3-D Modeling from Concept Sketches of Human Characters with Minimal User Interaction

Adrian Johnston[1], Gustavo Carneiro[1], Ren Ding[2], Luiz Velho[3]
[1]Australian Centre for Visual Technologies, The University of Adelaide, Australia
[2]Esuperfund, Melbourne, Australia
[3]Instituto de Matemática Pura e Aplicada, Rio de Janeiro, Brazil

***Abstract*—We propose a new methodology for creating 3-D models for computer graphics applications from 2-D concept sketches of human characters using minimal user interaction. This methodology will facilitate the fast production of high quality 3-D models by non-expert users involved in the development process of video games and movies. The workflow starts with an image containing the sketch of the human character from a single viewpoint, in which a 2-D body pose detector is run to infer the positions of the skeleton joints of the character. Then the 3-D body pose and camera motion are estimated from the 2-D body pose detected from above, where we take a recently proposed methodology that works with real humans and adapt it to work with concept sketches of human characters. The final step of our methodology consists of an optimization process based on a sampling importance re-sampling method that takes as input the estimated 3-D body pose and camera motion and builds a 3-D mesh of the body shape, which is then matched to the concept sketch image. Our main contributions are: 1) a novel adaptation of the 3-D from 2-D body pose estimation methods to work with sketches of humans that have non-standard body part proportions and constrained camera motion; and 2) a new optimization (that estimates a 3-D body mesh using an underlying low-dimensional linear model of human shape) guided by the quality of the matching between the 3-D mesh of the body shape and the concept sketch. We show qualitative results based on seven 3-D models inferred from 2-D concept sketches, and also quantitative results, where we take seven different 3-D meshes to generate concept sketches, and use our method to infer the 3-D model from these sketches, which allows us to measure the average Euclidean distance between the original and estimated 3-D models. Both qualitative and quantitative results show that our model has potential in the fast production of 3-D models from concept sketches.**

I. Introduction

We propose a new methodology that produces high-quality 3-D human character meshes from 2-D concept sketches in an almost fully automatic manner. Such approach will enable fast and easy production of 3-D base meshes of video game and movie characters, streamlining their production process. Currently, in these industries, the base mesh production starts with the creation of the concept sketch and the subsequent 3-D modeling process, where the artist sets the sketch as a background image in a 3-D modeling tool, such as 3DS Max or Maya, and deforms and subdivides simple primitives (e.g., cubes or spheres) to fit the silhouette and edge details of that sketch. Once this base mesh has been developed, it is added to a 3-D sculpting software such as ZBrush or Mudbox, which allows artists to add details to the base 3-D mesh. Our methodology will reduce the time to produce such base meshes by the automation of the following three steps (see Fig. 1): 1) automatic estimation of the 2-D body pose from the concept sketch (Fig. 1-(a-b)) [1]; 2) automatic estimation of the 3-D body pose and camera motion from 2-D landmarks (Fig. 1-(c)) [2]; and 3) optimization of the 3-D body mesh by minimizing the matching error between the projected mesh and the concept sketch (Fig. 1-(d),(a)).

Our main contributions in the development of this methodology are as follows:

1) adaptation of the method that estimates the 3-D body pose and camera motion from 2-D landmarks [2] such that it can deal with non-standard human body proportions and constrained camera motion (representing the concept sketch frontal view); and
2) design and implementation of the optimization method based on a sampling importance re-sampling algorithm, which searches for a 3-D body mesh (on an underlying low-dimensional linear model of human shapes) using the matching quality between the concept sketch and the projection of the estimated 3-D pose.

We show a qualitative experiment of the 3-D modeling process using seven concept sketches of male and female characters. We also show a quantitative experiment showing the matching error performance (average Euclidean distance) of the methodology with respect to a known 3-D model that is used to generate an artificial concept sketch. The results from the qualitative and quantitative experiments show that our proposed methodology has a potential to facilitate the production of 3-D base meshes.

A. Literature Review

The (semi-)automated creation of 3-D models is one of the major focuses of current research in 3-D computer graphics. Sketch-based 3-D modeling has become an important alternative to the complex freeform surface manipulation [3], but its underlying principle of sketching in 3-D is one of the main issues of this approach because artists are used to work in 2-D. Tools that enable artists to working effectively in 2-D to produce 3-D models, such as SketchUp, are becoming popular, but are usually not ideal for the design of complex human characters.

Recently, some research groups have proposed methods especifically designed to estimate 3-D models of human char-

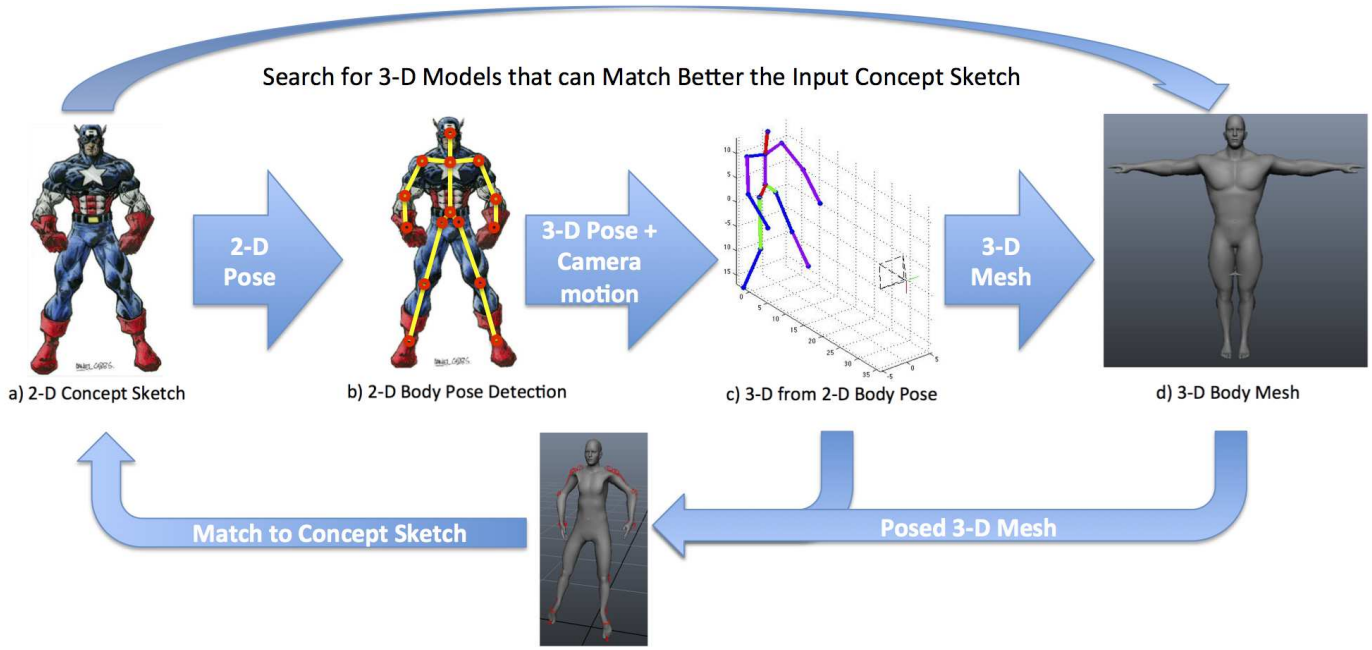

Fig. 1. Our methodology takes an input concept sketch (a) and estimates the 2-D pose, where the user can manually adjust the joint positions (b). This result is used to infer the 3-D pose and camera motion (c), and then we run a sampling importance resampling algorithm using a learned model of 3-D meshes and the inferred 3-D pose to match the 3-D body mesh to the concept sketch (d-a).

acters. For instance, ArtiSketch [4] is a method that can take several 2-D views of a human character in a video game and their respective user-input 3-D pose to produce a 3-D articulated model. This work is similar to ours, but the main difference is that our proposal requires less user interaction, compared with ArtiSketch, which requires intensive user interaction. Another important work that is related to our proposal is the 3-Sweep [5], which consists of a semi-automated way to produce complex 3-D models with little user interaction. Essentially, with 3 clicks non-artists can obtain the underlying 3-D model from a single 2-D picture of an object (note that complex objects require several stages of 3 clicks to model all their parts). However, the strong assumptions about the type of objects that can be modelled with this technique renders the modelling process of a human character extremely complicated. Kraevoy et al. [6] propose a method for estimating 3-D human meshes by matching a 3-D template shape prior model to contour drawings of human figures. The interesting contribution about this work is the technique used to find the correspondences between contour points and model vertices. However, this approach has many limitations compared to our methodology, such as the need for having a clean background and a constrained input data, based on a contour drawing, instead of an actual concept sketch. Moreover, this method by Kraevoy et al. [6] appears to work only with real humans, as opposed to the ”deformed” human figures present in concept sketches.

The most relevant paper to our work is the method by Guan et al. [7], which is based on similar steps to our proposed methodology. Essentially, the user first guides the estimation of the initial 3-D articulated body pose and shape, then a segmentation algorithm detects the boundaries of the human body on the image. A low-dimensional linear model of human shape is then used for the estimation of the 3-D body shape, where the optimization takes into account body pose, shape, reflectance, and scene lighting, in order to produce a synthesized body that is then used to match the input image. The main limitations that affect this approach, and not ours are: 1) it only works with naked or minimally clothed people; 2) it does not work properly with the challenging shading present in concept sketches considered in our work; 3) it does not take into account non-standard body proportions usually present in concept sketches; and 4) the 2-D body joint detector is not present, so it requires more user input than our method. Finally, recently and developed in parallel to our own work, Kazmi et al. [8] have proposed a methodology that estimates 3-D body meshes from concept sketches based on the following steps. First the methodology uses an input concept sketch based on occluded contours, which are then associated with 3-D human models using a model that takes several poses and human body shapes as its training set. Then, the user selects manually anchor points to be matched to the 3-D model, so that it can match the input image, similarly to the approach by Kraevoy et al. [6]. Though similar to our own work in terms of the final goals, Kazmi et al.’s method has many limitations that do not apply to our own, such as: 1) the need of a quite constrained input based on occluded contours, which is not the usual type of image in a concept sketch; and 2) the need of relatively intense user interaction in selecting anchor points to drive the deformation from the 3-D mesh to the input concept

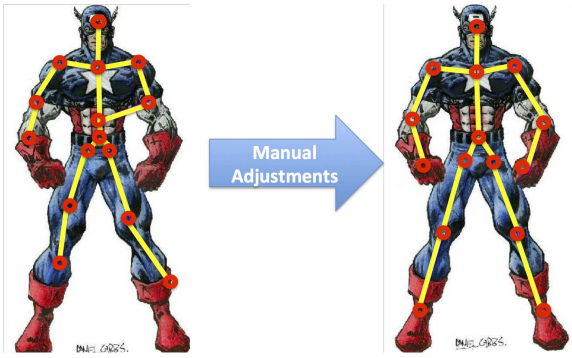

Fig. 2. Example of the automatic 2-D pose detection [1] (left) and the manual adjustments of the joint locations (right).

sketch.

II. Methodology

In this section, we first describe the 2-D body pose detection, followed by the camera motion estimation and 3-D from 2-D pose estimation, which is adapted to take into account the non-standard body pose proportions found in concept skecthes. Finally, we then describe the matching of the 3-D mesh to concept sketches using the sampling importance re-sampling method guided by the matching quality between the projected 3-D body mesh and the concept sketch.

A. 2-D Pose Detection

The problem of 2-D body pose detection has been addressed for inputs consisting of real pictures of humans [1], [9]. For example, Yang and Ramanan [1] describe a method for 2-D body pose detection based on a flexible, non-oriented mixture of parts model, whose parameters are learned with a structured support vector machine solver [10]. This approach produces solid results on several public databases and its main advantage lies in its run-time efficiency. Alternatively, the method proposed by Zuffi et al. [9] is based on an extension of the pictorial structure model and can capture the non-rigid shape deformation of the parts. We have tested these two models on concept sketches and noticed that the former [1] produces more reliable results. Nevertheless, the performance achieved is not ideal mostly because the input images we use are markedly different from the ones used to train the original model [1], which essentially consists of real pictures of humans on relatively cluttered backgrounds. Ideally, we would need to retrain this model for concept sketch images, but we are not aware of databases that contain a sufficient number of annotated concept sketches that could be used to train such model, so we adopt an alternative method, consisting of requesting the user to manually adjust the joint positions, if necessary (see Fig. 2). This is the only user interaction required in our system - all remaining steps below are fully automatic.

B. 3-D Pose and Camera Motion Estimation

The 3-D body pose detection and camera motion estimation from 2-D body pose landmarks is also another problem being intensively studied in computer vision [2], [11], where the main challenge lies in the ambiguities present in this estimation. Salzmann and Urtasun [11] propose an approach based on Gaussian process that requires large amounts of training data from varied viewpoints and body deformation to reliably recover the underlying 3-D pose, but in general it does not generalize well for cases not present in the training set, which is problematic for our application given that the "bodies" present in concept sketches will not be part of any training set. Ramakrishna et al. [2] propose a more robust approach by formulating it as an optimization problem that minimizes the re-projection error by adjusting the 3-D position of the landmarks and the camera motion. This method generalizes better to concept sketches, particularly because the original optimization problem can be extended in two ways: 1) the camera motion can be constrained to a frontal view setup, which is natural for concept sketches; and 2) the anthropometric regularities can be relaxed, which allows our method to adapt better for the non-standard human body proportions usually found in concept sketches. Therefore, we extend the method proposed by Ramakrishna et al. [2].

We consider that the P joints of the 3-D human body skeleton is denoted by $\mathbf{X} = [\mathbf{X}_1^\top, ..., \mathbf{X}_P^\top]^\top \in \mathbb{R}^{3P\times 1}$ [2]. Assuming a weak perspective projection, the 2-D coordinates of the projected 3-D points onto the image plane are produced by:

$$\mathbf{x} = \left(\mathbf{I}_{P\times P} \otimes \begin{bmatrix} S_x & 0 \\ 0 & S_y \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \mathbf{R}\right) \mathbf{X} + \mathbf{t} \otimes \mathbf{1}_{P\times 1}, \quad (1)$$

where $\mathbf{I}_{P\times P}$ is an identity matrix of size $P \times P$, $\mathbf{1}_{P\times 1}$ is a vector of P ones, $\mathbf{x} \in \mathbb{R}^{2P\times 1}$, $\otimes$ represents the Kronecker product operator, $\mathbf{s} = \left(\begin{smallmatrix} S_x & 0 \\ 0 & S_y \end{smallmatrix}\right)$ is a diagonal scale matrix, and $\mathbf{R} \in SO(3)$ and $\mathbf{t} \in \mathbb{R}^2$ are the respective rotation and translation parameters of the camera. Considering that the camera intrinsics are assumed to be known, the goal of the method described in this section is the estimation of the 3-D positions $\mathbf{X}$ and camera motion $\mathcal{C} = \{\mathbf{s}, \mathbf{R}, \mathbf{t}\}$ in (1).

The human body joints are represented by a weighted sum of action dependent basis poses, as follows [2]:

$$\mathbf{X} = \mu^{(\mathbf{X})} + \sum_{i\in I_{\mathbf{B}^*}} \mathbf{b}_i^{(\mathbf{X})} \omega_i^{(\mathbf{X})} \quad (2)$$

where $\{\mathbf{b}_i^{(\mathbf{X})}\}_{i\in I_{\mathbf{B}^*}} \in \mathbf{B}^* \subset \mathcal{B}$ represents the basis poses, $\mu^{(\mathbf{X})} \in \mathbb{R}^{3P\times 1}$ is the mean pose, and $\omega_i^{(\mathbf{X})}$ are the weights assigned to each basis. Note that $\mathcal{B} \in \mathbb{R}^{3P\times(\sum_{i=1}^{N_a} N_b^i)}$ is an overcomplete dictionary of basis components, which is created by concatenating N_b^i bases from N_a actions, $\mathbf{B}^*$ is an optimal subset of $\mathcal{B}$, and $I_{\mathbf{B}^*}$ are the indices of the optimal basis $\mathbf{B}^* \in \mathcal{B}$.

Our proposed optimization, which is an extension of the

original formulation in [2], is denoted by:

$$\begin{aligned} \underset{\Omega,\mathcal{C},I_{\mathbf{B}^*}}{\text{minimize}} \quad & \|\mathbf{x} - (\mathbf{I} \otimes \mathbf{sR})(\mathbf{B}^*\Omega + \mu^{(\mathbf{X})}) - \mathbf{t} \otimes \mathbf{1}\|_2 \\ \text{subject to} \quad & \sum_{\forall(i,j)\in\mathcal{L}} \|\mathbf{X}_i - \mathbf{X}_j\|_2^2 - \sum_{\forall(i,j)\in\mathcal{L}} l_{ij}^2 \geq -\kappa \\ & \sum_{\forall(i,j)\in\mathcal{L}} \|\mathbf{X}_i - \mathbf{X}_j\|_2^2 - \sum_{\forall(i,j)\in\mathcal{L}} l_{ij}^2 \leq +\kappa \\ & \mathbf{B}^* \in \mathcal{B} \\ & \mathbf{R} \text{ is close to identity,} \end{aligned} \tag{3}$$

where the main differences (compared to [2]) are the constraint on the camera rotation matrix $\mathbf{R}$, and the replacement of the original equality constraint (denoted by $\kappa = 0$) by the inequality constraint $\kappa > 0$. The new constraint on the camera rotation matrix, which makes it close to identity, conveys the assumption about the frontal view of the concept sketch and the new inequality constraint reflects the anthropometric irregularities typically observed in concept sketches. Also in (3), $\mathcal{L}$ denotes the set of edges present in the 3-D skeleton model and l_{ij} represents the average length of the limb between nodes i and j of the skeleton.

The optimization in (3) is solved with a matching pursuit algorithm [2], which is an iterative method that, at each step, adds a new basis to $I_{\mathbf{B}^*}$ fixing $\Omega, \mathcal{C}$, and then solve for each of the remaining parameters (Ω and $\mathcal{C}$) assuming the other two fixed. Note that in the original formulation [2], the camera parameters are found by first re-writing the vectors $\mathbf{x}$ and $\mathbf{X}$ in matrix form, as in $x \in \mathbb{R}^{2\times P}$ and $\mathcal{X} \in \mathbb{R}^{3\times P}$. Denoting the mean-centered projections by $\hat{x} = \mathbf{sR}\mathcal{X}$, we can use the singular value decomposition (SVD) to find $\mathbf{M} = \hat{x}\mathcal{X}^\top(\mathcal{X}\mathcal{X}^\top)^{-1} = \mathbf{UDV}^\top$. The scale parameter $\mathbf{s}$ is estimated from the first 2×2 submatrix of $\mathbf{D}$ [2] and the rotation is estimated with [12]:

$$\widehat{\mathbf{R}} = \alpha\mathbf{I}_{3\times 3} + (1-\alpha)\mathbf{UV}^\top, \tag{4}$$

where $\alpha \in [0, 1]$, which is re-adjusted using the SVD of $\widehat{\mathbf{R}} = \widehat{\mathbf{U}}\widehat{\mathbf{D}}\widehat{\mathbf{V}}^\top$, as follows: $\mathbf{R} = \widehat{\mathbf{U}}\widehat{\mathbf{V}}^\top$. This adjusts the estimated rotation matrix to be close to the identity rotation matrix $\mathbf{I}_{3\times 3}$, which expresses the idea of little camera rotation.

C. Matching the 3-D Mesh to the Concept Sketch

Fitting the 3-D body mesh to the input concept sketch requires a search process that estimates the 3-D mesh that visually matches the input data. We propose an optimization method that maximizes the density of edges found in the concept sketch falling inside the area delimited by the silhouette of the projected 3-D mesh, which implicitly assumes that the background of the concept sketch contains a significantly smaller density of edges compared to the foreground, but at the same time allows some clutter in the background. Our approach consists of a sampling importance re-sampling (SIR) method [13], where the training set for estimating the body mesh distribution is automatically generated with the meshes from the open source software package `Makehuman` (http://www.makehuman.org/).

Our SIR method is an iterative process (Fig. 3) consisting of three steps. The first step consists of generating S samples of 3-D body meshes according to

$$\{\mathbf{Y}_t^{(s)}\}_{s=1}^S \sim P(\mathbf{Y}_t|\mathbf{Y}_{t-1}), \tag{5}$$

where $P(\mathbf{Y}_t|\mathbf{Y}_{t-1})$ is a generative model defined below in (9), $\mathbf{Y}_t^{(s)} \in \mathbb{R}^Y$, t indexes the iteration step, and s denotes the sample index given previously generated samples.

The second step transforms the underlying skeleton of each of these samples from a canonical pose (notice the T-shape of the samples in Fig. 3) into the estimated 3-D pose $\mathcal{X}$, defined in Sec. II-B, via inverse kinematics [14] and linear skinning [15]. Specifically, inverse kinematics works by estimating the joint positions of the skeleton of a human model, maintaining the joint constraints, given that the desirable locations for the P joints are given by $\mathbf{X}_1, ..., \mathbf{X}_P$. We use `Havok` [16] for the implementation of the inverse kinematics, which outputs a set of matrices that define the forward kinematic transformations of each joint. Linear skinning [15] defines the changes in the surface of the 3-D mesh as a function of the skeleton joint transformations. That is, given that each mesh vertex can be affected by more than one bone, it is necessary to interpolate the respective joint transformations, and linear skining [15] simply interpolates linearly among the transformation applied to each bone.

The third step computes the weight of each sample with

$$w_t^{(s)} = w_{t-1}^{(s)} P(\text{match}|\mathbf{Y}_t^{(s)}), \tag{6}$$

where

$$P(\text{match}|\mathbf{Y}_t^{(s)}) = (\mathbf{z}_p \cap \mathbf{z}_c)/(\mathbf{z}_p \cup \mathbf{z}_c), \tag{7}$$

$\mathbf{z}_p, \mathbf{z}_c : \Omega \to \{0, 1\}$ represents two binary maps that transforms an index $i \in \Omega$ from image space to one or zero, with $\mathbf{z}_p(i) = 1$ representing that the image index i lies within the boundaries of the projected mesh (and $\mathbf{z}_p(i) = 0$ otherwise), and $\mathbf{z}_c(i) = 1$ denoting that there is an edge detected from the concept sketch image using Canny edge detector [17] (and $\mathbf{z}_c(i) = 0$ otherwise). Note that $P(\text{match}|\mathbf{Y}_t^{(s)})$ in (7) denotes the overlap ratio between the projected mesh binary map $\mathbf{z}_p$ and the edge map $\mathbf{z}_c$ of the concept sketch. The weights in (6) are subsequently re-normalized, and they are equal to $1/S$ in the first iteration.

These three steps are iterated until convergence, and the algorithm returns the result as $\mathbf{Y}^* = \sum_s w_T^{(s)}\mathbf{Y}_T^{(s)}$, where we assume that T is the final iteration step. In this method, the body 3-D mesh $\mathbf{V} \in \mathbb{R}^{3V\times 1}$, where V is the number of vertices in the mesh, is represented by a combination of Y basis meshes, learned with PCA, with $\mathbf{V} = \mu^{(\mathbf{V})} + \sum_{i=1}^Y \mathbf{b}_i^{(\mathbf{V})}\omega_i^{(\mathbf{V})}$. This mesh can be represented by a point in the Y-dimensional space with $\mathbf{Y} = [\omega_1^{(\mathbf{V})}, ..., \omega_Y^{(\mathbf{V})}]^\top$, which allows us to estimate the distribution of meshes with a Gaussian mixture model (GMM) [18], as in:

$$P(\mathbf{Y}|\theta) = \sum_i \pi_i \mathcal{N}(\mathbf{Y}; \theta_i), \tag{8}$$

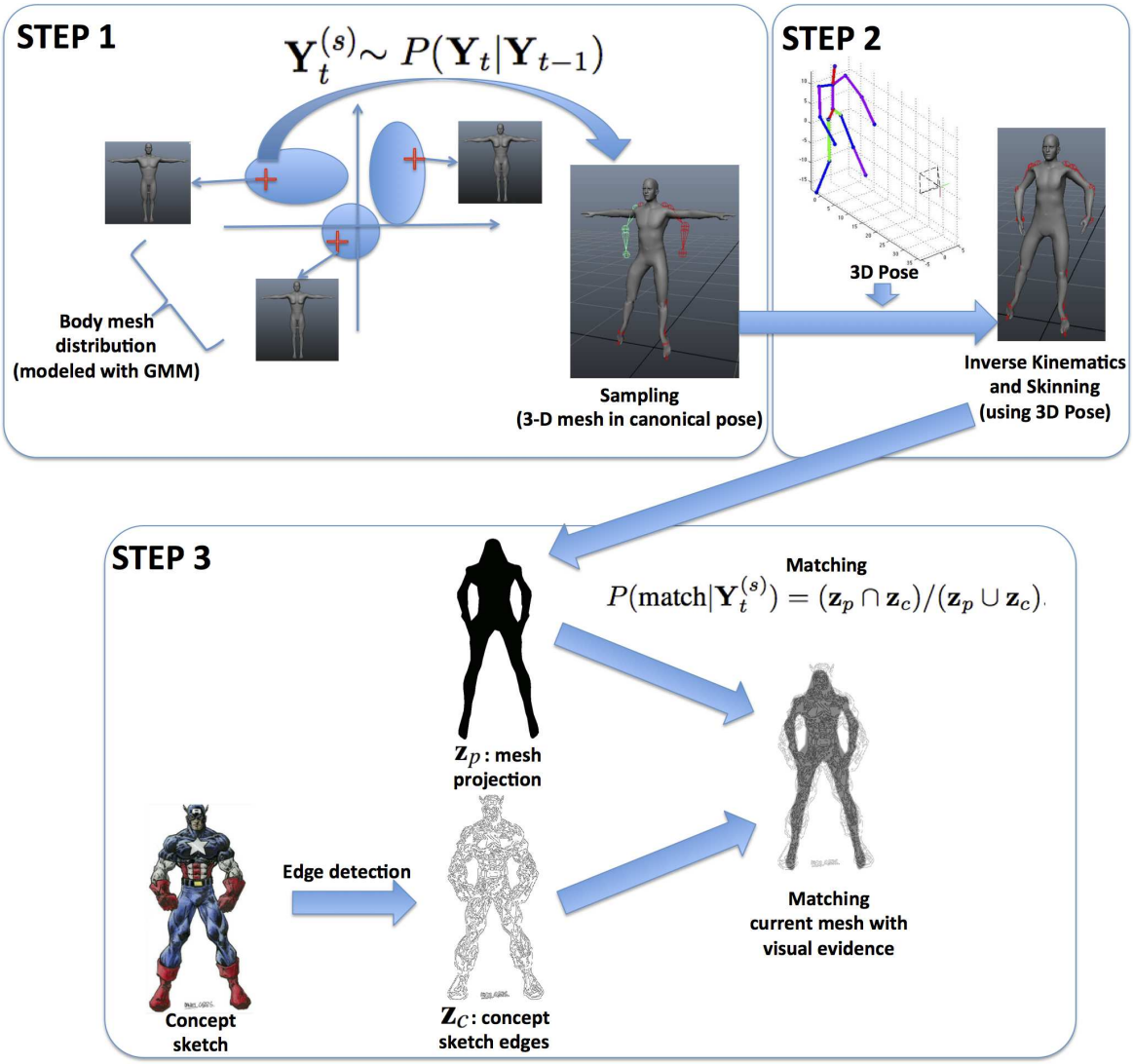

Fig. 3. Three steps of the matching SIR process: 1) generation of S samples of 3-D body meshes; 2) transformation of the underlying skeleton of each of the generated samples from a canonical T-shaped pose into the estimated 3-D pose $\mathcal{X}$; and 3) computation of the weight of each sample, using $P(\text{match}|\mathbf{Y}_t^{(s)})$.

where $\mathcal{N}(.)$ denotes the Gaussian function, π_i is the ownership of the component, and θ_i denotes the Gaussian parameters (mean and covariance). Finally,

$$P(\mathbf{Y}_t|\mathbf{Y}_{t-1}) = \mathcal{N}(\mathbf{Y}_t|\mathbf{Y}_{t-1}, \Sigma)P(\mathbf{Y}_t|\theta), \tag{9}$$

where $\Sigma = \mathbf{I}$ is an identity covariance matrix. This means that new samples are drawn using a distribution that depends on the learned GMM and on their previous location (note that at the first iteration step, the sampling is performed using only $P(\mathbf{Y}|\theta)$).

III. Experiments

For the experiments, we have downloaded seven concept sketches from the web to analyze the performance of our method qualitatively, by visual inspection of the estimated 3-D meshes, and quantitatively by projecting these estimated meshes to form new concept sketches with a known reference 3-D mesh. This allows us to compute the average Euclidean distance (AED), defined by:

$$AED = \frac{1}{V}\sum_{i=1}^{V} \|\mathbf{v}_i^{(r)} - \mathbf{v}_i^{(e)}\|_2, \tag{10}$$

where $\mathbf{v}_i^{(r)}$ denotes the i^{th} vertex of the reference 3-D mesh, $\mathbf{v}_i^{(e)}$ represents the i^{th} vertex of the 3-D mesh estimated by our method, with the mesh being represented by $\mathbf{V} = [\mathbf{v}_1^\top, ..., \mathbf{v}_V^\top]^\top \in \mathbb{R}^{3V\times 1}$, and both meshes normalized by the height of the reference mesh. The concept sketches are: `Captain America`[1], `Black Widow`[2], `The Flash`[3],

[1] http://static.comicvine.com/uploads/scale_medium/8/85629/1868635-captain_america_comic_drawings.jpg

[2] http://advancedgraphics.com/wp-content/uploads/2013/08/1589_BlackWidow_AvengersAssemble_28.jpg

[3] http://www.beloeil-jones.com/tag/the-flash/

Fig. 4. Qualitative results showing the estimated 3-D mesh on a canonical T-shape pose (left of each panel) for seven concept sketches (right of each panel).

`Tomb Raider`[4], `Ironman`[5], `Thor`[6], `Spiderman`[7], and `Terminator`[8].

In all experiments, meshes contain $V = 15000$ vertices, the PCA model of the meshes is trained with 200 samples to keep 96% of the energy of the principal components (which means that $Y = 5$ in our experiments), the number of skeleton joints is $P = 15$, and $\kappa = 8$ and $\alpha = 0.8$ for the optimization in (3). The Gaussian mixture model in (8) is trained with 30 components with the same samples as the ones used to train the PCA model above. Even though our implementation allows us to model male and female populations jointly, we model the model genders independently because we notice that the pose does not provide enough information to return the correct gender accurately.

A. Results

We first present the qualitative results in Fig. 4, where we show seven examples of concept sketches (right on each panel) and their respective estimated 3-D meshes (left on each panel) using our proposed methodology. The quantitative experiments in Fig. 5 take as input the image on the left of each panel, which is obtained by taking the estimated 3-D mesh from the qualitative experiment for each concept sketch, and projecting it to form a new concept sketch. Note that this new sketch has a reference 3-D ground truth mesh, and our method estimates the 3-D mesh shown on the right of each panel, which is used to compute the AED in (10) after normalizing both meshes by the height of the known (reference) 3-D mesh, as mentioned above. The AED result for each concept sketch is displayed on the top of the right image of each panel.

B. Discussions

The qualitative results in Fig. 4 shows that our proposed methodology produces, in general, 3-D meshes that are visually similar to the underlying (but hidden) 3-D model present in the concept sketch. In particular, the method differentiates well models that are very muscular (e.g., Thor and Terminator) from models that are not so muscular, such as Spiderman of

[4]http://francinedelgado.deviantart.com/art/Lara-Croft-Tomb-Raider-9-189389580
[5]http://www.deviantart.com/art/Iron-Man-Modular-Armor-62488829
[6]http://comics.cosmicbooknews.com/content/guardians-galaxy-origin-announced-gamora-ron-lim-concept-art
[7]http://comics.cosmicbooknews.com/content/guardians-galaxy-origin-announced-gamora-ron-lim-concept-art
[8]http://www.comicbookmovie.com/fansites/nailbiter111/news/?a=90782

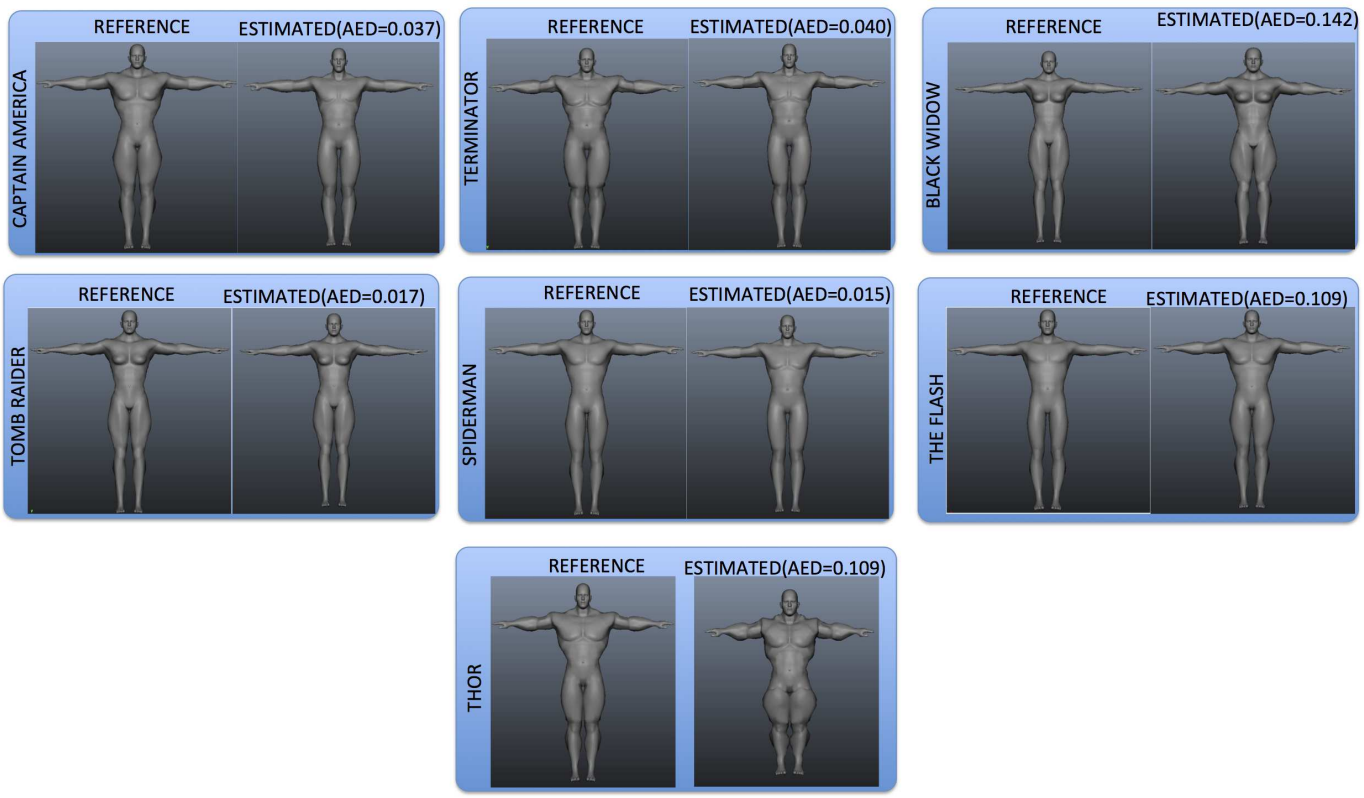

Fig. 5. Quantitative results showing a concept sketch that contains a ground truth reference 3-D mesh, on the left of each panel (obtained by projecting the 3-D mesh estimated using the original concept sketch). On the right of each panel, the new estimation of the the 3-D mesh is shown using the sketch on the left (i.e., not the original sketch) with the corresponding AED, computed with (10).

The Flash. It is also interesting to note that the system is robust to background noise, as clearly see in the models for Black Widow, Tomb Raider and Thor. In particular, the cape present in the Thor sketch produces a large number of edges that could have potentially confused the matching algorithm. The method also seems to be robust to the non-standard human proportions present in all of these sketches. Actually, the fact that the method can generate models with non-standard human proportions is interesting, given that it is trained with meshes produced by `Makehuman`, which in general generates standard human models. This happens because the sampling in the PCA space generates meshes that can deviate reasonably from the training samples. Finally, the quantitative experiment in Fig. 4 shows that in general, the estimated 3-D mesh is close to the reference meshes not only in terms of the AED results, but also in visual terms.

IV. Conclusions

In this paper, we present a methodology for estimating 3-D meshes from concept sketches using minimal user interaction. We believe the qualitative and quantitative results indicate that the proposed methodology has potential to be further investigated. We plan to study this method further, by addressing some of the issues noticed during the experimental evaluation. For instance, we plan to make the 2-D pose detection more effective and less dependent on human interaction. This will be achieved by building our own annotated database of concept sketches and re-training the method developed by Ramakrishna et al. [2]. We also intend to replace the linear skinning process to avoid the typical artifacts seen in our examples [19] (see for instance the skinning in Fig. 3). We also plan to use more sophisticated non-linear dimensionality reduction methods [20] that can lead to more effective matching processes. Finally, we plan to texture the 3-D mesh using the patterns present in the concept sketch.

Acknowledgements

This research was partly funded by the Data 2 Decisions Cooperative Research Centre. G. Carneiro acknowledges the APV (Apoio a Pesquisador Visitante) fellowship received from CNPq-Brazil. R. Ding developed this work while he was with the University of Adelaide.

References

[1] Y. Yang and D. Ramanan, "Articulated human detection with flexible mixtures of parts," in *CVPR*. IEEE, 2011, p. 1.

[2] V. Ramakrishna, T. Kanade, and Y. Sheikh, "Reconstructing 3d human pose from 2d image landmarks," in *ECCV*, 2012, p. 1.

[3] L. Olsen, F. Samavati, M. Sousa, and J. Jorge, "Sketch-based modeling: A survey," *Computers and Graphics*, vol. 33, pp. 85–103, 2009.

[4] Z. Levi and C. Gotsman, "Artisketch: A system for articulated sketch modeling," in *Computer Graphics Forum*, vol. 32, no. 2pt2. Wiley Online Library, 2013, pp. 235–244.

[5] T. Chen, Z. Zhu, A. Shamir, S.-M. Hu, and D. Cohen-Or, "3-sweep: extracting editable objects from a single photo," *ACM TOG*, vol. 32, 2013.
[6] V. Kraevoy, A. Sheffer, and M. van de Panne, "Modeling from contour drawings," in *Proceedings of the 6th Eurographics Symposium on Sketch-Based Interfaces and Modeling*, ser. SBIM '09. New York, NY, USA: ACM, 2009, pp. 37–44. [Online]. Available: http://doi.acm.org/10.1145/1572741.1572749
[7] P. Guan, A. Weiss, A. Balan, and M. Black, "Estimating human shape and pose from a single image," in *Computer Vision, 2009 IEEE 12th International Conference on*, Sept 2009, pp. 1381–1388.
[8] I. K. Kazmi, L. You, X. Yang, X. Jin, and J. J. Zhang, "Efficient sketch-based creation of detailed character models through data-driven mesh deformations," *Computer Animation and Virtual Worlds*, vol. 26, no. 3-4, pp. 469–481, 2015.
[9] S. Zuffi, O. Freifeld, and M. Black, "From pictorial structures to deformable structures," in *CVPR*. IEEE, 2012, p. 1.
[10] I. Tsochantaridis, T. Hofmann, T. Joachims, and Y. Altun, "Support vector machine learning for interdependent and structured output spaces," in *ICML*. ACM, 2004, p. 104.
[11] M. Salzmann and R. Urtasun, "Implicitly constrained gaussian process regression for monocular non-rigid pose estimation," in *NIPS*, 2010, p. 1.
[12] N. J. Higham, *Matrix nearness problems and applications*. University of Manchester Dept. of Mathematics, 1988.
[13] A. Doucet, S. Godsill, and C. Andrieu, "On sequential monte carlo sampling methods for bayesian filtering," *Statistics and computing*, vol. 10, no. 3, pp. 197–208, 2000.
[14] J. M. McCarthy, *Introduction to theoretical kinematics*. MIT press, 1990.
[15] L. Kavan, P.-P. Sloan, and C. O'Sullivan, "Fast and efficient skinning of animated meshes," in *Computer Graphics Forum*, vol. 29, no. 2. Wiley Online Library, 2010, pp. 327–336.
[16] H. Inc., havok API User Guide, 2013.
[17] J. Canny, "A computational approach to edge detection," *IEEE TPAMI*, no. 6, pp. 679–698, 1986.
[18] C. M. Bishop *et al.*, *Pattern recognition and machine learning*. springer New York, 2006, vol. 1.
[19] R. Y. Wang, K. Pulli, and J. Popović, "Real-time enveloping with rotational regression," in *ACM TOG*, vol. 26, 2007.
[20] J. A. Lee and M. Verleysen, *Nonlinear dimensionality reduction*. Springer, 2007.