Knowledge-based extraction of control skeletons for animation

F. Dellas, L. Moccozet, N. Magnenat-Thalmann
MIRALab - University of Geneva, Switzerland

M. Mortara, G. Patanè, M. Spagnuolo, B. Falcidieno
CNR-IMATI - Genova, Italy

Abstract

In this paper we propose a method for the automatic extraction and annotation of the animation control skeleton of virtual humans, which relies on an a-priori knowledge of the human anatomy. The method is based on a segmentation of the virtual human shape into semantically meaningful features, like arms or legs, and on an automatic location and labeling of joints of the control skeleton. The method is particularly relevant for computer animation where the process still largely relies on manual tasks, and especially for virtual characters built on real scanned data. Several examples will show the results obtained with our approach.

1 Introduction

The control skeleton is the major structural model which defines and drives the animation of virtual characters. From an abstract point of view, it is a directed acyclic graph, and more precisely a tree, whose arcs correspond to body segments and whose nodes locate the joints among the body parts. The animation of virtual characters is obtained by defining a mapping between the movement of the control skeleton and the resulting movement and deformation of the $3D$ body surface associated to the skeleton. Several methods exist to enrich the animation of a virtual character; they are based on the control skeleton and levels of sophistication in the deformation model (e.g., surface skinning). We refer the reader to [7] for a survey on the topic. The main advantage of skeleton-based animation is that high-resolution meshes can be animated by a motion computation that applies to the simpler skeleton structure, where the root node handles the global transformation and provides an initial frame for the sequence of transformations to be applied to the other nodes and segments.

The two main steps in the animation pipeline are the *design of the skeleton* and the *surface skinning* or *mapping*. In the first step, the animator has to design the skeleton of the input shape, that is, construct a hierarchy of labeled joints which can be animated using key-frames or pre-recorded sequences from motion capture. In the second step, in order to obtain a successful skeleton-driven deformation, the animator has to deliver an appropriate attachment of the *skin*, i.e. the surface mesh, to the underlying skeleton in order to transmit the appropriate deformation to the various surface segments during the animation. Intermediate data structures could also be used to mimic the presence of muscles. These two steps are typically done with manual interactions, using professional software [2] which allows to define the correspondence between patches of the body surface and segments of the control skeleton. This process is particularly tedious, especially when the model to be articulated is given only as a boundary representation (e.g., an acquired body surface). Few attempts have been proposed to automatically generate the articulated skeleton from a $3D$ surface mesh. Generally speaking, they rely on the computation of some kind of geometric skeleton, derived by shape segmentation or other skeletonization techniques. The main difficulty is to extract the control skeleton from the geometric skeleton, particularly when the topology of the articulated skeleton is predefined, as for virtual humans whose control skeleton has to adhere to the H-ANIM standard [1].

In this paper, we want to push forward the idea that better results can be obtained by taking into account the semantics of the shape at the various stages of the process. The first observation is that methods for automatic extraction of the skeleton are based either on *geometric transformation* (e.g., the medial axis transform) or on *geometric segmentation* (e.g., the fuzzy clustering technique), which do not consider the semantics that the shape segments should portray. In the virtual character domain, indeed, the relevant shape segments are those corresponding to natural segments, such as arms or legs, and none of the traditional geometric segmentation methods is devised to detect these features [5]. In our approach, we exploit the results of a shape segmentation method that is tailored to extract and annotate human body

parts. The second observation is that the *a-priori* knowledge about the shape domain can be exploited to better tune the joint location: reasoning on the functional meaning and anatomy of the physical joints, it is possible to deduce joint re-location rules which can be used to optimize the position of the automatically extracted joints. In our approach, we have experimented that the shape variations correspond to areas where skeletal joints occur. In particular, a curvature-based analysis of the shape at different scales is exploited to infer the location of the physical joints and to drive the definition of additional spinal joints on the control skeleton.

Overview and contribution The main contribution of the paper is the exploitation of knowledge and semantic information in the process of control skeleton extraction. In particular, a shape annotation process is effectively used as an automatic pre-processing of the control skeleton definition, either automatic or manual. To the best of our knowledge, the extraction and adaptation of the control skeleton driven by local semantics-based reasoning is the first attempt in the field and its main expected advantage is to have a direct mapping of the skin to the skeleton. Our approach is constituted of two main steps:

- *semantic shape annotation*: based on a decomposition of the human body surface presented in [13], we derive a set of semantic annotation rules to classify each segments with respect to a set of body parts;
- *control skeleton extraction*: the segmentation and annotation results are combined with a multi-scale curvature analysis, which is applied to body parts, in order to detect the location of joints using the *a-priori* knowledge on human anatomy to select the best candidates.

The shape segmentation method, called *Plumber*, has been firstly presented in [13] and further specialized in [11]. In this paper, the rules for automatically annotating the segmentation in the domain of virtual humans are presented. The method is able to handle real $3D$ input data for variable body shapes in terms of corpulence, postures, or age. The multi-scale curvature analysis, used at the second stage, is based on the method called *Tailor* presented in [12]. Here, we discuss on the experimental study that allowed us to derive rules for detecting and adjusting joints in the segmented body surface. Obviously, since we deal here with semantic features that inherently lack of a precise mathematical formulation, we have to accept a level of heuristics in the definition of the rules and methods for extracting the skeleton. Nonetheless, the shape segmentation and the curvature classification are based on sound mathematical definitions and provide a rich workbench of tools for processing human body models.

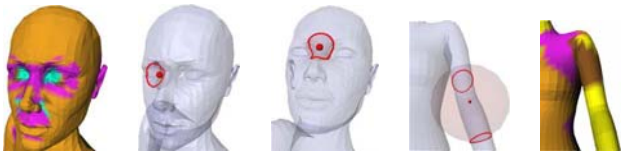
The paper is organized as follows. In Section 2, we present an overview of previous work in the field and explain the difficulty to obtain satisfying results. In Section 3, we briefly sketch the segmentation algorithm and the morphological analysis (detailed in [13] and [12]), that are exploited in our approach. Then, the core of our methodology, i.e. the annotation, the skeleton extraction and refinement, is detailed in Section 4. Finally in Section 5 and 6, we demonstrate our algorithm on representative scanned bodies. The body data were obtained from a full $3D$ body scanner and they exhibit different genders, morphologies and acquisition postures in order to establish the efficiency and robustness of the method.

2 Previous work on segmentation and abstraction of $3D$ shapes and virtual characters.

In the following paragraphs, we briefly summarize the main categories of methods to segment and extract a skeleton of arbitrary shapes. The first ones aim at providing a compact representation of a $3D$ surface that allows to retrieve an approximation of the original shape. The second ones generate segmentations for animation.

Skeleton extraction as a shape analysis process. Most of these approaches are based on the Medial Axis Transform (MAT, for short): here, the skeleton is defined as the union of the centers of the maximal interior spheres which approximate the input surface. Even though the MAT provides a short as well as a skeletal similar representation of a $3D$ shape, it is very sensitive to noise and produces complex skeletons for animation purpose. A Modified MAT, called *Power Crust* [4], has been proposed to reduce this sensibility, but the final skeleton remains very complex to be handled when dealing with real data. Another notable example of topology-driven skeleton is the *Reeb graph* [15], [8], which codes the topology of a given shape $\mathcal{M}$ through the study of the evolution of the iso-contours of a mapping function defined on $\mathcal{M}$. In particular, the Reeb graph provides a segmentation of the shape into slices and the branches of the graph identify connected components of the surface; however, the segmentation is computationally expensive and does not offer a simple control over the scale at which the shape is analyzed.

Shape segmentation for animation. Differently from previous approaches, which give an accurate skeleton to retrieve back the original $3D$ shape, other methods identify primitives to explicitly animate the shape. The extracted segmentation is not necessarily an exact simplification of the shape, but allows to animate the model. Teichmann et



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

Figure 1. (a) Vertex labeling as described in Table 1; in particular concave (b) and saddle-like (c) configurations. On the arm, each sphere having two intersection curves determines a limb vertex (d,e).

Table 1. The main morphological labels among those described in [12].

Feature Type	Color	$\sharp\cap$	Description
Tip	Red	1	Sharp, convex
Pit	Blue	1	Sharp, concave
Mount	Orange	1	Rounded, convex
Dip	Cyan	1	Rounded, concave
Blend	Pink	1	Blend
Limb	Yellow	2	Cylindrical
Joint	Brown	2	Conic
Split	Green	more	–

al. [16] first use the MAT for retrieving the skeleton of animation. From any closed polygonal mesh, they extract a simplified and cleaned MAT with manual intervention to select the main joints on this MAT. The use of a simplified MAT as a control skeleton has been also studied in [19], but did not offer the same simplicity of use because of the skeleton complexity. It is also difficult to use this kind of methods in a standard animation pipeline. In [10], [9], the shape segmentation is used for skeleton driven animation; however, the original input mesh is synthetic and relatively simple. In a different context, vision-based modeling and tracking, the authors of [6] propose spines to represent the shape and topology of 3D objects with a branching axial structure in order to track the limbs motion of moving articulated creatures from videos.

Hybrid approaches. Mixing geometric algorithms tends to increase the robustness and some of the presented approaches are quite successful in providing rough skeletons for simple shapes [16, 18, 10]. One of these recent approaches [18] provides sufficiently clean skeletons for animation, by connecting "domain control points" obtained using geometry shrinking with radial basis functions. These links are made by using *Snakes*, which are also difficult and non-intuitive to tune. It is also more difficult to label and identify joints in this type of skeleton.

Existing methods start from a given $3D$ shape and mainly extract the skeletal structure solely based on the surface properties. The main drawback is that the topology and geometry of the skeleton can not be pre-defined. For Virtual Humans, and particularly for animating Virtual Humans, the expected skeleton must satisfy certain constraints to be directly usable for animation, i.e. it has to adhere to a reference skeleton such as H-ANIM. We also have requirements on the way the skeleton must be placed with respect to the skin surface. Precisely recovering the reference structure is mandatory in order to further apply animation algorithms (skinning, motion control, motion capture, ...) that rely on the assumption that the skin is appropriately attached to the human skeleton. Our objective is therefore to use segmentation algorithms in conjunction with a set of semantic rules based on the knowledge of the skeleton structure and on the human anatomy in order to systematically extract the expected virtual human skeleton from a single $3D$ body shape.

Mortara et al. [11] have shown that multi-scale shape analysis is appropriate for studying and segmenting human body shapes into semantic meaningful components, and they also expose that an implicit skeleton may be extracted. In this paper we define additional rules and processes for precisely controlling the skeleton extraction in order to match the pre-defined structure and constraints.

3 The *Tailor* and *Plumber* methods

The *Tailor* method [12] classifies the vertices of a $3D$ surface represented by a triangle mesh $\mathcal{M}$ according to geometric and morphological descriptors evaluated on neighborhoods of increasing size. The set of neighborhoods associated to each vertex v is defined by a set of spheres $\{S(v,r)\}_r$, centered at v, and whose radii $\{r\}$ represent the scales at which the shape is analyzed. Chosen a scale r, we consider the surface region containing v and delimited by the intersection γ between $\mathcal{M}$ and $S(v,r)$; we discard other regions of intersection between the sphere and the mesh that might occur but do not contain v. The number of connected components of γ gives a qualitative characterization of the shape in a neighborhood of each vertex, and the evolution of the length ratio of the boundary components of γ with respect to the radius of the spheres is used to compute geometric attributes (curvature, concavity/convexity, etc...) to detect specific features such as sharp protrusions or wells, mounts or dips, blends or branching parts (see Figure 1 and Table 1).

Starting from the above-mentioned vertex classification, *Plumber* defines a shape decomposition into connected components that are either *tube* features, identified by re-

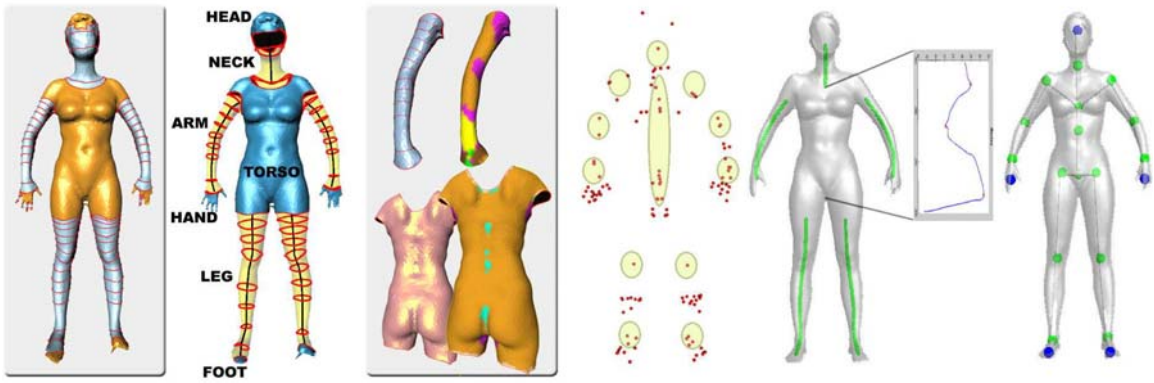


Figure 2. From left to right, pipeline of the proposed approach. The segmentation and morphological analysis phases (in the grey boxes) are already published in [12, 13]. The other steps constitute the original contributions of the paper.

gions which can be described as generalized cylinders (e.g., handle-like and protrusion-like features, together with their concave counterparts), and *blob* regions identified by patches which connect tubular features. The tube construction process starts from *Limb* vertices located by *Tailor* as candidate seed regions of tubular features.

Plumber [14] works in a multi-scale setting (i.e., using a fine-to-coarse strategy), starting with the extraction of small tubes first; the set of radii can be automatically set by uniformly sampling the interval between the minimum edge length and the diagonal of the bounding box of $\mathcal{M}$.

4 The proposed approach

The animation control skeleton is a specialized graph with a pre-defined topology that approximates the human skeleton. In our study, we use H-ANIM as reference description of the control skeleton, which is a widely adopted and standardized model of control skeleton included also in MPEG-4 [1]. Our objective is therefore twofold: we have to extract a skeletal structure from the $3D$ scan data of the body shape that

- precisely reproduces the topology of the H-ANIM skeleton and
- accurately locates the nodes at the natural joints inside the $3D$ shape.

A method for automatically locating the joints must cope with the different form, structure and functionality of body parts as well as with the different postures in which the data have been acquired. Geometric extraction algorithms work under the hypothesis that for tubular shaped parts, like arms or legs, it is reasonable to consider the joint positions on the centerline of the corresponding shape segments. Other joints, as for example vertebra, are located at positions that are only roughly approximated by the trunk centerline. It is straightforward that different strategies must be applied to detect joints, depending on the body part they belong to.

Figure 2 depicts the whole pipeline of control skeleton extraction. The body model is firstly segmented into tubular and non-tubular shaped components using the *Plumber* tool. For each tubular part, the approximated axis and cross section are computed; they will be used as a first skeleton draft on which candidate joints will be projected. Then, as we will detail in Section 4.2, each segment is automatically annotated with a semantic label that indicates which body part it represents. Our experiments have shown that saddle-like or blend local features nicely detect variations on limbs and therefore are useful to detect elbow or keen joints, while small concave regions locate joint useful for the spinal part of the control skeleton. Therefore, each annotated segment is processed using the *Tailor* algorithm to locate the candidate skeletal joints (e.g., spine for the torso, knee and ankle for legs.). The candidate locations are processed further in order to construct the spine curve and to define the final joints. Finally the corresponding control skeleton is constructed. Figure 3 shows the H-ANIM structure and highlights in red the 16 joints and 5 tip joints that our method is able to detect from $3D$ scan data. The various steps are discussed in details in the following Sections.

4.1 Semantic annotation of human body models

Decomposing an object into parts having a precise morphological meaning, such as tubular parts, has a deep impact in the classification of articulated shapes. Such a morpho-

(a) (b)

Figure 3. (a) Detectable set of joints on a scanned body, and their h-anim correspondences: in red detectable joints, in blue tip joints. (b) General spinal shape and detectable joints in green.

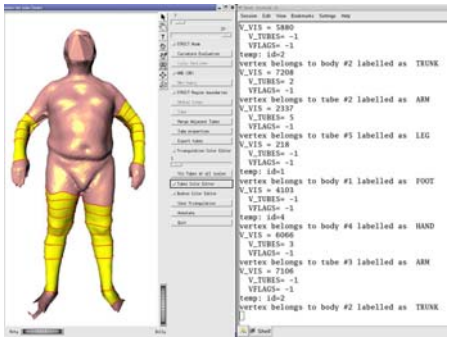


Figure 4. Graphical user interface: on the left, shape segmentation with the corresponding annotation, on the right.

logical segmentation is expressive enough to allow an automatic annotation of components with semantic content, at least in well specified knowledge domains like that of human body models. In fact, while geometric attributes may vary from a model to another, the human body structure is well defined and the basic components are predominantly tubular (e.g., arms, legs, fingers, neck). Therefore, we implemented an automatic *semantic annotator* for human body parts based on the segmentation provided by *Plumber*.

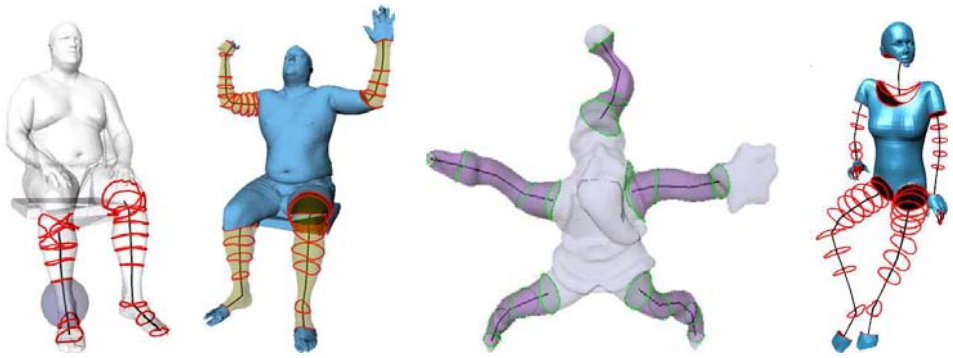
In the following, we are going to describe the algorithm and show the results of the segmentation and annotation on real scan data. The annotation can be defined as a function $\alpha : \mathcal{S} \rightarrow \mathcal{L}$ from the set of segments $\mathcal{S}$ into the set of labels $\mathcal{L}$. In our case, the segments are those given by *Plumber* and they may be either tubular or non-tubular parts; the labels are defined in order to make the annotation exhaustive with respect to the segmentation and the set of labels is chosen as $\mathcal{L} := \{$*trunk, arm, hand, palm,finger, fingertip, leg, foot, neck, head*$\}$.

In general, some of the labels in $\mathcal{S}$ might not appear in the annotation because they have not been identified by the segmentation due to the posture, the poor quality of the scans, or the selection of level of details which do not enable to characterize small features such as fingers. In this last case, the hand segment will be labeled as *hand*, discarding the *palm*, *fingers*, and *fingertip* labels. Conversely, fingers, fingertips, and palm will be instantiated at the expense of *hand*, unless we deduce afterwards that adjacent regions labeled as *palm*, *finger*, and *fingertip* form a *hand*.

The annotator exploits the geometric attributes of parts, computed during the segmentation phase. For tubes, these are the axis length and the maximum, minimum and average length of cross sections, while for blobs the volume is considered. We point out that a tube segment has always two adjacent segments, while a blob segment may be adjacent to one or more tubes; in particular, we will call *cap* a blob segment adjacent to one tube exactly. Given a segmented shape, we define as *shape-graph* the graph whose nodes are the identified segments and the arcs code the adjacency among them. The adjacency relations among segments coded in the shape graph and the a-priory knowledge on human anatomy are exploited through annotation rules of parts, based on geometric attributes of segments. The annotation rules come from the following considerations and imply a sequence of applications:

- the *trunk* is the blob segment of maximum volume [1].
- If the *trunk* is adjacent to 4 tubes, those are *legs* and *arms*; if it is adjacent to 5 tubes, also the *neck* has been segmented.
- If the *neck* has been segmented, it is the tube adjacent to the *trunk*, also adjacent to a cap, having minimum length; the *head* is the cap adjacent to the *neck*.
- Among the four tubes adjacent to the *trunk*, not yet labeled (i.e. except the *neck*, if segmented) *arms* are those having maximum length, maximum section (greater value of maximum section length) and adjacent to a cap, that will be labeled as *foot*.
- The two tubes adjacent to the *trunk* still unlabeled will be annotated as *arms*.

[1] This statement always holds: if the model is undersegmented, the trunk segment will have the maximum volume, at the expense of the arms, legs, and neck.



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

Figure 5. (a-b) Shape segmentation of a human body in different postures; the presence of the seat stops the identification of the legs at the knees. (c) The Klaus' beret is recognized as tubular feature by *Plumber* but cannot be associated to a label in $\mathcal{L}$; (d) segmentation of a seated virtual human. Comparing (a-b) with (d), we conclude that the segmentation provided by *Plumber* is not affected by the posture and the annotation is right if we suppose that the input shape does not contain artificial (e.g., seat, beret) and self-intersecting parts.

- If a cap is adjacent to an *arm*, it will be labeled as *hand*; otherwise, the body segment adjacent to an *arm* (beyond the *trunk*) will be annotated as *palm*, and its adjacent tubes as *fingers*. Finally, caps adjacent to *fingers* will be annotated as *fingertips*.

Results Once the model is annotated, a mouse click over a segment will cause the corresponding label to be printed on the screen. In Figure 4, the graphical user interface of the whole work-flow (morphological analysis, segmentation, and annotation) is shown, side by side with the command shell where the main computation steps are reported by the program. Also, the output of some queries on segment labeling by the user are displayed. In all our test cases, *Plumber* segments at least legs and arms, whatever the posture of the model, provided that limbs do not lay over other body parts or objects (for instance in Figure 5(a)). In particular, as shown in Figure 5(b) *Plumber* generates the leg segments despite the presence of a seat, but they circumstantially stop at the knees. The misleading/inaccurate geometric parameters of the legs put to risk the whole annotation results. We point out that the seat itself raises difficulty, while just the seated posture do not affect the annotation (see Figure 5(d)). Thus, we admit arms and legs in general position but we require that they do not touch other body parts; moreover, for getting a correct annotation we require that no other tubular shaped object touches the body. In this case (e.g., a man holding a stick, or santa Klaus beret as shown in Figure 5(c)), the segmentation would generate tubular parts that do not correspond to any body part. If the requirements just described are met, all the segments will be labeled. To conclude, we remark that the automatic semantic annotation gave the right labeling for 16 models over the 19 in our repository, that is, all except the three shown in Figure 5(a-c) and previously discarded.

4.2 Control skeleton extraction

The segmentation step has identified and annotated the main anatomical parts of the human body: arms, legs, and neck considered as tubular features, and also the torso and extremities, like the tip of hands and feet, and the head. In the second stage, each annotated segment is processed with a multi-scale curvature analysis in order to detect candidate regions for joint locations. *Tailor* provides the tools for the identification of curvature variation zones across scales. In particular, regions whose vertices have a uniform label of type *blends* or *dips* in the *Tailor* terminology (see Section 3) are identified as candidate regions which contain the joints of that segment. For each of these areas, its centroid is defined as the mean of triangles barycenters.

The next step consists in the analysis of the variation of the centroid locations across scales. We compute centroids across the different scales and filter them according to their position invariance. In particular, we filter across all the scales according to the standard deviation to remove undesired centroids, if they exist, which most probably do not correspond to joints. From our experiments on the average centroids displacement across scales, we noticed that they are approximately located at a same position considering a set of bodies, regardless of fatness, size and morphology, which confirm our hypothesis (see Figure 6). Moreover, by taking into account only the blends and concave features we have noticed that irrelevant feature points are easily filtered

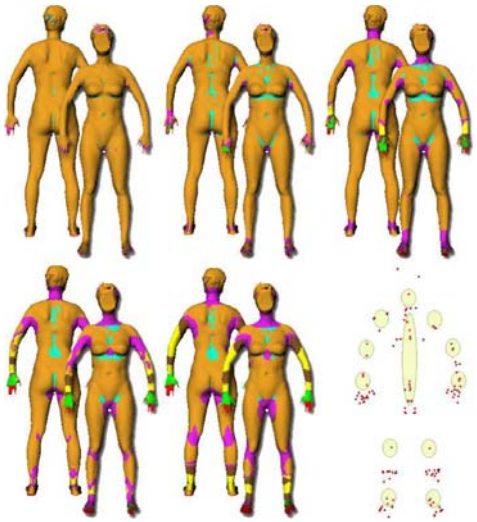


Figure 6. Detected areas on a scanned body (front and back) at all scales; in the most right picture, the centroids corresponding to these regions are located by the circles and they globally correspond to the anatomical joints.

Figure 7. Detected areas on a scanned body with a different morphology: similar features are identified.

out. It is important to underline that the detection of joint areas is consistent and stable with respect to different corpulence and posture variations (see Figure 7), and therefore the initial assumptions seem to be confirmed by the experiments.

The proposed method attempts firstly to detect variations on limbs, therefore, in this first phase we are interested in detecting stable *blend* features (see Figure 1(b-c)) which are identified by one curve of intersection between the sphere and the surface, at the considered scale. Secondly, the methods targets the detection of the spine, characterized by features on the back labeled as *dip*, and which correspond to one curve of concave intersection between the sphere and the surface (see Figure 1(b)). This general processing framework has been specialized to the different type of body segments, as detailed in the following paragraphs.

Case of limbs Each limb part contains only two joints: elbow and wrist for the arms, and ankle and knee for the leg. For each of these parts, we get a certain amount of centroids. Since we know in advance the number of joints we will get, and since we pre-filtered the centroids to avoid outliers, a K-mean algorithm is applied to obtain the desired number of areas for each part. We thus obtain two distinct areas of centroids. For each area, we classify the most probable couple of centroids that could be the expected joints. If centroids are detected at different scales with the smallest displacement between scales, they corresponds to the most plausible candidate to be selected as the joint. Then, for each K-mean area, we obtain the most plausible couple of centroids, and we take simply the average as the joint position. The remaining centroids are re-projected onto the nearest point of the centerline tubes and give finally the exact location of each joint. It is important to notice this joint detection is independent from the posture of the arms and legs.

Case of the torso: knowledge-based extraction of the spine As previously discussed, it is not anatomically realistic to consider the spine as the centerline of the torso as it lays, anatomically, directly under the skin of the back. Also in this case, the multi-scale approach is able to detect relevant feature points, but we process them in a different way in order to extract skeleton joints from the torso. The joints detection on the torso is the key of our method. It induces the way we connect the different joints with the limbs, as it also detects the branching parts between the torso and the limbs. It means that we are able to retrieve the pelvis and the shoulders based on simple anatomic rules on a standard torso posture: when we normally stand up, our spinal has a naturally four curvatures where some of the maxima indicate the locations of arms and hip (see Figure 3(b)).

Depending on the human corpulence fatness and back curvature, not all the spine is detected. We typically obtain the areas where the spinal curvature is maximum and the main direction of the spinal is simply the average of their main axis. We then extrapolate the lowest and the highest point along this axis on the torso part. At this stage, we need to know the nearest points to this axis, which give the points

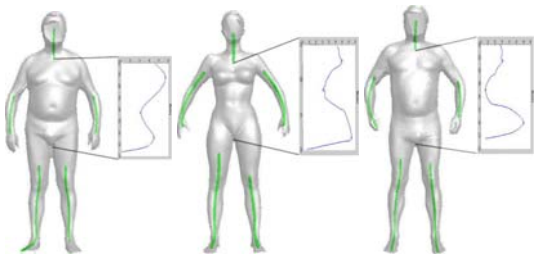


Figure 8. Three scanned bodies with their tube centerlines and associated spine.

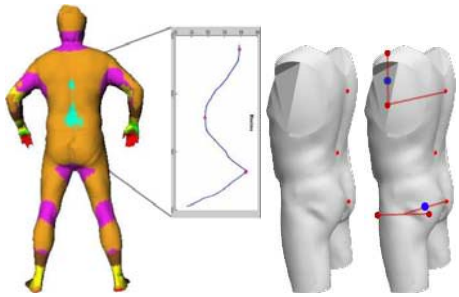


Figure 9. Detected areas on the body back at one of the level range in cyan (protrusion areas); detection of joint height and final position of the joints on the torso using the re-sampled 2D Catmull-Rom spline in red, and deduction of shoulders and hips joints locations

of the mesh that constitute the spinal (L_s the list of points on the scanned mesh belonging to the spinal and I_s the corresponding indexed points). To obtain a uniform representation, we re-sample this set of points using a Catmull-Rom spline in $2D$, using only the length and the depth dimensions. This step is important, because we need a continuous and uniform representation in order to evaluate the two main curvatures of the spine whose maxima correspond to the basin and the height of the shoulder joints.

Then, we take the center of each area that is cut at this height to locate the joint position inside the limbs for the hips and the shoulders. We then have to find the last three maxima on this curve. The first one corresponds to arms height and the last one to basin's height. The middle one is the global minimum of the depth spinal. Usually, we have a dorsal curvature amplitude bigger than the sacral one. But certain pathologies induce the opposite. The curvature amplitude stays in all cases similar, except for major diseases. So as to avoid noisy data and to look for the right maxima, we cut the curve into three main areas according to the standard depth deviation. Then,we cut the curve into three main areas according to the standard depth deviation in order to find the right maximum in each range and to locate precisely the branching joints (see Figure 9).

Case of tips and neck As discussed in Section 3 not all relevant body parts are precisely detected as tubular features: this is the case of tips and neck. The scanned data are in fact usually not very precise to guarantee a good reconstruction of hand or foot fingers, that often are glued together in the reconstructed mesh. Therefore it is not possible to segment each finger independently. The method we follow in this case is simply to compute the mean for each of these parts in order to locate the extremity joints and add them to the joint hierarchy in the last step. For the skull base joint, we compute its position the same way and reproject it to the nearest point of the corresponding *Plumber* tube.

Skeleton integrity check and refinement To build the final skeleton, and annotate it properly, we simply start from the spine and link the hips and shoulders to the nearest joints of the corresponding arms and legs using the Euclidean distance. Since we know which joints have to be retrieved, the labeling is quite immediate. The first resulting skeleton is not necessary symmetric and may contain incoherencies. In order to control and correct these incoherencies as much as possible, we use anatomic proportions based on statistical data from [17, 3]. These data show that the proportions between the long bones of the human skeleton are very stable with a negligible variation: the ratio fibula/femur and ulna/humerus are respectively called crural and brachial index. Their values are respectively equal to 81% and 78% with a standard deviation at 1%. However these statistics reflect the proportions between the real bones' length, and not between the control segments' lengths. There is a slight difference between the control skeleton (located in the shape center for members) and the real anatomical one: the proportions need to be adapted to our type of skeleton. The ratios increase a bit and it can be proved that they remain less than 100%. In mean, we estimate them respectively at 95% and 90%. If the two segments of symmetric bones are not fitting the human proportions, we then apply the mean proportions to each segments to correct the skeleton.

5 Results and discussion

We first compare our results with the expected simplified H-ANIM hierarchy. Figure 11 shows that we are able to detect 16 joints, plus 5 tip joints which are approximations of the concerned parts;the corresponding expected joints into the H-ANIM hierarchy are depicted in Figure 3(a).

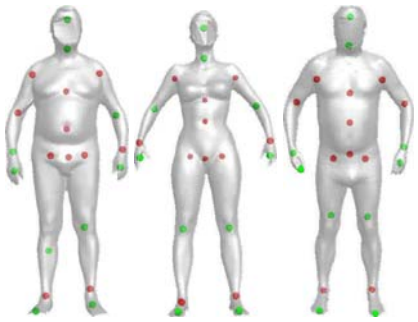


Figure 10. Detected set of projected joints onto tube centerlines on three scanned bodies.

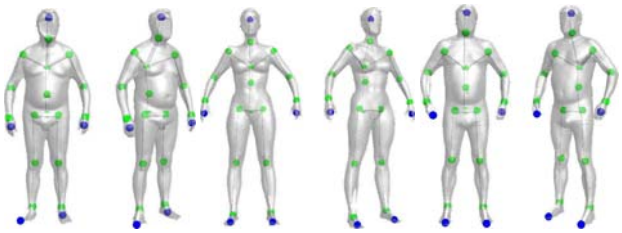


Figure 11. Optimal set of joints on the three scanned bodies, according to statistical proportions.

All our models are about $20K$ vertices closed mesh. The computational time is the following on a P4 3GHz with 1 GB RAM: about 1 minute to process the first segmentation, then around 10 seconds for each tubular members and for each scale of the range and 1 minute for the rest of the analysis and to build the final skeleton. Thus, it takes 5 minutes of pure computation to achieve the overall process for one body. We present below three different scanned bodies with different morphotypes: first with the detected tubular center-lines (see Figure 8) and spinal curve, then the detected joints re-projected onto the tubes centerlines (see Figure 10 and finally the corrected one according to statistical proportions (see Figure 11). Some of the joint locations are not precise: such as the knees or wrists, because small variations are not at the same exact location than bigger ones. Once we re-project the detected joint onto the curves, the error is less important. Another kind of inaccuracy occurs when we get very sharp variation induced by muscles for instance (like the third scanned model) just near bigger variations. In that case, the joint is considered upper than it should be. A robust way to correct the joints position is to check body proportions. This error disappear totally once we check crural and brachial indices and choose the most plausible joints. Our method has been successfully applied to real body scans; the multi-scale approach allows to compensate irregularities of this kind the data. Moreover, the approach is basically robust with respect to variations in morphologies and postures.

- *Morphologies*: The detection works with various morphologies as demonstrated in the results. We expect that it also globally works with various ages (children or elderly) as the basic hypothesis should also apply. This should anyway be tested as fat and muscles volumes and repartition are varying with age. Moreover, the spine detection may need to be adapted to specific ranges of ages. The spine curvature assumption may not exactly apply for children or elderly people. Future works will include the evaluation of the method for different range ages. If required the knowledge-based detection should be adapted and tuned according to specific anatomical configurations.

- *Postures*: for the spine detection, our assumption about the curvature of the spine curve implies that the candidate for scanning should be standing up. Regarding the limbs, we have shown that the detection relies on shape analysis tools that are robust with respect to limb posture. Therefore we can state that there are not any strong constraints about the posture as long as the candidate is standing (which is the usual posture inside a 3D body scanner).

6 Conclusions

The skeleton extraction mainly remains a tedious interactive task that greatly relies on the skills and experience of designers. The rules of thumb that designers are usually applying consists of locating each joint approximately at the level (height) of the joint (relying on visual clues provided by the $3D$ surface of the body shape) and in the center of the $3D$ surface section. Our approach basically implements these rules and the resulting joints are closely located to the ones that a designer would initially select. However, we must state that once designers have approximately located the joints using the previous rules, they need to slightly tune and adjust their locations until the $3D$ skinning is satisfying and accurate. Therefore, our approach provides designers with an initial skeleton configuration that they may further need to tune according to the skinning. We presented in this paper the extracted skeletons on different scanned meshes. We have shown that exploiting a semantic driven segmentation it is possible to automatically generate a 16 joints skeleton for animation for $3D$ human scans, regardless to the body posture.

With respect to the current implementation, significant

improvements are desirable, either by increasing algorithm robustness or speed. The joint location is dependent from the segmentation into body parts; moreover, it can only handle the same amount of joints on separate limbs. When processing a very corpulent body having for instance one arm much more bent than the other, it may happens that the arm segments extension differ much despite the volumetric symmetry of the limbs. This is because the tube radius increases notably in correspondence of the bending zone. Therefore two symmetric tubular parts may not contain the same amount of joints. In this scenario, we have no way to improve the detection locally, but when we rebuild the whole skeleton, we can decide if we missed or built one more joint.

Another notable improvement can also be achieved in the scale range choice for the multi-scale analysis. The interesting scales seem to be always located around a fixed scale and, with further analysis, it could be possible to refine the range so that we may divide the computational time of this step by two. Another alternative is to automatically set the step range to divide into less scales. Nevertheless, we believe that this methodology can be extended if we adapt the pre-defined joints hierarchy. We believe also that realistic creatures could be processed: for instance when scanning figurines for 3D video games. Future works include the extension of the method to various ages (from children to elderly people). A next stage would also consist in attaching the $3D$ surface/skin to the extracted skeleton in order to control the skinning and animate the 3D body shape according to control skeleton motions.

Acknowledgments. This work has been supported by the EC-IST FP6 Network of Excellence "AIM@SHAPE".

References

[1] H-anim specification. http://www.h-Anim.org.

[2] Poser. http://www.e-frontier.com/.

[3] L. Aiello and C. Dean. *An Introduction to Human Evolutionnary Anatomy*. London Academic Press, 1990.

[4] N. Amenta, S. Choi, and R. K. Kolluri. The power crust. In *Proceedings of the sixth ACM symposium on Solid modeling and applications*, pages 249–266. ACM Press, 2001.

[5] M. Attene, S. Katz, M. Mortara, G. Patanè, M. Spagnuolo, and A. Tal. Mesh segmentation - a comparative study. In *SMI '06: Proceedings of the IEEE International Conference on Shape Modeling and Applications 2006 (SMI'06)*, page 7, Washington, DC, USA, 2006. IEEE Computer Society.

[6] G. J. Brostow, I. Essa, D. Steedly, and V. Kwatra. Novel skeletal representation for articulated creatures. In *ECCV04*, pages Vol III: 66–78, 2004.

[7] T. D. Giacomo, H. Kim, L. Moccozet, and N. Magnenat-Thalmann. Control structure and multi-resolution techniques for virtual human representation. In L. D. Floriani and M. Spagnuolo, editors, *Shape Analysis and Structuring*. Springer-Verlag, 2007.

[8] M. Hilaga, Y. Shinagawa, T. Kohmura, and T. L. Kunii. Topology matching for fully automatic similarity estimation of 3d shapes. In *SIGGRAPH '01: Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, pages 203–212, New York, NY, USA, 2001. ACM Press.

[9] S. Katz, G. Leifman, and A. Tal. Mesh segmentation using feature point and core extraction. *The Visual Computer*, 21(8):649–658, 2005.

[10] S. Katz and A. Tal. Hierarchical mesh decomposition using fuzzy clustering and cuts. In *ACM Transactions on Graphics*, volume 22, pages 954–961, 2003.

[11] M. Mortara, G. Patanè, and M. Spagnuolo. From geometric to semantic human body models. *Computers & Graphics*, 30(2):185–196, 2006.

[12] M. Mortara, G. Patanè, M. Spagnuolo, B. Falcidieno, and J. Rossignac. Blowing bubbles for multi-scale analysis and decomposition of triangle meshes. In *Algorithmica*, volume 38, pages 227–248, 2003.

[13] M. Mortara, G. Patanè, M. Spagnuolo, B. Falcidieno, and J. Rossignac. Plumber: a method for a multi-scale decomposition of 3d shapes into tubular primitives and bodies. In *In: Ninth ACM Symposium on Solid Modeling and Applications SM'04*, pages pp.339–344, 2004.

[14] M. Mortara, G. Patanè, M. Spagnuolo, B. Falcidieno, and J. Rossignac. Plumber: A method for a multi-scale decomposition of 3d shapes into tubular primitives and bodies. In *ACM Symposium on Solid Modeling, 2004*. ACM Press, 2004.

[15] G. Reeb. Sur les points singuliers d'une forme de pfaff completement integrable ou d'une fonction numerique. In *Comptes Rendu Acad. Sciences*, pages 847–849. Sciences Park, 1946.

[16] M. Teichmann and S. Teller. Assisted articulation of closed polygonal models. In *Proc. 9th Eurographics Workshop on Animation and Simulation*, Lisbon, Portugal, August 31 - September 1 1998.

[17] L. Testut. *Trait d'anatomie humaine. Tome 1*. Paris, O. Doin, 1899.

[18] M. O. Wan-Chun Ma, Fu-Che Wu. Skeleton extraction of 3d objects with radial basis functions. In G. Elber and V. Shapiro, editors, *Proceedings: Symposium on Shape Modeling International (SMI-03)*, pages 207–215, Seoul, Korea, 2003. ACM.

[19] S. Yoshizawa, A. Belyaev, and H.-P. Seidel. Free-form skeleton-driven mesh deformations. In G. Elber and V. Shapiro, editors, *Proceedings: 8th ACM Symposium on Solid Modeling and Applications (SM-03)*, pages 247–253, Seattle, USA, 2003. ACM.