Estimating Anthropometry and Pose fr om a Single Image

Carlos Barron and Ioannis A. Kakadiaris
Department of Computer Science
University of Houston
4800 Calhoun, Houston, TX 77204-3475
{cbarron, ioannisk}@uh.edu

Abstract

In this paper, we present a four-step technique for simultaneously estimating a human's anthropometric measurements (up to a scale parameter) and pose from a single image. The user initially selects a set of image points that constitute the projection of selected landmarks. Using this information, along with a priori statistical information about the human body, a set of plausible segment length estimates are generated. The third step produces a set of plausible poses based on joint limit constraints using a geometric method. In the fourth step, pose and anthropometric measurements are obtained by minimizing an appropriate cost function subject to the associated constraints. The novelty of our approach is the use of anthropometric statistics to constrain the estimation process that allows the simultaneous estimation of both anthropometry and pose. We demonstrate the accuracy, advantages and limitations of our method for various classes of both synthetic and real input data.

1. Introduction

Video-based three-dimensional human motion tracking is an important and challenging research problem. Its importance stems from numerous applications such as: 1) performance measurement for human factors engineering, 2) posture and gait analysis for training athletes and physically challenged persons, 3) human body, hands and face animation, and 4) automatic annotation of human activities in video databases. The challenges towards the general applicability of a vision-based 3-D tracking system on real data include the following:

- *Data from one camera only:* There are several applications for which the video recordings from only one view are available (e.g., for analyzing the motion of astronauts during extravehicular activities in previous missions). In addition, the camera might be moving, possibly zooming in and out.
- *Model-Acquisition:* There is no such thing such as "average" human and that makes the selection of a geometric model for model-based tracking difficult.
- *Modeling:* The human models that are currently used for motion estimation do not incorporate statistical anthropometric information.

Our long term goal is to develop a model-based system for tracking humans from monocular images. In this paper, we present a technique for simultaneous anthropometry and pose estimation from the first frame of an image sequence. The input to the algorithm are the image coordinates of the visible landmarks from the human subject (as selected by the user) in the image under examination (Figure 6(a)). The output is the subject's anthropometric measurements (up to a scale parameter) and his/her pose in the specific image (Figure 6(c)). The novelty of our approach is the use of anthropometric statistics to constrain the estimation process. The impact of our method lies in the ability to semi-automate the initialization phase for model-based human tracking methods from a single camera. As it will explained in later sections, our method can handle images like the one depicted in Figure 1(a), but not images like the one depicted in Figure 1(b).

The remainder of this paper describes our technique in more detail. In Section 2 we review previous work in the area, and in Section 3 we formulate the problem. In Section 4 we describe our method in detail, while in Section 5 we illustrate results from our system.

2. Previous W ork

Two of the challenges in model-based human tracking algorithms are: 1) the acquisition of an accurate human body model that will be employed as the model, and 2) the initialization of the model in the first frame of the image se-

(a) (b)

Figure 1. (a) Instance of an image that can be handled by our algorithm, (b) Instance of an image that cannot be handled by our algorithm.

quence. Concerning model acquisition, existing approaches use models of the human body whose parts are either approximated with simple shapes and their dimensions have been manually measured [9, 18] or models whose shape and/or dimensions have been determined based on camera input data. In this second category, methods have been developed to obtain models of human body parts from multiple cameras [10, 11, 14] or range data [7]. Concerning posture estimation, methods have been presented that use either one [4, 5, 15, 19], or multiple cameras [1, 6, 8, 12, 13]. However, in most of the existing tracking approaches the user specifies an approximate position and posture from the human model at the first frame of the image sequence [5, 13, 17]. In contrast, Bregler and Malik [4] for the initialization step of their human tracking method, they minimize a cost function over position, angles and body dimensions. In particular, a user selects the 2D joint locations and then a 3D pose is found by minimizing the sum of the squared differences between the projected model joint locations and the corresponding model joint locations. The authors mention that they had good results with a Quasi-Newton method and a mixed quadratic and line search procedure. However, no information is provided about the accuracy and repeatability of their method, nor for what class of postures and human body dimensions does the method work. The contribution of our paper is a systematic study and a technique that takes into consideration statistical anthropometric information to constrain the estimation process.

3. Problem Statement

The human musculosketelal system is composed from a series of jointed links, which can be approximated as rigid bodies. Human motion estimation is aimed at quantitatively describing the spatial motion of body segments and the movements of the joints connecting those segments. A hallmark of the individuality of the people that we encounter in our daily life is the variation of their anthropometric measurements. If we assume that we have no anthropometric information for the subject that we are observing, the problem of anthropometry and pose estimation from a single image can be formulated as follows: *Given a set of points in an image that correspond to the projection of landmark points of a human subject, estimate both the anthropometric measurements (up to a scale) of the subject and his/her pose that best match the observed image.* By the term "up to a scale", we refer to the fact that from a single camera (under perspective projection) we cannot infer absolute lengths (like "upper-leg-length" and "shoulder-width") but ratios of lengths. Therefore, in the following when we refer to the estimation of the anthropometric measurements, we imply the estimation of ratios of lengths like "upper-leg-length" over "shoulder-width".

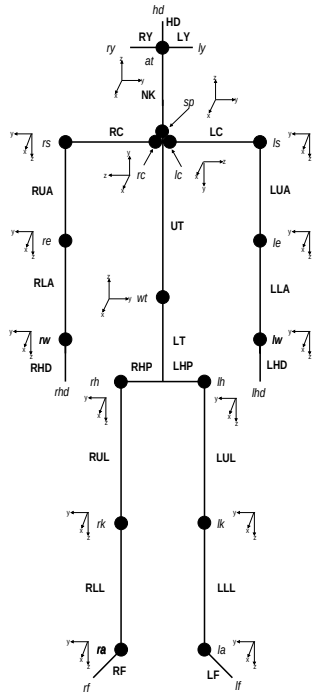
4. Methods

Our algorithm has the following steps:
Algorithm: Anthropometry and Pose Estimation
Step 1: Selection of projected landmarks.
Step 2: Choice of initial Stick Model.
Step 3: Initial estimates for pose.
Step 4: Iterative minimization over lengths and angles.
First, we present the Stick Human Body Model (SM) that we have developed, and then we will elaborate in the above steps in detail.

4.1. Stick Human Body Model

For the purposes of this research, we have developed a generic Stick Human Body Model inspired by the human body model employed at the Human Modeling and Simulation Center at University of Pennsylvania [3]. The model consists of a set of segments connected by joints. Specifically, a Stick Model is a tree (w, V, A), where V is a set of sites/landmarks and A is a collection of edges (segments) with endpoints in V, and $w \in V$ is the root. In our case, A ={ HD, RY, LY, NK, UT, RC, LC, RUA, LUA, RLA, LLA, RHD, LHD, LT, RHP, LHP, RUL, LUL, RLL, LLL, RF, LF} as enumerated in Figure 2(b), and the set of landmarks consists of a set of joints J={at, sp, la, lc, le, lh, lk, ls, lw, ra, rc, re, rh, rk, rs, rw, wt} (information about the SM's joints is provided in Table 1), and other landmarks M ={ry (right eye), ly (left eye), rhd (base of the right middle finger), lhd (base of the left middle finger), rf (tip of the right foot), lf (tip of the left foot)} (V = J ∪ L).

A local coordinate system is attached to each body part. The kinematics are represented by a transformation tree



(a)

ID	Segment
HD	Head
RY	Right Eye
LY	Left Eye
NK	Neck
UT	Upper Torso
RC	Right Clavicle
LC	Left Clavicle
RUA	Right Upper Arm
LUA	Left Upper Arm
RLA	Right Lower Arm
LLA	Left Lower Arm
RHD	Right Hand
LHD	Left Hand
LT	Lower Torso
RHP	Right Hip
LHP	Left Hip
RUL	Right Upper Leg
LUL	Left Upper Leg
RLL	Right Lower Leg
LLL	Left Lower Leg
RF	Right Foot
LF	Left Foot

(b)

Figure 2. (a) Stick Human Body Model (SM) and its associated coordinate systems, (b) Names of the SM's segments.

whose root is the subject coordinate system and whose leaves are the coordinate systems of head, hands, and feet. The origin of the subject coordinate system is the waist joint. Figure 2(a) depicts the local coordinate systems of the stick human model which correspond to the joints listed in Table 1. Notice that every joint has both rotational degrees of freedom and translational degrees of freedom to allow for segment scaling. For each joint an upper limit and a lower limit are required. The default data for the joints are extracted from [16]. In addition, using the anthropometric measurements in [16], we have build a database that contains statistical information related to the segment lengths of our simplified model. Using this statistical information, we have computed a cadre family, also known as the boundary family, of simplified models [2]. The cadre family is a multivariate representation of the extremes of the population distribution. It has the ability to span the multivariate space in a systematic fashion and to capture a significant amount of the variance in the space using a small number of sample human models. The probability density function of the multivariate normal distribution is defined by:

$$f(\mathbf{x}) = ((2\pi)^k |\mathbf{\Sigma}|)^{-\frac{1}{2}} \exp^{(-\frac{1}{2}(\mathbf{x}-\mathbf{m})^\top \mathbf{\Sigma}^{-1}(\mathbf{x}-\mathbf{m}))}, \quad (1)$$

where k is the number of dimensions (in our case the variables are the lengths of the 22 segments of our Stick Model Figure 2(b)), $\mathbf{x}$ is a random vector, $\mathbf{m}$ and $\mathbf{\Sigma}$ are the mean and the covariance matrix of the population. If we define the quadratic form $Q(\mathbf{x}) = \mathbf{x}^\top \mathbf{\Sigma} \mathbf{x}$, then $Q(\mathbf{x})$ defines a conic surface whose shape depends on the diagonal elements of $\mathbf{\Sigma}$, and since $\mathbf{\Sigma}$ is a variance-covariance matrix this surface is a hyperellipsoid. As we know, the principal components $P_i (i = 1, \ldots, 7)$ are linear combinations of the original k variables (the lengths) as follows: $P_i = \alpha_{i1} l_1 + \alpha_{i2} l_2 + \ldots + \alpha_{ik} l_k$. Computation of the weight coefficients α_{ip} is constrained by:

$$\sum_{p=1}^{k} \alpha_{ip}^2 = 1 \quad (2)$$

We prioritize the principal components according to their associated captured variance and we keep seven components to limit the sum of the associated variances to be equal to the desired total captured variance. Next, the component scores matrix is multiplied by a binary matrix formed by all combinations of ± 7. This yields standard score vectors. Using these vectors, we can generate a family of stick models that includes not only hyperellipsoid surface points but also the axial points and the mean. The total number of SMs produced is 139 (in general $2^s + 2s + 1$, where s is the number of principal component vectors kept). A sample of these SMs is depicted in Figure 3. For these 139 models we maintain information concerning the segment lengths $\{l_i^q\}$ where $i = 1, \ldots, 22$ (the segments' names are given in Figure 2(b)) and $q = 1, \ldots, 139$. Using this information we compute means and the standard deviations.

ID	Joint	From	To	DOF	PR
at	atlanto occipital	NK	HD	Tz*Rz*Ry*Rx	3
sp	solar plexus	UT	NK	Tz*Ry*Rz*x	2
la	left ankle	LLL	LF	Tx*Rz*Rx*Ry	4
lc	left clavicle	UT	LC	Tz*Rx*Ry	3
le	left elbow	LUA	LLA	Tz*Ry	5
lh	left hip	LT	LUL	Tz*Rz*Rx*Ry	2
lk	left knee	LUL	LLL	Tz*R-y	3
ls	left shoulder	LC	LUA	Tz*Rz*Rx*Ry	4
lw	left wrist	LLA	LHD	Tz*Ry*Rx*Rz	6
ra	right ankle	RLL	RF	Tx*R-z*R-x*Ry	4
rc	right clavicle	UT	RC	Tz*R-x*Ry	3
re	right elbow	RUA	RLA	Tz*Ry	5
rh	right hip	LT	RUL	Tz*R-z*R-x*Ry	2
rk	right knee	RUL	RLL	Tz*R-y	3
rs	right shoulder	RC	RUA	Tz*R-z*R-x*Ry	4
rw	right wrist	RLA	RHD	Tz*Ry*R-x*R-z	6
wt	waist	LT	UT	Tz*Ry*Rz*Rx	1

Table 1. Information related to the joints of the Stick Model.

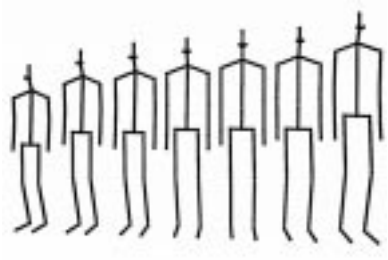


Figure 3. Sample models fr om the distribution of SMs (7 out of 139).

4.2. Step 1: Input Data

We have developed a simple user interface that allows the user to select the projection of visible landmarks of the subject's body (see Figure 6(a)). In addition, the user marks the segments whose orientation is almost parallel to the image plane. For example, in Figure 6(a) the white dots depict projection of landmarks associated with segments whose orientation is almost parallel to the image plane, and the black dots depict all other selected landmarks. Although, information from both types of landmarks will be used for pose estimation, initial length estimates will be based on the projected length of the segments whose orientation is almost parallel to the image plane only.

4.3. Step 2: Initial Anthropometric Estimates

Our basic assumption is that there is a number of segments whose orientation is almost parallel to the image plane and therefore we can obtain good approximation ratios for them. Thus, we cannot handle images like the one depicted in Figure 1(b) since one cannot locate segments that are almost parallel to image plane to obtain reliable initial anthropometric estimates. Let l_i and p_i be the length and the projected length of a segment i, respectively. Let $\mathrm{I} \subset [1, 22]$ be the index set of the segments whose orientation is almost parallel to the image plane, and $n = |\mathrm{I}| \leq 22$. Using the measurements p_k ($k \in \mathrm{I}$), we compute all possible ($\frac{n(n-1)}{2}$) ratios $p_{ij} = \frac{p_i}{p_j}$ where $i < j$, and $i, j \in \mathrm{I}$. Based on these ratios, we select one SM from our family of 139 SMs whose length ratios closely match the ratios computed from the image. To accomplish this goal, we determine $q^\star = \arg\min_q \sum_{i,j \in \mathrm{I}} (r_{ij}^q - p_{ij})^2$, where $r_{ij}^q = \frac{l_i^q}{l_j^q}$, $i < j$, and $q = 1, \ldots, 139$. The length measurements of the selected SM are the initial segment length estimates for our algorithm.

To facilitate the overall understanding of our algorithm, we first present the fourth step in the next section and then the third step in Section 4.5.

4.4. Step 4: Minimization of the cost function

The variables that we want to estimate their values are the lengths of the body segments and their pose. Therefore, we will solve a system of equations were prior information about the human body (e.g., relations between lengths of segments) will provide constraints to an optimization that minimizes the discrepancy between the synthesized appearance of the SM (for that pose) and the image data of the subject in the given image.

As mentioned earlier, the user selects a set of points on the image that correspond to the projection of the sites of the Stick Model. For each of these points, we setup a point-to-line constraint, since the site will lie on a line that goes through the center of the camera and the projection of a landmark. Let $\mathbf{o} \in \mathbf{R}^3$ be the camera's center of projection, $\mathbf{m}_i \in \mathbf{R}^3$ be the position of a SM's site, and $\mathbf{m}_i^p \in \mathbf{R}^3$ be the corresponding projection point selected by the user. Then, the constraint line is given by: $\mathbf{c}_i = \mathbf{o} + \lambda \mathbf{d}_i$, where $\mathbf{d}_i = \frac{(\mathbf{m}_i^p - \mathbf{o})}{\|\mathbf{m}_i^p - \mathbf{o}\|}$. Our objective function is $O = \sum_i distance(\mathbf{m}_i, \mathbf{c}_i)$. We seek to minimize the value of this function using a BFGS nonlinear solver [20]. Due to the large number of degrees of freedom, in order not to be trapped in local minima and to obtain an anthropometrically plausible correct answer, we apply the solver in a hierarchical manner and we employ a number of constraints. Statistical information about the proportions of the human body and the range of motion of each joint are integrated into the hierarchical optimization method as a set of constraints.

Hierarchical Solver: First, to facilitate and expedite the minimization process, we assign a priority to each joint and end effector, and we schedule our optimization to proceed in a hierarchical manner starting with joints closer to the waist moving outwards. The priorities for each joint are detailed in the column named PR in Table 1.

Constraints: Three classes of constraints are applied: 1) constraints derived from the joint limit information associated with the range of motion of a joint, 2) constraints that enforce the symmetry between the left and right sides of the subject (e.g., the length of the left upper arm is equal to the length of the right upper arm), and 3) constraints that enforce proportions. For the symmetry constraints in particular, we require that the ratios $\{\frac{\mathrm{LY}}{\mathrm{RY}}, \frac{\mathrm{LC}}{\mathrm{RC}}, \frac{\mathrm{LUA}}{\mathrm{RUA}}, \frac{\mathrm{LLA}}{\mathrm{RLA}}, \frac{\mathrm{LHD}}{\mathrm{RHD}}, \frac{\mathrm{LHP}}{\mathrm{RHP}}, \frac{\mathrm{LUL}}{\mathrm{RUL}}, \frac{\mathrm{LLL}}{\mathrm{RLL}}, \frac{\mathrm{LF}}{\mathrm{RF}}\}$ are within ϵ distance from the value one. Thus, the variables whose values will be estimated are the lengths of the following 12 segments: HD, NK, LC, LUA, LLA, LHD, UT, LT, LHP, LUL, LLL, and LF.

Proportions: Our goal is to guide the minimization solver to a solution for the pose that is feasible (hence the use of joint limit constraints) and anthropometrically plausible. Hence, we seek to find ratios of measurements that need to be maintained during the minimization process. In the following, we describe our algorithm to determine which ratios will be used to constrain the estimation process. The objective of this algorithm is to find a minimum set of ratios that constrain all segment lengths. We formulate the problem as a *set covering* problem. If R is the set of all possible ratios, we want to find a set B $\subset$ R to cover the set L of the segment lengths. Let c_{ij} be the absolute value of the correlation between ratios $r_i \in$ R and $r_j \in$ R, and let $l_k \in$ L be the length of segment k. For each ratio $r_i = \frac{l_i^1}{l_i^2}$ ($l_i^1, l_i^2 \in L$), we define the quantities of weight, degree, cover and goodness. Let $w_i = weight(r_i) = \frac{\sigma(r_i)}{\mu(r_i)}$. The variance $\sigma^2(r_i)$ is an indication of the precision of the statistical information concerning the ratio r_i. Therefore, the smaller the quantity $\frac{\sigma(r_i)}{\mu(r_i)}$, the more constrained the ratio. Let $cover(r_i, l_k)$ be the quantity that measures to which extend the ratio r_i constrains the length l_k. Then, if we define $d_i = degree(r_i) = \sum_{l_k} cover(r_i, l_k)$, the goodness g_i of a ratio is given by: $g_i = goodness(r_i) = \frac{d_i}{w_i}$. In the following, we outline the steps of the algorithm.

Algorithm: Ratio Selection

Step 1: Set B = $\emptyset$.

Step 2: $\forall\ (r_i, l_k) \in$ (R $\times$ L) set:

cover(r_i, l_k)=$\begin{cases} 1 & \text{if } (r_i = \frac{l_i^1}{l_i^2} \wedge (l_k = l_i^1 \vee l_k = l_i^2)) \\ c_{ib} & \text{otherwise.} \end{cases}$

where $c_{ib} = \max_j\{c_{ij}\}, and$

$(r_j = \frac{l_j^1}{l_j^2} \wedge (l_k = l_j^1 \vee l_k = l_j^2))$.

Step 3: $\forall\ l_k \in L$ set $care(l_k)$ = 0 .

Step 4: $\forall j, r_j \in$ R\ B, compute d_j and g_j .

Step 5: Select r_m such that $g_m = \max(g_j)\ \forall j, r_j \in$ R\B .

Step 6: B = B$\cup\{r_m = \frac{l_m^1}{l_m^2}\}$, $care(l_m^1)$ += cover (r_m, l_m^1), and $care(l_m^2)$ += cover (r_m, l_m^2).

Step 7: If $care(l_k) \geq 1\ \forall k,\ l_k \in L$ then done. (That means that we have found a set of ratios B that constrain all the lengths). Otherwise, $\forall j, r_j \in$R\B $(j \neq m)$ cover(r_j, l_k) = $\max\{0,$ cover $(r_j, l_k) - care(l_k)\}$ and goto step 4.

The set B contains the following ratios: $\frac{UT}{LF}, \frac{UT}{LLL}, \frac{UT}{LLA}, \frac{UT}{LUA}, \frac{LUL}{LLL}$, and $\frac{LUA}{LLA}$. Based on this information, the constraint for a ratio $r_j = \frac{l_j^1}{l_j^2}$ ($\forall r_j \in$ B) takes the form:

$$l_j - \sigma_j \leq l_j \leq l_j + \sigma_j\ ,$$

where l_j is selected from $\{l_j^1, l_j^2\}$ and corresponds to the segment with the minimum value of $\frac{\mu_j}{\sigma_j}$.

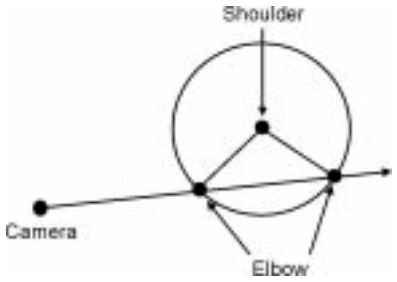


Figure 4. Initial pose estimates.

4.5. Step 3: Initial estimates for the pose

In order for the nonlinear solver not to get trapped to local minima, we use a geometric method for providing an initial guess for the pose of some segments. This procedure is applied only to segments for which both endpoints were selected by the user. Let $\mathbf{m}_i^p \in \mathbf{R}^3$ be the projection of a site m_i in the image, $l_i > 0$ be the length of the segment of which this landmark is the end-effector, and $\mathbf{j} \in \mathbf{R}^3$ be the position of the parent joint of that landmark on the Stick Model. By construction, the following equation applies:

$$\|\mathbf{o} + \lambda \mathbf{d}_i - \mathbf{j}\| = l_i$$

where $\mathbf{d}_i$ is the unit direction between the camera and $\mathbf{m}_i^p$, as defined earlier. This quadratic equation has two solutions (see Figure 4):

$$\lambda_1 = \mathbf{d}_i \bullet (\mathbf{j} - \mathbf{o}) + \sqrt{[\mathbf{d}_i \bullet (\mathbf{o} - \mathbf{j})]^2 - ||\mathbf{o} - \mathbf{j}||^2 + l_i^2}$$

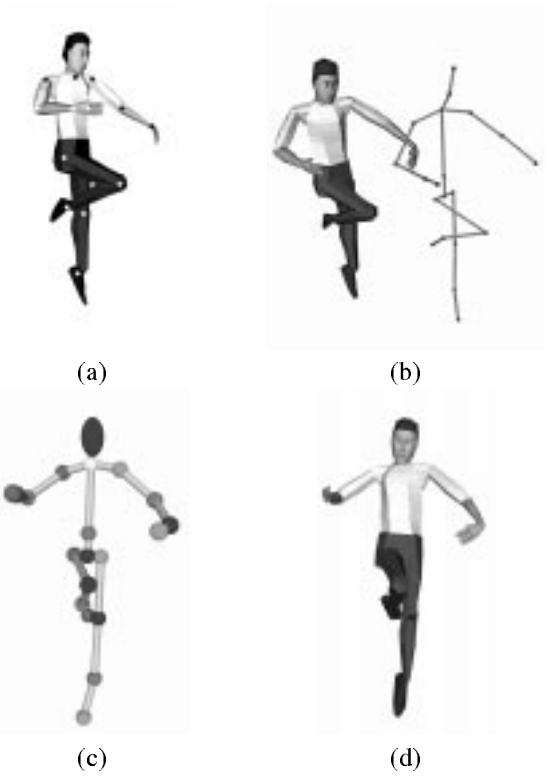
$$\lambda_2 = \mathbf{d}_i \bullet (\mathbf{j} - \mathbf{o}) - \sqrt{[\mathbf{d}_i \bullet (\mathbf{o} - \mathbf{j})]^2 - ||\mathbf{o} - \mathbf{j}||^2 + l_i^2}$$

that correspond to the intersection of the line $\mathbf{o} + \lambda_k \mathbf{d}_i$ with the sphere of radius l_i centered at $\mathbf{j}$. The two possible initial guesses for the position of site m_i are $\mathbf{m}_{i1} = \mathbf{o} + \lambda_1 \mathbf{d}_i$ and $\mathbf{m}_{i2} = \mathbf{o} + \lambda_2 \mathbf{d}_i$. Then, joint limit information is used to prune the solutions that are not feasible. If both positions are feasible, then both are used as initial values for the nonlinear solver.

5. Results and Discussion

We have performed a number of experiments on synthetic and real data to assess the accuracy, limitations and advantages of our approach. In all the example input images, the white dots depict projection of landmarks associated with segments whose orientation is almost parallel to the image plane, and the black dots depict all other selected landmarks (see for example the input image depicted in Figure 6(a)). In the first experiment, we applied our technique to an image created using the virtual human modeling tool EAI Jack®. Figure 5 (a) depicts the selected points in the

input image, while Figure 5(b) depicts the reconstructed 3D model. Figures 5(c,d) depict the reconstructed 3D model in novel views. Tables 2 and 3 contain statistical information related to the accuracy of the estimation process.



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

Figure 5. (a) Input image and selected points, (b) reconstructed 3D model, (c) novel view of the SM model along with (d) the virtual human model.

	$\frac{\text{UT+LT}}{\text{LF}}$	$\frac{\text{UT+LT}}{\text{LLL}}$	$\frac{\text{UT+LT}}{\text{LLA}}$	$\frac{\text{UT+LT}}{\text{LUA}}$	$\frac{\text{UL}}{\text{LLL}}$	$\frac{\text{LLL}}{\text{LLA}}$
Actual	2.32	1.47	2.04	2.00	0.99	1.38
Estimated	2.30	1.49	2.10	1.99	0.97	1.41
PE %	0.86	1.36	2.94	0.50	2.02	2.17

Table 2. Accuracy of the length estimates for the synthetic experiment.

In the second experiment, we applied our technique to a real image from the subject *Vanessa* whose anthropometric dimensions were manually measured. Figure 6(a) depicts the selected points, Figure 6(b) depicts the reconstructed model overlayed to the image and Figures 6(c,d) depict the model from novel views. Table 4 captures the percentage errors (PE) of the estimation process. We observe that the estimation of anthropometric information is within 3.3% of

Joint	Actual Values	Estimated	PE %
at	$(0.00^o, 0.00^o, 0.00^o)$	$(0.00^o, 0.00^o, 0.00^o)$	0.00
sp	$(-0.50^o, -2.00^o, 0.00^o)$	$(-0.48^o, -1.98^o, 0.00^o)$	1.37
la	$(12.97^o, -15.37^o, -63.40^o)$	$(13.60^o, -16.50^o, -65.00^o)$	3.09
lc	$(12.24^o, 0.99^o)$	$(12.01^o, 0.99^o)$	1.87
le	(41.84^o)	(42.31^o)	1.12
lh	$(-16.80^o, 2.14^o, -2.73^o)$	$(-16.53^o, 2.12^o, -2.69^o)$	1.60
lk	(0.00^o)	(0.0^o)	0.00
ls	$(2.50^o, 36.47^o, 3.41^o)$	$(2.50^o, 36.87^o, 3.37^o)$	1.09
lw	$(42.67^o, -39.12^o, 1.85^o)$	$(43.02^o, -40.12^o, 1.65^o)$	1.86
ra	$(32.28^o, 33.54^o, -50.13^o)$	$(31.98^o, 33.32^o, -49.81^o)$	0.72
rc	$(11.70^o, 11.41^o)$	$(11.61^o, 11.51^o)$	0.82
re	(62.74^o)	(63.05^o)	0.49
rh	$(-8.63^o, -13.40^o, 63.87^o)$	$(-8.62^o, -13.40^o, 63.77^o)$	0.15
rk	(129.13^o)	(131.02^o)	1.46
rs	$(26.20^o, 31.11^o, 42.44^o)$	$(26.20^o, 31.07^o, 43.01^o)$	0.97
rw	$(-5.35^o, -6.55^o, -54.53^o)$	$(-5.15^o, -7.05^o, -54.50^o)$	0.98
wt	$(0.00^o, 0.00^o, 0.00^o)$	$(0.00^o, 0.00^o, 0.00^o)$	0.00

Table 3. Accuracy of the pose estimates for the synthetic experiment.

the anthropometric dimensions of the subject. In general, we have performed numerous other experiments with a variety of subjects whose anthropometric dimensions are known with similar very encouraging results.

	$\frac{\text{UT+LT}}{\text{LF}}$	$\frac{\text{UT+LT}}{\text{LLL}}$	$\frac{\text{UT+LT}}{\text{LLA}}$	$\frac{\text{UT+LT}}{\text{LUA}}$	$\frac{\text{UL}}{\text{LLL}}$	$\frac{\text{LLL}}{\text{LLA}}$
Actual	2.43	1.16	2.59	2.23	0.87	2.23
Estimated	2.35	1.15	2.55	2.17	0.87	2.22
PE %	3.29	0.86	1.54	2.69	0.00	0.45

Table 4. Accuracy of the length estimates for the subject *Vanessa*.

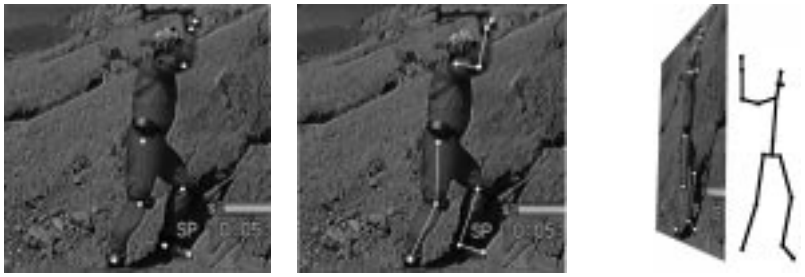
In the third experiment, we applied our algorithm to a variety of images from a variety of application domains, where anthropometric information about the subjects was not available, Figures 7(a), 8(a,f), 9(a), depict the input images along with the selected points, while Figures 7(b-c), 8(b-f), 9(b-d), depict the reconstructed model from various viewpoints.

6. Conclusion

In this paper, we have presented a four step technique for generating anthropometric and posture information for a human subject from a single image. The user initially selects a set of image points that constitute the projection of selected landmarks. Based on the image coordinates of the selected points and anthropometric statistics, pose and anthropometric measurements are obtained by minimizing an appropriate cost function subject to the associated constraints. The novelty of our approach is the use of anthropometric statistics to constrain the estimation process that allows the simultaneous estimation of both anthropometry and pose. We have demonstrated the accuracy, advantages and limitations of our method for various classes of both synthetic and real input data.

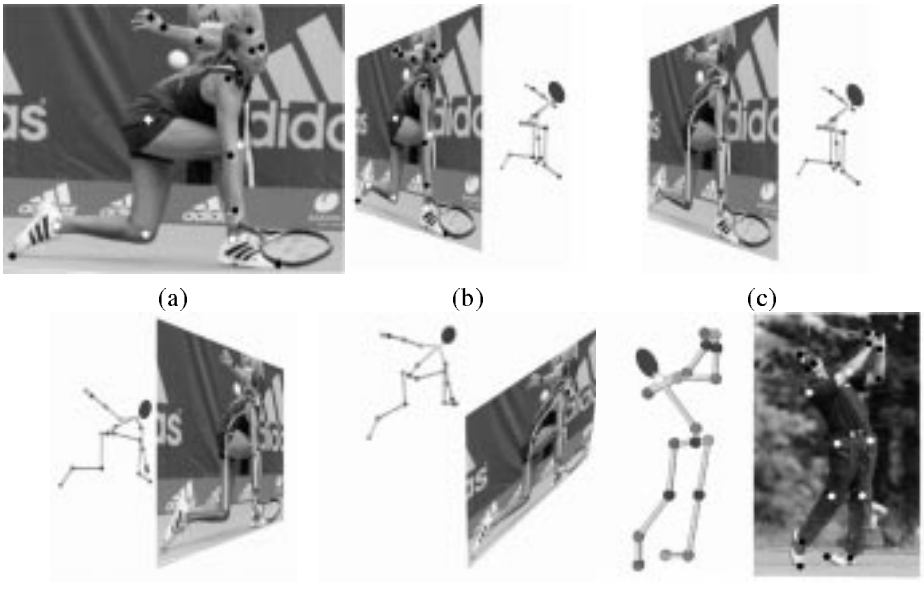
(a) (b) (c) (d)

Figure 6. (a) Selected points, (b) reconstructed model overlayed to the image, and (c,d) the model from novel views.



(a) (b) (c)

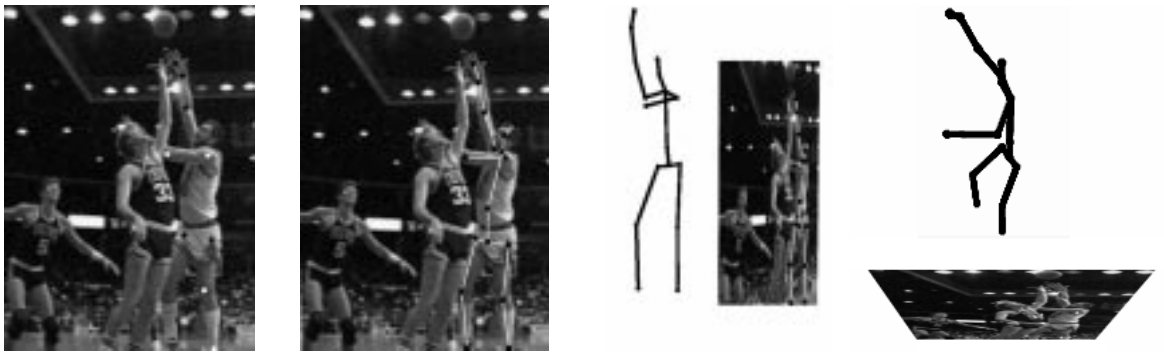
Figure 7. (a) Input landmarks, and (b,c) two views of the reconstructed model.



(a) (b) (c)

(d) (e) (f)

Figure 8. (a) Input landmarks, (b-e) various views of the reconstructed model, an (f) an example from golf.



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

Figure 9. (a) Input landmarks, and (b-d) various views of the reconstructed model.

References

[1] A. Azarbayejani, C. Wren, and A. Pentland. Real-time 3-D tracking of the human body. In *Proceedings of the IMAGE'COM 96*, Bordeaux, France, May 1996.

[2] F. Azuola. *Error in the Representation of Anthropometric Data By Human Figure Models*. PhD thesis, University of Pennsylvania, Philadelphia, PA, 1996.

[3] N. I. Badler, C. B. Phillips, and B. L. Webber. *Simulating humans: computer graphics animation and control*. Oxford University Press, New York, NY, 1993.

[4] C. Bregler and J. Malik. Tracking people with twists and exponential maps. In *Proceedings of the 1998 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 8–15, Santa Barbara, CA, June 23-25 1998.

[5] T. Cham and J. Rehg. A multiple hypothesis approach to figure tracking. In *Proceedings of the 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume II, pages 239–245, Fort Collins, CO, June 23-25 1999.

[6] Q. Delamarre and O. Faugeras. 3D articulated models and multi-view tracking with silhouettes. In *Proceedings of the 7th International Conference of Computer Vision*, pages 716–721, Kerkyra, Greece, September 20-27 1999.

[7] I. Douros, L. Dekker, and B. F. Buxton. An improved algorithm for reconstruction of the surface of the human body from the 3D scanner data using local B-spline patches. In *IEEE International Workshop on Modeling People*, pages 29–36, Corfu, Greece, September 20 1999.

[8] D. M. Gavrila and L. S. Davis. 3-D model-based tracking of humans in action: a multi-view approach. In *Proceedings of the 1996 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 73–80, San Francisco, CA, June 18-20 1996.

[9] L. Goncalves, E. D. Bernardom, E. Ursella, and P. Perona. Monocular tracking of the human arm in 3D. In *Proceedings of the Fifth International Conference on Computer Vision*, pages 764–770, Boston, MA, June 20-22 1995.

[10] J. Gu, T. Chang, I. Mak, S. Gopalsamy, H. C. Shen, and M. F. Yuen. A 3D reconstruction system for human body modeling. In N. Magnenat-Thalmann and D. Thalmann, editors, *Proceedings of CAPTECH 98*, pages 229–241, Geneva, Switzerland, November 26-27 1998.

[11] A. Hilton. Towards model-based capture of a person's shape, appearance and motion. In *IEEE International Workshop on Modeling People*, pages 37–44, Corfu, Greece, September 20 1999.

[12] S. Iwasawa, J. Ohya, K. Takahashi, T. Sakaguchi, S. Kawato, K. Ebihara, and S. Morishima. Real-time, 3D estimation of human body postures from trinocular images. In *IEEE International Workshop on Modeling People*, pages 3–10, Corfu, Greece, September 20 1999.

[13] I. A. Kakadiaris and D. Metaxas. Model-based estimation of 3D human motion with occlusion based on active multi-viewpoint selection. In *Proceedings of the 1996 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 81–87, San Francisco, CA, June 18-20 1996.

[14] I. A. Kakadiaris and D. Metaxas. 3D Human body model acquisition from multiple views. *International Journal of Computer Vision*, 30(3):191–218, 1998.

[15] H. J. Lee and Z. Chen. Determination of 3D human body postures from a single view. *Computer Vision, Graphics, and Image Processing*, 30:148–168, May 1985.

[16] National Aeronautics and Space Administration. Man systems integration standards. Technical report, 1987.

[17] R. Plankers, P. Fua, and N. D'Apuzzo. Automated body modeling from video sequences. In *IEEE International Workshop on Modeling People*, pages 45–52, Corfu, Greece, September 20 1999.

[18] J. M. Rehg and T. Kanade. Model-based tracking of self-occluding articulated objects. In *Proceedings of the Fifth International Conference on Computer Vision*, pages 612–617, Boston, MA, June 20-22 1995.

[19] S. Wachter and H. H. Nagel. Tracking of persons in monocular image sequences. In *Proceedings of IEEE Nonrigid and Articulated Motion Workshop*, pages 2–9, Puerto Rico, June 16 1997.

[20] J. Zhao and N. I. Badler. Nonlinear programming for highly articulated figures. *ACM Transactions on Graphics*, 13(4):313–336, October 1994.