0**	**0.084**	0.68	**1.0**	**0.064**
FFMI [kg/m²]	0.78	**0.99**	**0.168**	1.22	**0.98**	**0.215**	0.37	**1.0**	**0.078**	0.44	**1.0**	**0.076**
Fat Mass [kg]	2.31	**0.99**	**0.469**	2.06	**0.99**	**0.512**	1.26	**1.0**	**0.252**	0.68	**1.0**	**0.168**
FFM [kg]	0.72	**0.99**	**0.469**	1.12	**0.98**	**0.512**	0.37	**1.0**	**0.232**	0.44	**1.0**	**0.199**
Percent Fat [%]	--	**0.97**	**0.502**	--	**0.94**	**0.671**	--	**1.0**	**0.242**	--	**0.99**	**0.243**
Visceral Fat [kg]	2.87	**0.98**	**0.016**	15.21	**0.60**	**0.065**	4.58	**0.96**	**0.023**	5.75	**0.96**	**0.022**
Trunk Fat Mass [kg]	2.67	**0.99**	**0.313**	2.76	**0.98**	**0.323**	2.21	**0.99**	**0.222**	1.73	**0.99**	**0.197**
Trunk FFM [kg]	0.62	**1.0**	**0.201**	1.68	**0.96**	**0.386**	0.93	**0.99**	**0.280**	0.817	**0.99**	**0.183**
Arms Fat Mass [kg]	3.64	**1.0**	**0.089**	3.96	**1.0**	**0.136**	2.49	**0.99**	**0.030**	2.12	**0.99**	**0.033**
Arms FFM [kg]	2.12	**0.96**	**0.195**	3.29	**0.90**	**0.173**	1.23	**0.99**	**0.052**	1.36	**0.98**	**0.032**
Legs Fat Mass [kg]	2.93	**0.98**	**0.193**	6.43	**0.85**	**0.636**	1.30	**1.0**	**0.042**	1.09	**1.0**	**0.050**
Legs FFM [kg]	0.92	**0.99**	**0.196**	1.25	**0.98**	**0.187**	0.93	**0.99**	**0.096**	0.80	**0.99**	**0.059**

TABLE III

RESULTS FOR ALL MEASURED COMPOSITION METRICS

Output Variable	Gender	Model 1: Combined 5-fold cross validation on all available scans		Model 2: BMI regression only on test set		Model 3: Height & Weight regression on test set		Model 4: PC to body metrics regression using projected PCs from test set scans		Model 5: PC to body metrics regression using predicted body shape from image	
		R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE
Fat Mass [kg]	Male	**0.90**	0.90	**0.74**	5.88	**0.75**	5.78	**0.96**	**2.27**	**0.90**	3.63
	Female	**0.94**	0.94	**0.86**	3.43	**0.91**	2.83	**0.94**	**2.35**	**0.94**	2.29
FFM [kg]	Male	**0.93**	0.93	**0.31**	9.26	**0.73**	5.78	**0.94**	**2.78**	**0.89**	3.63
	Female	**0.91**	0.91	**0.57**	5.20	**0.87**	2.83	**0.89**	**2.70**	**0.92**	2.29
% Fat	Male	**0.68** (0.76)	0.68 (0.76)	**0.41** (0.46)	5.72 5.44	**0.41** (0.50)	5.71 (5.27)	**0.90** (0.90)	**2.36** (2.45)	**0.725** (0.806)	3.90 (3.27)
	Female	**0.77** (0.76)	0.77 (0.76)	**0.50** (0.54)	4.56 (4.43)	**0.65** (0.56)	3.86 (4.31)	**0.75** (0.74)	**3.06** (3.38)	**0.74** (0.631)	3.29 (3.94)
FMI [kg/m²]	Male	**0.89**	0.89	**0.75**	1.86	**0.75**	1.87	**0.96**	**0.73**	**0.90**	1.19
	Female	**0.94**	0.94	**0.88**	1.23	**0.91**	1.05	**0.93**	**0.91**	**0.94**	0.85
FFMI [kg/m²]	Male	**0.91**	0.91	−0.42	3.26	**0.53**	1.87	**0.90**	**0.90**	**0.81**	1.19
	Female	**0.89**	0.89	**0.43**	2.07	**0.85**	1.05	**0.87**	**1.02**	**0.91**	0.85
Visceral Fat Mass [kg]	Male	**0.67**	0.67	**0.12**	0.23	**0.18**	0.23	**0.75**	**0.13**	**0.66**	0.15
	Female	**0.76**	0.76	**0.25**	1.89	**0.27**	1.86	**0.71**	**0.12**	**0.36**	0.17
Trunk Fat Mass [kg]	Male	**0.92**	0.92	**0.72**	3.35	**0.74**	3.25	**0.95**	**1.40**	**0.92**	1.76
	Female	**0.94**	0.94	**0.84**	1.98	**0.89**	1.64	**0.92**	**1.37**	**0.91**	1.48
Trunk FFM [kg]	Male	**0.90**	0.90	**0.39**	4.26	**0.80**	2.45	**0.87**	**1.97**	**0.87**	1.97
	Female	**0.88**	0.88	**0.46**	2.91	**0.83**	1.64	**0.84**	**1.63**	**0.87**	1.43
Arms Fat Mass [kg]	Male	**0.80**	0.80	**0.70**	0.74	**0.74**	0.70	**0.89**	**0.45**	**0.81**	0.59
	Female	**0.88**	0.88	**0.76**	0.69	**0.81**	0.61	**0.83**	**0.58**	**0.84**	0.57
Arms FFM [kg]	Male	**0.88**	0.88	−0.02	1.80	**0.31**	1.48	**0.87**	**0.65**	**0.71**	0.96
	Female	**0.80**	0.80	**0.47**	0.76	**0.64**	0.63	**0.71**	**0.57**	**0.66**	0.61
Legs Fat Mass [kg]	Male	**0.78**	0.78	**0.63**	2.53	**0.62**	2.56	**0.91**	**1.26**	**0.75**	2.07
	Female	**0.91**	0.91	**0.66**	2.02	**0.67**	2.00	**0.90**	**1.13**	**0.85**	1.32
Legs FFM [kg]	Male	**0.90**	0.90	**0.26**	3.48	**0.67**	2.34	**0.87**	**1.44**	**0.84**	1.61
	Female	**0.88**	0.88	**0.59**	2.05	**0.80**	1.45	**0.87**	**1.19**	**0.85**	1.26

For % fat, we included two methods of prediction, % fat = predicted Fat Mass / scale weight, and the linear regression of % fat from PCA vectors from (4) below in parentheses.

TABLE IV

COMPARISON AGAINST NG ET AL.

Output Variable	Gender	Ng et al, 3D PC Only; Stepwise Regression 5-fold CV		This work; prediction on 2D image, reported on test set scans	
		R^2	RMSE	R^2	RMSE
Fat Mass [kg]	Male	**0.88**	3.38	**0.90**	3.63
	Female	**0.93**	2.96	**0.94**	2.29
FFM [kg]	Male	**0.93**	3.38	**0.89**	3.63
	Female	**0.90**	2.95	**0.92**	2.29
% Fat	Male	**0.65**	3.83	**0.725 (0.806)**	3.90 (3.27)
	Female	**0.70**	4.10	**0.74 (0.631)**	3.29 (3.94)
FMI [kg/m^2]	Male	**0.87**	1.11	**0.90**	1.19
	Female	**0.93**	1.13	**0.94**	0.85
FFMI [kg/m^2]	Male	**0.90**	1.11	**0.81**	1.19
	Female	**0.88**	1.12	**0.91**	0.85
Visceral Fat Mass [kg]	Male	**0.67**	0.16	**0.66**	0.15
	Female	**0.75**	0.14	**0.36**	0.17
Trunk Fat Mass [kg]	Male	**0.91**	1.68	**0.92**	1.76
	Female	**0.94**	1.43	**0.91**	1.48
Trunk FFM [kg]	Male	**0.90**	1.94	**0.87**	1.97
	Female	**0.87**	1.72	**0.87**	1.43
Arms Fat Mass [kg]	Male	**0.84**	0.26	**0.81**	0.59
	Female	**0.70**	0.58	**0.84**	0.57
Arms FFM [kg]	Male	**0.76**	0.52	**0.71**	0.96
	Female	**0.67**	0.33	**0.66**	0.61
Legs Fat Mass [kg]	Male	**0.71**	0.87	**0.75**	2.07
	Female	**0.83**	0.86	**0.85**	1.32
Legs FFM [kg]	Male	**0.89**	0.76	**0.84**	1.61
	Female	**0.83**	0.71	**0.85**	1.26

Comparison of our work with Ng *et al.* which only reports % fat as predicted Fat Mass / scale weight.

TABLE V

P-VALUES OF PAIRED T-TESTS

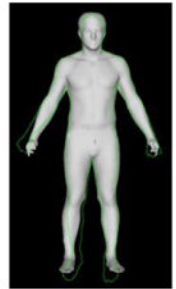| Output Variable	Gender	DXA1 vs DXA2	T1 vs T2	T1 vs DXA1
Fat Mass [kg]	Male	0.21	0.50	0.35
	Female	0.58	0.30	0.45
FFM [kg]	Male	0.30	0.50	0.35
	Female	0.68	0.30	0.45
% Fat	Male	0.17	0.71	0.46
	Female	0.44	0.78	0.50
FMI [kg/m^2]	Male	0.18	0.40	0.35
	Female	0.52	0.28	0.53
FFMI [kg/m^2]	Male	0.30	0.40	0.35
	Female	0.60	0.28	0.53
Visceral Fat Mass [kg]	Male	0.10	0.51	0.20
	Female	0.13	0.13	0.03
Trunk Fat Mass [kg]	Male	0.74	0.56	0.76
	Female	0.24	0.11	0.62
Trunk FFM [kg]	Male	0.76	0.82	0.01
	Female	0.10	0.19	0.03
Arms Fat Mass [kg]	Male	0.30	0.59	0.02
	Female	0.23	0.66	0.50
Arms FFM [kg]	Male	0.86	0.54	0.07
	Female	0.21	0.22	0.87
Legs Fat Mass [kg]	Male	0.06	0.49	0.18
	Female	0.43	0.10	0.35
Legs FFM [kg]	Male	0.02	0.38	0.92
	Female	0.48	0.22	0.31
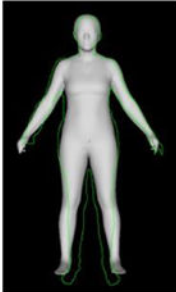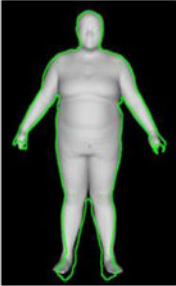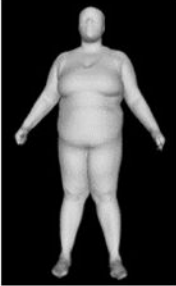
p-values for paired t-tests. $p < 0.004$ was used to test for statistically significant differences.

DXA1 and DXA2 are the two DXA measurements, T1 and T2 are the two trials of our method on separate sets of photographs.

T1 was the test set that we treated as the result trial in Table III, IV, and VI.

TABLE VI

VISUALIZED RESULTS

Results viewed under camera projection π. Columns in order show: a) The camera image input b) the seed shape defined by the known height and weight c) the seed shape optimized for the rigid transformation to align best to the joint positions d) the final optimized shape deformation and transformation e) the ground truth scan. Note that participants are not scanned in the exact same position they were photographed in. f) Predicted and ground truth % fat values from the direct regression method, picked for consistency.