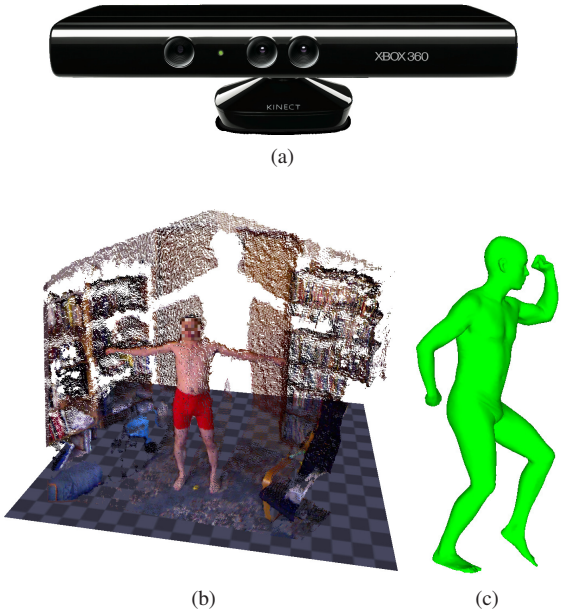
Home 3D Body Scans from Noisy Image and Range Data

Alexander Weiss* David Hirshberg† Michael J. Black†
*Dept. of Computer Science, Brown University, Providence, RI, USA
†Perceiving Systems Dept., Max Planck Institute for Intelligent Systems, Tübingen, Germany
aweiss@cs.brown.edu david.hirshberg@is.mpg.de black@tuebingen.mpg.de

Abstract

The 3D shape of the human body is useful for applications in fitness, games and apparel. Accurate body scanners, however, are expensive, limiting the availability of 3D body models. We present a method for human shape reconstruction from noisy monocular image and range data using a single inexpensive commodity sensor. The approach combines low-resolution image silhouettes with coarse range data to estimate a parametric model of the body. Accurate 3D shape estimates are obtained by combining multiple monocular views of a person moving in front of the sensor. To cope with varying body pose, we use a SCAPE body model which factors 3D body shape and pose variations. This enables the estimation of a single consistent shape while allowing pose to vary. Additionally, we describe a novel method to minimize the distance between the projected 3D body contour and the image silhouette that uses analytic derivatives of the objective function. We propose a simple method to estimate standard body measurements from the recovered SCAPE model and show that the accuracy of our method is competitive with commercial body scanning systems costing orders of magnitude more.



(a) (b) (c)

Figure 1: **Overview**. (1a) Microsoft Kinect [1]. (1b) 3D point cloud of a human in a cluttered home environment. (1c) Recovered shape transformed into a new pose.

1. Introduction

For many applications an accurate 3D model of the human body is needed. The standard approach involves scanning the body using a commercial system such as a laser range scanner or special-purpose structured-light system. Several such body scanners exist, costing anywhere from $35,000 to $500,000. The size and cost of such scanners limit the applications for 3D body models. Many computer vision solutions suffer the same problems and require calibrated multi-camera capture systems. Here we describe a solution that produces accurate body scans using consumer hardware that can work in a person's living room (Fig. 1). This opens the door to a wide range of new applications.

Recently there have been several approaches to capturing 3D body shape from a monocular image [15, 16, 19, 26], a small number of synchronized camera images [5], or from several unsynchronized cameras [17]. We restrict our attention to the monocular case, where the common approach is to segment the person from the background and to estimate the 3D shape of the body such that the silhouette of the body matches the image silhouette. The wide variation in body shape, the articulated nature of the body, and self occlusions in a single view, however, all limit the usefulness of image silhouettes alone. To cope with these issues we combine image silhouettes with coarse monocular range data captured with a single Microsoft Kinect sensor [1].

The resolution and accuracy of the sensor is relatively poor and our key contribution is a method to accurately estimate human body pose and shape from a set of monocular low resolution images with aligned but noisy depth information. To be scanned, a person moves in front of a

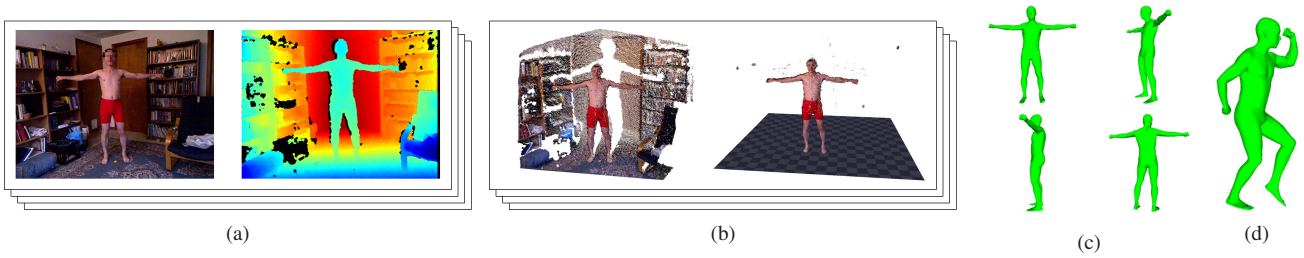


Figure 2: **Overview**. (2a) Four views of the body in different poses are captured from a single Kinect. (2b) 3D point cloud and segmented 3D point cloud with ground plane for four frames (one shown). (2c) Recovered pose and shape (4 frames). (2d) Recovered shape in new pose.

single sensor to capture a sequence of monocular images and depth maps that show the body from multiple angles (Fig. 2). As the person moves, their body shape changes making rigid 3D alignment impossible. We solve for the pose in each frame and for a single common shape across all frames. To do so, we use the SCAPE model [4] which is a parametric 3D model that factors the complex non-rigid deformations induced by both pose and shape variation and is learned from a database of several thousand laser scans.

We estimate model parameters in a generative framework using an objective function that combines a silhouette overlap term, the difference between the observed range data and the depth predicted by our model, and an optional pose prior that favors similarity of poses between frames. The silhouette term uses a novel symmetric shape dissimilarity function that we locally minimize using a standard quasi-Newton method. Our silhouette formulation has significant advantages over previous methods (such as ICP) and enables accurate optimization of body shape and pose in a very high-dimensional space.

In summary our contributions are: 1) A system for at home body scanning; 2) The combination of multiple low-resolution, noisy, monocular views (range and/or silhouettes) to estimate a consistent 3D body shape with varying pose; 3) A new method for matching 3D models to silhouettes using an objective function that is correspondence-free, bidirectional, and can be optimized with standard methods requiring derivatives of the objective function; 4) A simple method to predict 3D body measurements from SCAPE model parameters using linear regression; 5) A quantitative comparison with a commercial state-of-the-art solution for scanning and measuring the body.

2. Related Work

The Microsoft Kinect provides one of the first inexpensive and widely available range sensors. Existing commercial and research systems solve the problem of rough body pose estimation from this type of sensor data [12, 23] but, to date, there are no methods for accurate body shape estimation. To estimate body shape accurately, we must deal with data that is monocular, low resolution, and noisy.

Anguelov *et al*. [4] describe a partial solution. They show that, given a high resolution range image from a single view, the SCAPE model can be fit to the observed data. The observed data constrains the full 3D shape, enabling them to hallucinate unseen parts of the body (shape completion). For our purposes, this is not sufficient since we seek an accurate model of the full body shape. We must therefore combine multiple views of the person and several low-resolution scans to obtain an accurate representation.

If the person remained rigid, or we used multiple sensors, then it would be straightforward to rigidly align multiple views to recover a composite 3D body. In our case, people move relative to a single sensor and, even if they try to maintain the same pose, there will be non-rigid variations in their shape. To cope with this we need to integrate the consistent 3D body shape information across views and poses. To do so we use the SCAPE model [4] which factors body shape and pose information.

Balan and Black [6] use a similar idea to solve for body shape under clothing. They capture dressed people in multiple poses with a multi-camera system and a green-screen background. Like us, they assume that body shape is consistent across pose variations and combine information from multiple poses to estimate body shape. Our work is different in that we use monocular data. Every time instant captures the person in a different pose, so we never see the same pose from multiple views.

There have been several recent methods that estimate body shape from monocular images. For example, image contours have been used in several graphics applications [19, 26] where the metric accuracy of the body shape is not important. The recovered shapes are used for animation purposes to change the shape of people in images or videos. To achieve metrically accurate results, more information is required. Guan *et al*. [15] show that silhouettes alone are not sufficient for this task and introduce two innovations. First they provide a height- and weight-constrained subspace of

body shape variation to constrain the problem. Second, and more importantly, they integrate a shape from shading cue into the body shape optimization (similar to [9]). The shading cue gives information about the shape inside the body contour and they show that adding this improves the recovered shape accuracy.

Shading is a relatively weak cue and if range data is available, it can provide more information about shape. In early work on body shape estimation, Plänckers and Fua [22] use depth information from a stereo camera to estimate rough body shape in a frontal view. Grest *et al*. [14] fit parameters of a simplified body model to silhouettes and then use these parameters to improve pose tracking from range data.

3. Sensor and preprocessing

The Microsoft Kinect sensor that we use consists of an IR camera, an RGB camera, and an IR projector that casts a fixed speckle pattern. Conversion of the pattern, as seen by the IR camera, to a depth map happens on the device. It has a USB interface and images can be captured using a library developed by the OpenKinect project [2]. This library provides access to both the depth map and the raw IR video, as well as to the RGB video and data from a built in accelerometer. The video streams are VGA resolution and both the RGB and IR (either raw or the depth map) can be captured synchronized at 30 fps. [1]

Intrinsic calibration. Intrinsic calibration of the RGB camera is carried out with a checkerboard and standard calibration techniques [7]. To calibrate the IR camera we cover the projector so that the calibration grid is not corrupted by the projected pattern; otherwise calibration is identical to that of the RGB camera. We correct for a known offset between the raw IR image and the depth map; see [21].

Stereo calibration. Stereo calibration between the depth and RGB cameras can be achieved with standard stereo calibration methods [7]. We use this only for visualization to map the color image onto the point cloud.

Depth calibration. The Kinect reports depth discretized into 2047 levels, with a final value reserved to mark pixels for which no depth can be calculated. These discrete levels are not uniformly distributed, but are much denser close to the device. We calibrate the depth by lining up a planar target parallel to the Kinect such that the depth values are as uniform as possible across its surface; the distance is then measured and the process repeated with depths ranging from 0.5m to 3m in 0.1m increments. A curve of the form: $d(x) = \frac{1}{ax+b}$ is fit to this data, yielding the distance $d(x)$ in meters given the discrete depth level x. The resulting depth maps can be visualized either as a range image or as a point cloud, see Fig. 2.

Ground plane. We obtain an estimate of the ground plane by robustly fitting a plane to the bottom of the point cloud, using the Kinect's on board accelerometer to initialize such that we locate the floor and not one of the walls.

Segmentation. We segment the body from the surrounding environment using background subtraction on the depth map. Given a depth map D_{bg} taken without the subject present and a depth map D_f associated with a frame f, we take the foreground to be $D_{bg} - D_f > \epsilon$, where ϵ is a few mm. We then apply a morphological opening operation to remove small isolated false positives.

4. Body model and fitting

In order to estimate a body shape that is invariant to pose, we need a model that accurately represents non-rigid shape deformations while factoring deformations caused by changes in intrinsic shape (height, weight, body type, etc.) from deformations caused by changes in pose. We use a SCAPE [4] model with 15 body parts connected in a kinematic tree with 3 degrees of freedom between each part. Non-rigid deformations due to pose variation are modeled using linear predictors learned from examples. Body shape deformation is modeled using principal component analysis (PCA) on an aligned database of several thousand bodies. We use the method described in [15] to constrain body shape to a subspace that is roughly orthogonal to height variation, allowing us to freely optimize within the subspace of bodies with the subject's reported height. Our model has 48 pose parameters per frame and 60 shape parameters (i.e. 252 parameters for 4 frames).

4.1. Pose initialization

We assume a gross initial pose estimate; a complete, end to end system would be obtained by combining the method we describe here with an existing coarse pose tracking algorithm [12, 23]. The subject provides their height and the initial body shape is taken to be the average shape for the subject's height and gender [15]. We initialize the body model in the scene using the ground plane and the centroid of the point cloud. Examples of initializations for two trials can be seen in Fig. 4.

4.2. Depth objective

For a body model represented as a triangulated 3D mesh with pose and shape parameters θ, we associate a triangle $t_x(\theta)$ with every pixel x in the overlap between the model silhouette $S(\theta)$ and observed silhouette T by finding the front most triangle that projects into x. Let $U(\theta) = \{(x_1, t_{x_1}(\theta)), \ldots\}$ for all x in $S(\theta) \cap T$. For each pixel we have the observed depth $\breve{D}_x$, and for the corresponding triangle t we find the depth, $D_{x,t}(\theta)$, along a ray through the

[1] As of this writing, software tools for working with the Kinect are evolving rapidly. In addition to the OpenKinect libraries we use, options now include OpenNI and Microsoft's Kinect SDK, both of which provide additional functionality.

pixel center to the plane of the triangle. Taking ρ to be a robust error function (here, Geman-McClure [13]), our depth objective is

$$E_d(\theta; U) = \frac{1}{|U|} \sum_{(x,t)\in U} \rho\left(D_{x,t}(\theta) - \breve{D}_x\right).$$

4.3. Silhouette objective

Methods for fitting 3D models to silhouettes usually approximate one of these two integrals

$$\int_{\vec{x}\in S} \min_{\vec{y}\in T} \rho(||\vec{x} - \vec{y}||) \tag{1}$$

$$\int_{\vec{x}\in \partial S} \min_{\vec{y}\in \partial T} \rho(||\vec{x} - \vec{y}||). \tag{2}$$

Here S and T are silhouettes, ∂S and ∂T are their boundaries, and ρ is a non-decreasing function (*e.g.* Geman-McClure [13]). Frequently, approximations to (1) use a discrete distance map [5, 24] and approximations to (2) use a discrete distance map or a correspondence-based scheme like ICP [10, 17]. The integrand of the latter is illustrated in Fig. 3. Integrals like these are often used to define shape distances [8], but are not widely used with parametric 3D models under projection.

Accurately fitting a body to the image evidence benefits from bi-directional shape distance functions [24] that compute the distance from the model to the image contour and vice versa. Minimizing the distance from the image to the model ensures that all image measurements are explained while minimizing the distance from the model to the image ensures that visible body parts are entirely explained by image evidence. Modeling the distance from the model to the image is straightforward using the Euclidean distance transform to approximate the distance function to the image silhouette, as this does not change during optimization. Modeling the distance from image to the model is more difficult because the distance function to the model's silhouette changes with the parameters being optimized; this makes an explicit computation of the derivatives difficult.

Consequently, many methods that use distance maps either use uni-directional distance, from model silhouette to static observed silhouette [20, 24] or use a derivative-free optimizer [5]. Problems with the uni-directional application of (1) have been discussed and addressed [24]. Similar problems arise with the use of (2) but are not often mentioned. The use of derivative free methods for a high-dimensional problem like ours is impractical, so we seek a method admitting explicit computation of the derivative.

ICP methods are frequently used to minimize (2) for 2D to 2D and 3D to 3D shape registration problems. They can be used bidirectionally and optimization is straightforward because the average point-to-shape distance is bounded by

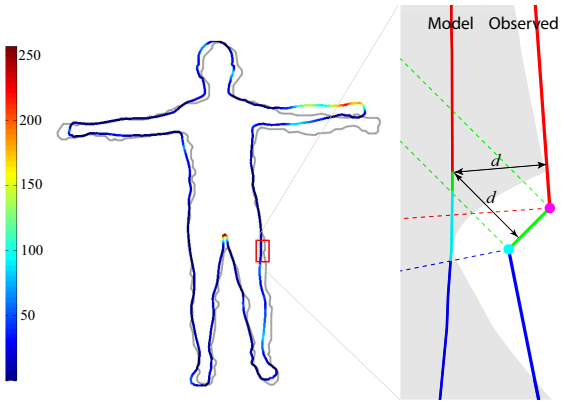


Figure 3: **Silhouette distance**. On the left, the silhouette of the body model is colored by squared distance to the grey observed silhouette. On the right, the implicit point and line correspondence on an arc of the left leg's silhouette is shown by coloring the arc to match the colors of points and lines on the observed silhouette. The squared distance function along this arc as a function of the y-coordinate is overlaid in grey to illustrate the effects of changes in correspondence. Colored dashed lines are used to indicate the boundary of the region where a segment's point-line distance applies.

the average distance between corresponding points, which is a smooth function of the vertices of both shapes. Under projection we lose this bound because points on the silhouette boundary no longer have a stable relationship to the 3D geometry. Without this, the use of ICP is problematic, especially with complex articulated and non-rigid objects.

If we have a set of correspondences between 3D model vertices on the silhouette boundary and points on the observed silhouette, as we minimize the average distance of the projected vertices to their corresponding 2D points, some vertices will disappear from the silhouette boundary and new vertices will appear. Since these newly visible vertices will not influence the objective function until we recompute correspondences, the optimizer may move them anywhere without penalty. When this happens, the parameters being optimized may jump away from low-error fixed points to a solution from which ICP cannot recover.

We address this problem with a well-behaved new formulation that uses implicit rather than explicit correspondences. We compute the line integral in (2) directly, replacing the explicit correspondences of ICP with the continuously changing ones implied by the min function. Symmetrizing this yields an objective function that is correspondence-free and bidirectional.

To compute this integral, we must know, for each point on the integration silhouette, the distance to the nearest point on the other (reference) silhouette. Each segment of the integration silhouette is broken up into pieces that are

nearest to the same geometric primitive (vertex or line segment interior) in the reference silhouette. These breaks, illustrated in Fig. 3, occur in two circumstances: 1) Along lines emanating from a segment's vertices and perpendicular to the segment. These lines define the region where perpendicular distance to the segment is defined (dashed lines in Fig. 3). 2) On linear or quadratic arcs where two points (quadratic), two segment interiors (linear), or a segment interior and a point (quadratic) are equidistant (arrows of equal distance d in Fig. 3).

The derivative of this integral is easily computed in terms of the derivative of the path of integration and the derivative of the integrand [11]. There is, however, a small problem. At the breaks the integrand is not differentiable with respect to the reference silhouette, as the the distance functions to the two equidistant primitives vary independently. Nor is it differentiable with respect to the point of evaluation x, as variation in one direction is dictated by one primitive's distance function and variation in another will be dictated by the other's. If these breaks occur only at points, as they do for almost every pair of silhouettes, they do not matter. There are finitely many such breaks, and the value of the integrand at finitely many points, so long as it is bounded, does not effect the value of an integral. But, if a segment on the integration silhouette lies along one of the arcs where two primitives are equidistant, the non-differentiability of the integrand is inherited by the integral. Because this happens only when two constraints are met – the integration path and arc of equidistance must be parallel and touching – manifolds where our objective function is non-smooth have dimension 2 less than the parameter space. There is nothing about these constraints that would push the optimization trajectory toward these manifolds. In practice we optimize using a method intended for smooth functions and do not encounter problems.

De la Gorce *et al*. [9] use a similar integration-based approach in the context of articulated hand tracking with a generative model and formulate a differentiable objective function. Their objective focuses on a generative model of image appearance across the interior of the object. They compute a 2D integral, which allows them differentiability despite a 1D discontinuity along the occluding contour of the body. We could similarly compute a differentiable version of the area integral in (1), but it would require us to compute $\arg\min_{\vec{y}\in T} ||\vec{x}-\vec{y}||$ inside a 2D region, which amounts, in our setting, to computing the Voronoi diagram for a set of line segments.

Our silhouette objective function is a symmetrized and scaled version of (2), averaging distance over each silhouette boundary to the other:

$$E_{uni}(A,B) = \frac{1}{2\,|\partial A|}\int_{\vec{x}\in\partial A}\min_{\vec{y}\in\partial B}\rho(||\vec{x}-\vec{y}||) \quad (3)$$

$$E_s(S(\theta),T) = E_{uni}(S(\theta),T)+E_{uni}(T,S(\theta)) \quad (4)$$

where $S(\theta)$ is the silhouette of the model with parameters θ and T is the image silhouette.

4.4. Optimization

To estimate θ, we alternately compute pixel-triangle correspondences $U_f(\theta_i)$ for every frame f and new model parameters θ_{i+1} by local minimization of $E_i(\theta;U_f(\theta_i)) = \sum_f E_d(\theta;U_f(\theta_i)) + \alpha\sum_f E_s(S_f(\theta),T_f) + \beta E_{\text{pose}}(\theta)$, where $E_{\text{pose}}(\theta)$ is a simple pose prior. For local minimization, we use a SR1 trust region method with exact solution of the trust region subproblem.

5. Results

We scanned four subjects, having each stand in a T pose four times: facing the camera, in profile, facing away from the camera, and rotated $45°$, halfway between frontal and profile. As demonstrated in Fig. 5, the choice of the four poses is relatively arbitrary; we found that more than four poses did not significantly improve the results and fewer made them worse.

Fitting results for two subjects are shown in Fig. 4. It is important to remember that these images are not multi-camera synchronous captures. Because these images are not captured simultaneously, and the subjects move from frame to frame, the pose cannot be assumed constant between frames. Consequently we let pose vary between frames and use a simple pose prior that penalizes frame-to-frame variation in the orientation of each body part independently. This helps keep the pose reasonable in cases like the third frame (profile view) for the female subject, where the right leg is not visible from the camera and is thus otherwise unconstrained. The foot pose of the female subject shown here is problematic, with portions of the feet incorrectly segmented as background and a large region of floor nearby incorrectly segmented as foreground inducing incorrect ankle rotation. Despite that, the fit to the remainder of the body is quite good. With the coarse range and silhouette data used here, any individual view may not be very accurate, but the robust combination of body shape across views provides sufficient constraints to recover shape well.

Figure 5 shows a subject scanned in several widely varying poses and fit without the pose constancy prior to highlight the ability of the method to integrate shape across multiple disparate poses. The pose error in the second frame, where the lower legs are pulled tightly up to the upper legs, is due to a segmentation error; the lower legs were incorrectly segmented as background, so there was no image evidence to drive the lower legs to remain visible.

Optimization takes approximately 65 minutes per body. This may seem excessive but recall that the optimization involves estimating 252 parameters.

From bodies to measurements. One of the reasons to fit a 3D body model is to extract standard measurements of the

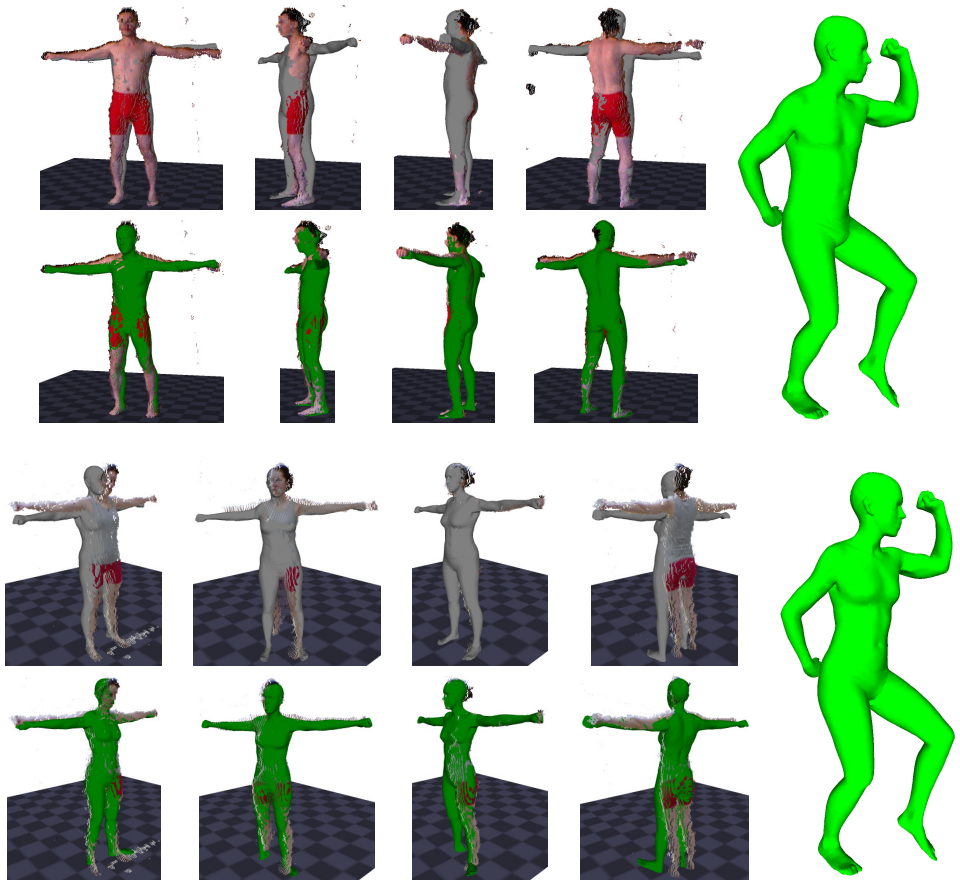


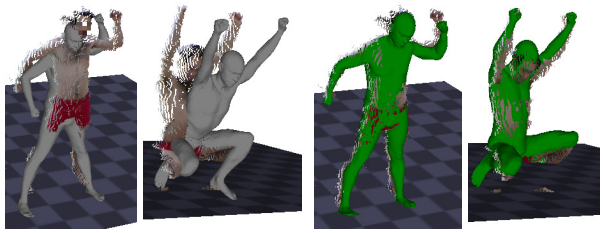
Figure 4: **Results**. Rows 1-2 male subject. Rows 3-4 female subject. Grey mesh is initialization. Green mesh is fitted result. Righthand column is fitted result reposed into novel pose.

body (arm length, chest circumference, etc.) that are useful in many applications. To calculate measurements from shape parameters, we use a method that follows Allen *et al*. [3] in spirit, but is the inverse of the problem they describe. Allen *et al*. learn a linear function from a set of measurements to shape parameters, allowing them to synthesize new people with specified measurements. We take the same data–shape parameters and hand measurements for the several thousand subjects of the CAESAR dataset–and perform linear regression to learn a function from shape parameters to measurements (with the exception of weight, where we find it more accurate to regress from the shape parameters to the cube root of weight).

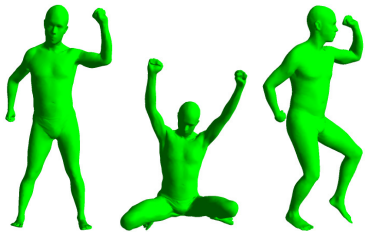
Accuracy relative to laser scans. We evaluate the metric accuracy of fitting body shape using just image contours and using both image contours and depth. To do so we captured reference scans of the subjects using a Vitus laser scanner (Human Solutions GmbH, Germany) (Fig. 6a). To test the accuracy of using the Kinect sensor versus a commercial laser scanner, we first fit the SCAPE model to the laser scans using a standard ICP method (Fig. 6b); we also fit to Kinect data as described above (Fig. 6c). This allows us to evaluate the accuracy of the fitting method and sensor data independent of the smoothing effect introduced by the SCAPE model which represents body shapes in a low-dimensional linear subspace. The SCAPE fit to the laser scan represents a "best case scenario" since the data is high resolution and highly accurate. The difference between a model fit to laser data and Kinect data is illustrated in Fig. 6d; the vertex to vertex distances are 0.53mm (minimum), 22.23mm (maximum), 10.17 (mean), 9.91 (median).

Linear measurement accuracy. The second source of ground truth we use to evaluate accuracy is hand measurements, taken by a professional with both tailoring and anthropometric measurement experience. These we compare to measurements calculated from the optimized shape parameters using the linear predictors described above.

Figure 7 compares the measurement accuracy from SCAPE bodies fit to silhouettes alone, silhouettes and range, and laser data. We find that range and silhouettes together are more accurate than silhouettes alone. The measurement accuracy using the Kinect-based fits is only

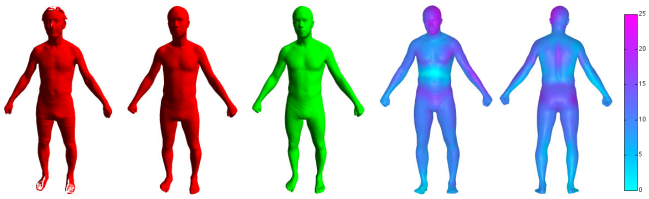


(a) (b)



(c) (d)

Figure 5: **Widely varying poses**. (5a) Initialization. (5b), (5c) Result. (5d) Result reposed into novel pose.



(a) (b) (c) (d)

Figure 6: **Comparison to laser scan**. (6a) Laser scan. (6b) SCAPE model fit to laser scan; pose and shape recovered. (6c) Contour + depth fit to 4 views, reposed to match pose of laser scan of same subject. (6d) Difference map showing areas of similarity (blue) and difference (purple) between 6b and 6c (scale in mm).

slightly worse than with the high-resolution full-body laser scans; median errors generally are within 1cm of the laser scan measurements.

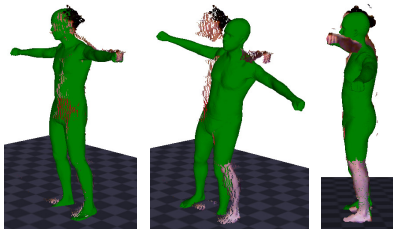
Additionally, we compare our accuracy with that of a commercially available laser scan measurement system, Human Solutions Anthroscan (Human Solutions GmbH, Germany). This system works on the raw laser scan and, consequently, factors out the effect of the SCAPE model. It is interesting to note that our inexpensive system is competitive and even outperforms the commercial system on all the circumference measurements.

6. Conclusions

Three-dimensional body scanning has so far had a limited range of applications due to the expense, complexity, and space requirements of existing scanning systems. All these systems are based on multiple calibrated cameras and structured light sources (including lasers). Here we show that we can achieve similar accuracy with a single inexpensive commodity sensor. We have demonstrated the feasibility of a body scanner that could work in a person's living room by combining information about body shape over several noisy frames. The key idea is to use the shape constancy of the body across frames to accurately estimate a single shape and varying pose. The approach combines silhouettes and depth with a novel silhouette dissimilarity term that overcomes problems of previous approaches. We show that measurements of the body can be reliably predicted using a simple linear regression approach and compare favorably to expensive commercial systems.

Future work should address the estimation of shape under clothing. This has been demonstrated in a synchronized multi-camera capture scenario with silhouettes [6] and with laser scans [18]. We believe that it should work with the Kinect sensor. We would also like to improve the optimization speed to make it interactive. An interactive system could provide the user with feedback about how to move to improve their body model.

Acknowledgments. We thank Loretta Reiss for her measurement expertise and Lisa Wang for mathematical discussions. This work was supported in part by NIH EUREKA award 1R01NS066311–01 and NSF award IIS–0812364.



(a) (b) (c)

Figure 8: **With and without** E_d. (8a) Fit with contour and depth terms. (8b) Fit with only contour term. (8c) Same as (8b) but seen from the camera, showing the quality of the contour match, despite the pose being wildly wrong.

References

[1] Microsoft Corp. http://www.xbox.com/kinect.

[2] OpenKinect project. http://openkinect.org.

[3] B. Allen, B. Curless, and Z. Popovic. The space of human body shapes: Reconstruction and parameterization from range scans. *ACM Trans. Graph.*, 22(3):587–594, 2003.

[4] D. Anguelov, P. Srinivasan, D. Koller, S. Thrun, J. Rodgers, and J. Davis. SCAPE: Shape completion and animation of people. *ACM Trans. Graph.*, 24(3):408–416, 2005.

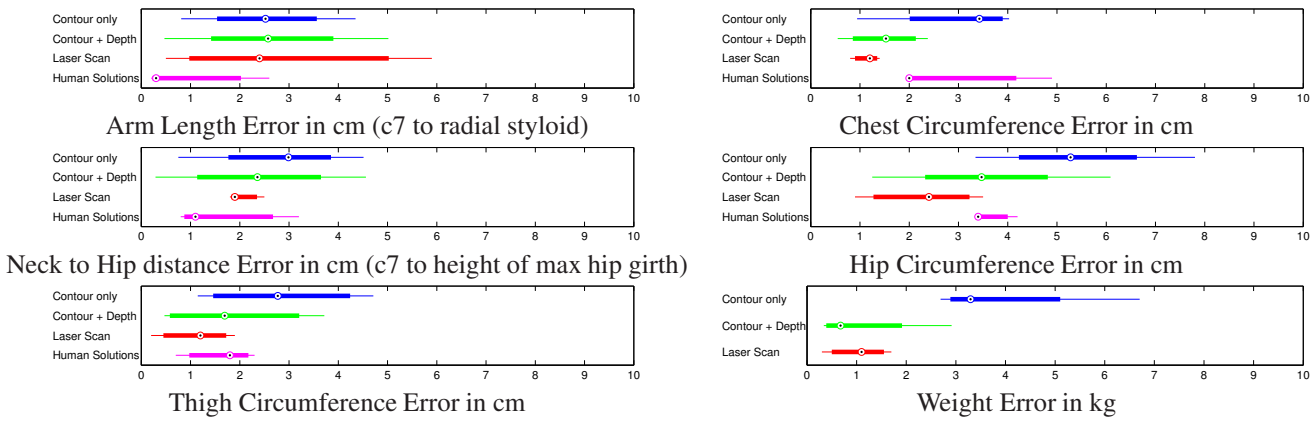


Figure 7: **Measurement accuracy**. Error of measurements found by regression from fitted shape parameters using contour cost only (blue, 4 subjects), using contour and depth costs (green, 4 subjects), and of SCAPE fit to laser scan (red, 3 subjects), with respect to ground truth obtained via hand measurement. For comparison, we also show measurement error between hand measurement and measurements calculated from the laser scans by a commercial scan measurement system (Human Solutions Anthroscan) (magenta, 3 subjects).

[5] A. Balan, L. Sigal, M. J. Black, J. Davis, and H. Haussecker. Detailed human shape and pose from images. *CVPR*, pp. 1–8, 2007.

[6] A. O. Balan and M. J. Black. The naked truth: Estimating body shape under clothing. *ECCV*, pp. 15–29, 2008.

[7] J.-Y. Bouguet. Camera Calibration Toolbox for Matlab. http://www.vision.caltech.edu/bouguetj/calib_doc.

[8] G. Charpiat, O. Faugeras, and R. Keriven. Approximations of shape metrics and application to shape warping and empirical shape statistics. *Foundations of Computational Mathematics*, 5(1):1–58, 2005.

[9] M. de La Gorce, N. Paragios, and D. Fleet. Model-based hand tracking with texture, shading and self-occlusions. *CVPR*, pp. 1–8, 2008.

[10] Q. Delamarre and O. Faugeras. 3D articulated models and multi-view tracking with silhouettes. *ICCV*, pp. 716–721, 1999.

[11] H. Flanders. Differentiation under the integral sign. *American Mathematical Monthly*, 80(6):615–627, 1973.

[12] V. Ganapathi, C. Plagemann, D. Koller, and S. Thrun. Real time motion capture using a single time-of-flight camera. *CVPR*, pp. 755–762, 2010.

[13] S. Geman and D. McClure. Statistical methods for tomographic image reconstruction. *Bulletin Int. Statistical Institute*, LII(4):5–21, 1987.

[14] D. Grest, D. Herzog, and R. Koch. Human model fitting from monocular posture images. *Proc. Vision, Modeling, Visualization*, 2005.

[15] P. Guan, A. Weiss, A. O. Balan, and M. J. Black. Estimating human shape and pose from a single image. *ICCV*, pp. 1381–1388, 2009.

[16] N. Hasler, H. Ackermann, B. Rosenhahn, T. Thormählen, and H.-P. Seidel. Multilinear pose and body shape estimation of dressed subjects from image sets. *CVPR*, pp. 1823–1830, 2010.

[17] N. Hasler, B. Rosenhahn, T. Thormählen, M. Wand, J. Gall, and H.-P. Seidel. Markerless motion capture with unsynchronized moving cameras. *CVPR*, pp. 224–231, 2009.

[18] N. Hasler, C. Stoll, B. Rosenhahn, T. Thormählen, H.-P. Seidel. Estimating body shape of dressed humans. *Computers & Graphics*, 33:211216, 2009.

[19] A. Jain, T. Thormählen, H.-P. Seidel, and C. Theobalt. Moviereshape: Tracking and reshaping of humans in videos. *ACM Trans. Graph.*, 29(5), 2010.

[20] D. Knossow, R. Ronfard, and R. Horaud. Human motion tracking with a kinematic parameterization of extremal contours. *IJCV*, 79(3):247–269, 2008.

[21] K. Konolige and P. Mihelich. ROS.org wiki: kinect_calibration/technical. [Online](Accessed 2/2011) Available: http://www.ros.org/wiki/kinect_calibration/technical.

[22] R. Plänkers and P. Fua. Model-based silhouette extraction for accurate people tracking. *ECCV*, pp. 325–339, 2002.

[23] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake. Real-time human pose recognition in parts from single depth images. *CVPR*, 2011.

[24] C. Sminchisescu and A. Telea. Human pose estimation from silhouettes. a consistent approach using distance level sets. *WSCG Int. Conf. Computer Graphics, Visualization and Computer Vision*, 2002.

[25] Z. Zhang. Iterative point matching for registration of free-form curves and surfaces. *IJCV*, 13(2):119–152, 1994.

[26] S. Zhou, H. Fu, L. Liu, D. Cohen-Or, and X. Han. Parametric reshaping of human bodies in images. *ACM Trans. Graph.*, 29(4), 2010.