AN ONTOLOGY OF VIRTUAL HUMANS: INCORPORATING SEMANTICS INTO HUMAN SHAPES

A. García-Rojas, D. Thalmann, F. Vexo[1], L. Moccozet, N. Magnenat-Thalmann[2],
M. Mortara, M. Spagnuolo[3] and M. Gutiérrez[4]
[1] VRlab - EPFL
[2] MIRALab - University of Geneva
[3] IMATI - CNR
[4] ITESM-Campus Toluca

keywords: Human shape reconstruction and synthesis, semantic annotations, ontologies.

Abstract

Most of the efforts concerning graphical representations of humans (Virtual Humans) have been focused on synthesizing geometry for static or animated shapes. The next step is to consider a human body not only as a 3D shape, but as an active semantic entity with features, functionalities, interaction skills, etc. The ontology for Virtual Humans we are defining will provide the "semantic layer" required to reconstruct, stock, retrieve and reuse content and knowledge related to Virtual Humans. The connection between the semantic and graphical data is achieved thanks to an intermediate layer based on anatomical features extracted form morphological shape analysis. The resulting shape descriptors can be used to derive higher-level descriptors from the raw geometric data. High-level descriptors can be used to control human models.

1. Introduction

Virtual Humans, as graphical representations of human beings have a large variety of applications. Within inhabited Virtual Environments, Virtual Humans (VHs) are a key technology that can provide virtual presenters, virtual guides, virtual actors, and be used to show how humans behave in various situations [24].

Creating Virtual Humans is a complex and time consuming task which involves several Computer Science areas: Artificial Intelligence, Computer Graphics, Geometric Modelling, Multimodal Interfaces, etc. In this article we give an overview of the state of the art on analysis and synthesis of human shapes. We present recent advances and underlaying difficulties on the creation of VHs. Our main contribution focuses on proposing a semantics-based method for organizing the various types of data that constitute a Virtual Human. The knowledge related to the synthesis, animation and functionalities of VHs is formally specified in the form of an ontology.

An ontology representation of a virtual human must be closely linked to the associated graphical one. It is particularly required to be able to go from the graphical representation to the ontology -semantic- one: with the analysis of the 3D graphical representation in order to query the 3D models for semantic information. It is also required to be able to go from the ontology description to the graphical representation: with the integration of the semantic descriptors in the modelling and animation process, which means that we need to construct the graphical representation of a virtual human from the semantic descriptors. This is made possible thanks to an intermediate layer of humans shape descriptors (features, landmarks, segments...). Nowadays, many detailed 3D datasets of human bodies are available, and with current scanning technology, new ones are relatively easy to produce. As a result, recent modelling approaches, based on real data and statistical analysis of shapes database, should allow for controlling the synthesis of human shapes with high-level body descriptors.

The next section of this paper is dedicated to the synthesis of human shapes based on real data. These are particularly suited to derive semantic data based on human shape synthesis methods. The third section surveys the requirements for modelling active virtual humans, which means turning virtual humans into entity able to interact inside and with the virtual environment. The fourth section demonstrates that the morphological analysis of human body shapes combined with anthropometric knowledge allows to extract accurate low-level semantic features. Finally we propose our preliminary ontology description for Virtual Humans.

2. Analysis and Synthesis of Human Shapes

2.1. Body shape reconstruction and synthesis

As described by T. Dey in [7], the recent advances in scanning technology [6] let rapidly emerge what he calls “sample based geometric modeling” for digital modeling of physical objects from sample points. The basic idea of these modeling methods relies on the usage of acquisition devices such as 3D scanners for extracting geometric data from a real instance. Because the source of the models is real data, they are suitable for producing realistic looking objects. However, acquisition devices do not provide “ready-to-use” results and post-processes are required in order to obtain an accurate shape. Acquired data are usually noisy, over-sampled and incomplete.

Two strategies are possible to manage the acquisition post-process: either consider any object shape as a soup of triangles and therefore only rely on geometric information such as curvature to drive and control noise removal, re-sampling and hole-filling, or consider that each object is an instance of a family of objects that share similarities and therefore use templates to correct and complete the acquired data.

This first category is general but does not catch the specificities of the reconstructed object. It can miss important features of the object or complete it in an inconsistent manner. The second strategy is already widely used by designers when they interactively model a complex shape: they start from an existing similar one and deform it according to the knowledge they have about this family of objects

The integration of pre-existing knowledge in automatic reconstruction process should greatly improve the accuracy of the resulting instance to capture all the features of the object’s family. Automating this approach requires the extraction of high-level information from the acquired data in order to apply knowledge-based methods. Such those proposed in [15] and [8] for hole-filling. But they are not always relevant and appropriate for extrapolating the missing shape surface at holes for human shape.

Therefore it is clear that an accurate and robust reconstruction of the human shape requires incorporating knowledge of morphology and anthropometry [14] (figure 1). Automating this approach requires the extraction of high-level information from the acquired data in order to apply knowledge-based methods.

Human body shape is a typical example family of articulated physical object: it does not have only one shape but many, corresponding to all the possible postures that the underlying articulated skeleton can take. When acquiring scan data, we only obtain a single static snapshot of the body shape. As such and for many range applications, this static snapshot is not sufficient as it does not capture all the possible degrees of flexibility of the human shape.

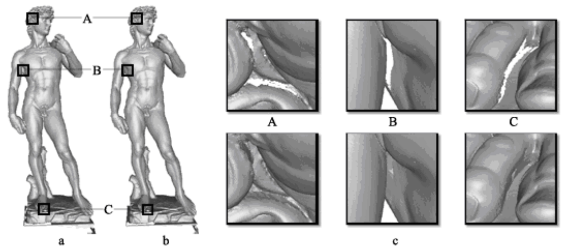

Figure 1. Example of small size holes filling with PolyMender.

To mimic the flexible and dynamic behavior of the human shape, the traditional approach uses skeleton-driven deformations, a classical method for the basic skin deformation that is among the most widely used techniques in 3D character animation. It binds a 3D shape to an articulated control skeleton. Binding information is then used to deform the body shape according to control skeleton motion (figure 2).

Figure 2. Human shape and its associated control articulated skeleton.

Therefore constructing an appropriate and accurate model of human shape should not only reconstruct a 3D shape that matches the body of the scanned volunteer but should also reconstruct all the related information that makes it possible to reproduce the human shape in any of its possible postures. This aspect includes a high degree of integration of the semantic of human body within the construction process.

From this short introduction, we can understand that human shape (re)construction involves:

- Reconstructing more than only the static shape information (and particularly the control animation structure: control skeleton and skin binding).
- Taking into account more than only the available input geometrical information in the reconstruction process.

Algorithms making use of domain knowledge are more accurate because they prevent the surface from erroneously being corrected in under-sampled areas. Examples are template-model based fitting strategies, where scan data is repaired with the geometry from a template surface. These approaches are gaining more and more interest. Examples include the methods proposed by Käahler et al. [17] for faces, using the template-to-data correspondence found using easily identifiable landmarks and Allen et al. [2],[1] who reconstruct bodies using correspondence markers attached to the scanned people.

In [20], Moccozet et al. proposed a full reconstruction pipeline that produces a close approximation of the scanned data of a human body. It is based on fitting a human template model defined in [22] by Seo et al. (which includes both the skin surface and the animation control information) to the scanned data.

In [3], Ben Azouz et al. propose an alternate approach for extracting the variations of the human shape from a 3-D anthropometric database using a volumetric representation of human shapes. More recently, Wang [25] described a feature-based parametrization of the human body for constructing mannequins from body scan data. However, none of these methods integrates the semantic information related to the animation structure required to control and animate the human shape.

The semantic description of human body shape requires defining the common features between human shapes. Alternatively, this should also bring to another open question: what makes each body shape different from each other or in other terms, how far is it possible to characterize the individualization of human shape morphologies?

3. Active human body representation

Once a virtual reconstructed object will be immersed inside a virtual environment, it will have to be able to act as its real counterpart. Therefore it can not be limited to a soup of triangles without higher level information. This is particularly true for virtual humans. Whenever they are included within a virtual environment, they are expected to move and interact with this environment. Obviously, the 3D body shape and even the control animation structure do not include the required level of information for modeling an “active” human shape.

3.1. First level of interaction: accessories

Accessories are objects such as clothes, jewels, hats, glasses that are attached to the human shape. Their motion and animation depends on the motions and animations of the human shape itself. Attaching accessories to a human shape involves locating where they should be placed on the body shape and extracting measurements information in order to fit the accessories to the body shape. At first glance, correctly placing an accessory intuitively relies on a morphological segmentation of the human shape, e.g. a watch will be placed at the border between the hand and the forearm; a shirt will cover the trunk and the arms.

Morphological segmentation of the human shape can also be used to optimize the simulation process when flexible accessories such as clothes are simulated according to the human shape animation. In [5], Cordier et al. automatically resized the garments worn on a 3D body model as the body changes its dimension.

3.2. Second level of interaction: animation

For animating characters there is the assumption that human shape is articulated. Is commonly used the H-Anim [13] standard to define a skeleton to the human shape with the information needed to deform properly the geometry to give a natural looking (see fig. 2). The traditional techniques to synthesize human movements trough the skeleton are kinematics, dynamics or a combination of them [16], [19]. They are used as tools for animating but don’t provide a system for autonomous behavior.

Synthesizing autonomous behavior consist in the techniques to provide to the character the ability of perform tasks, communicate, socialize, take decisions etc. by itself. In this level animations involves many others scientific fields of knowledge as artificial intelligence, psychology, biology, etc. Therefore for this preliminary version of the ontology we limit ourselves to consider that the human body can be animated through its articulations; and the existence of animation sequences and behavioral controllers that drive its movements.

3.3. Third level of interaction: manipulation of objects

The necessity to model interactions between an object and a virtual human appears in most applications of computer animation and simulation. Such applications encompass several domains, as for example: virtual autonomous agents living and working in virtual environments, human factors analysis, training, education, virtual prototyping, and simulation-based design.

Smart Objects are an interesting way to model general agent-object interactions based on objects containing interaction information of various kinds: intrinsic object properties, information on how to interact with it, object behaviors, and also expected agent behaviors. The smart object approach, introduced by Kallmann and Thalmann [18] extends the idea of having a database of interaction information. For each object modelled, we include the functionality of its moving parts and detailed commands describing each desired interaction, by means of a dedicated script language. A feature modelling approach [23] is used to include all desired information in objects.

4. Features extraction and morphological decomposition

In the previous sections we have depicted a scenario where VHs are created starting by an acquisition process; then, the acquired models must undergo post-process reconstruction phases; last, available accessories must be defined, and virtual objects that can be manipulated (and how they can be manipulated) must be specified. We have underlined that in all of this stages the ability to extract semantic information from human body shapes is crucial, and this mainly results in decomposing the shape into meaningful segmented parts or in locating anthropometric landmarks over the body model.

The most common features involved in human shape synthesis are landmarks. Landmarks and segments provide low-level semantic descriptors from which it is possible to derive higher-level ones. Landmarks are points of correspondence on each object of the same kind that match between and within populations [4]. The widely adopted landmarks structure for human shape is the one proposed in the H-ANIM standard [13] description, as shown in figure 3 (top).

The computational methods involved in the extraction of features such as landmarks or shape segmentation must comply with the following constraints:

- Landmarks extraction and morphological segmentation results must be anthropometrically consistent. Extracted features and segmentation must be associated to anthropometric features and segments.
- Landmarks extraction and morphological segmentation results must be consistent and almost invariant from one data set to another.

In figure 3, we show some examples of landmarks extraction based on a multi-scale morphological analysis of the human shape. These features are extracted with a tool called Tailor, which is based on a multi-scale morphological analysis method [21]. This method decomposes the surface into meaningful shape features, like tips, tubular protrusions, concave regions, sharp points, etc.

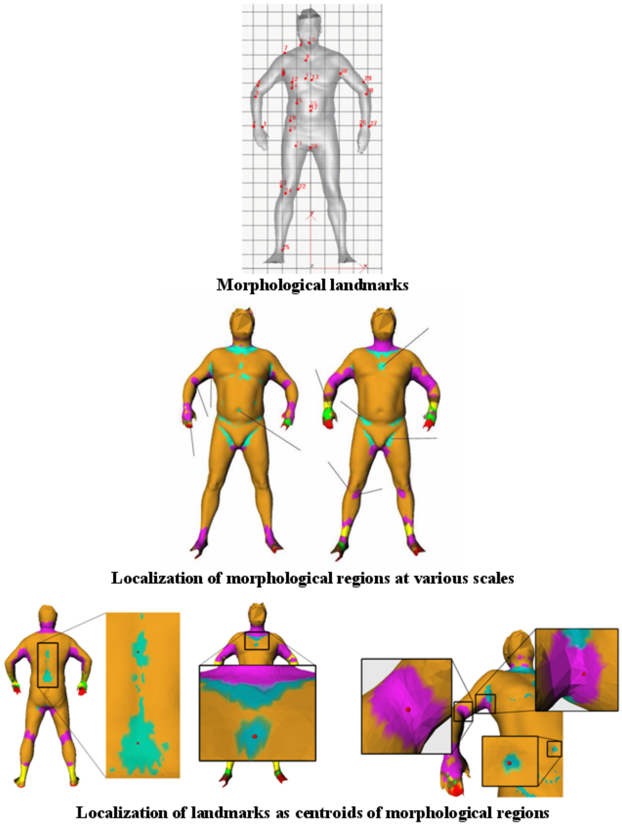

Figure 3. Morphological landmarks.

In figure 4 the shape analysis of two datasets with different morphologies shows that the resulting identified morphological regions are located at anthropometrically meaningful sites and that they are greatly consistent between the two data sets. We expect that this kind of segmentation could also allow to query the shape model for higher level information. For example, the variation of the size of region could also be used to derive some information regarding the morphology. The differences of the region configuration corresponding to the belly-button at scale 4 between the two dataset may allow estimating fat.

In the previous section we have analyzed all the implications, methodologies and algorithms used in the creation of Virtual Humans. As we can se is not a simple work and a lot of information is required, generated and at the end lost because it does not have semantic information associated.

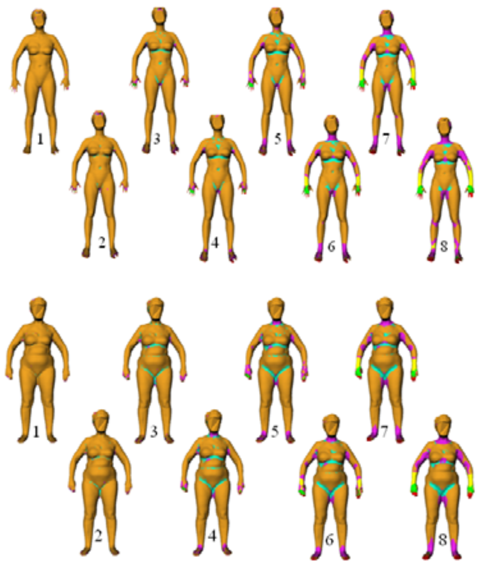

Figure 4. Consistent analysis of body shape among various morphologies.

5. An Ontology for Virtual Humans

According to Gruber [9], an ontology is a formal specification of a shared conceptualization. Virtual Humans are complex entities composed by well defined features, and functionalities. Concepts and techniques related to the creation and exploitation of VHs such as those described in previous sections are shared by the research community. A formal representation refers to the fact that VH representations and their associated semantics shall be both human and machine readable -this is achieved by means of an XML-based representation.

Associating semantic information to the components of a virtual environment has proved to be useful in terms of component reuse, content adaptation, etc. In [12], Gutiérrez et al. defined an object representation based on the semantics and functionality of interactive digital items - virtual objects- within a Virtual Environment (VE), see figure 5.

Every object participating in a VE application is a dynamic entity with multiple visual representations and functionalities. This allows for dynamically scaling and adapting the object's geometry and functions to different scenarios. In [11], the semantic model presented in [12] was complemented with an ontology of objects that allowed for expressing the relationships between interaction devices and virtual entities in a VE. The present work builds upon the acquired experience and focuses on a single type of virtual entity: Virtual Humans.

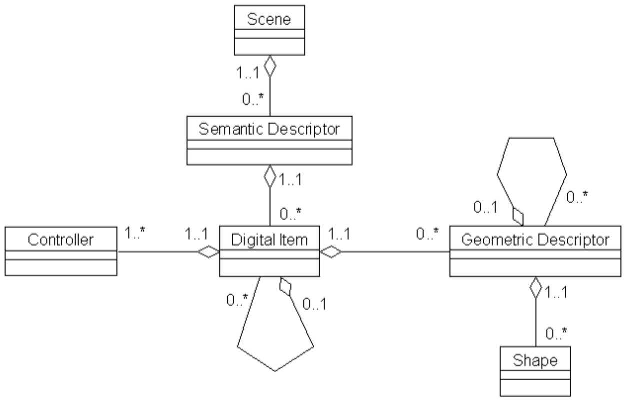

Figure 5. Semantic representation of an interactive virtual environment.

5.1. Developing the Ontology

The development of an ontology usually starts by defining its domain and scope. That is, answer several basic questions known as *competency questions*. Competency questions (CQs) are one of the best ways to determine the scope of the ontology. CQs consist on a list of questions that a knowledge base based on the ontology should be able to answer [10]. The proposed ontology should be able to answer the following categories of competency questions:

Model history
Is this model obtained by editing another model?
What features have been changed on model X?
What tools where involved in the synthesis/modification of this VH?
Who performed the task T on the model X?

Features listing
What is the height of the model?
Is the model male or female?
Is the model European?
What are the features of this model?
Is this model obtained artificially or it represents a real person?
Which VH have a landmark description?
Which are the available structural descriptors for a particular VH?
Which are the standing(seating, walking, .) VH?
How is the body model represented? (a mesh/ a point set/...)
Is the VH complete? (does it have a skeleton/ a hierarchy of body parts/ a set of landmarks attached to it?)

Questions whose answer is a function of low/high level features
Most of the answers to these questions cannot be directly

answered by the ontology -at least not in the current version. Answers will be provided by external algorithms which will take as input the data retrieved through the ontology.
Which are the VH that are fat/slim/short?
Is this VH a child or an adult?
Does it have a long nose?
Does it miss any body part?
Do they have similar anthropomorphic measures (in terms of landmarks?)
Is the model suitable for animation?
How will this VH look like after 20 years? With 20 kg more? With another nose?
Does this model fit this cloth?
What VH do I get if I put the head of VH1 on the body of VH2?

Animation sequences

What model does this animation use?
What are the joints affected by this animation sequence?
Are there any animation sequences lasting more than 1 minute suitable for this VH?
Are there any "running"/"football playing" animation sequences for this kind of VH?
Can the animation sequence X be applied to the VH Y? (in the case of key-frames for skeleton-based animation this would basically depend on the possibility to match the key-frame data to the skeleton of the VH).

Animation algorithms

What are the input and output channels of a particular Behavior controller (animation algorithm)?
What are the models suitable to be animated with this algorithm?
Does this VH have a vision sensor attached?
Can this VH react to sound events in its virtual environment?

Interaction with objects

What capabilities does an object provide?
What are the actions the human can execute on the object?
What are the characteristics of an object? (structure, physical properties, etc.)
How can the object be grasped?

5.2. Ontology components

We have defined a first version of the VH ontology based on the competency questions listed above. The Ontology for Virtual Humans aims at organizing the knowledge and data of three main research topics and applications involving graphical representations of humans:

- Human body modeling and analysis: morphological analysis, measuring similarity, model editing/reconstruction.
- Animation of virtual humans: autonomous or pre-set animation of VH.
- Interaction of virtual humans with virtual objects: virtual -smart- objects that contain the semantic information indicating how interactions between virtual humans and objects are to be carried out.

Figure 6 presents a diagram of the main components of the ontology. The primary class defined in the ontology is the Virtual Human which is a full-body or partial representation of a human being. The model can be synthesized in a variety of ways and can represent a real or a virtual person. VHs are characterized by a set of general attributes (sex, nationality, race, etc.), and structural descriptors.

The structural descriptor is an abstract class that defines the entry points to a variety of descriptors such as animation oriented structures like human skeleton, Smart Object skeleton and others such as topological graphs. This human skeleton definition is based on the H-Anim specification.

The human body consists of a number of segments (such as the forearm, hand and foot) which are connected to each other by joints (such as the elbow, wrist and ankle). In order for an application to animate a humanoid, it needs to obtain access to the joints and alter the joint angles. Each Joint node can contain other Joint nodes, and may also contain a Segment node which describes the body part associated with that joint.

The shape representation of the virtual humans is placed in the abstract class geometry that contains a general description of 3D Shapes used to represent a VH body or parts of it. A set of landmarks can be associated to any geometry as well as textures and other generic information.

Virtual Human animation could be created by means of pre-recorded animation sequences or behavioral controllers. Animation sequences can be applied to one or many VHs. They could be constrained to some requirements to indicate whether this sequence can be applied or not to other VH.

As virtual humans are capable of interacting with objects in their environment, the Smart Object class encloses all those objects that can be manipulated by VHs. This class is constituted by a hierarchical collection of nodes. The hierarchical organization specifies the relations between different Geometry and Attribute Sets composing an object.

6. Conclusions

The current version of the ontology for Virtual Humans is work in progress. As stated before, there are still missing components which are required to fulfill all the needs of a complex and multidisciplinary task such as the creation and use of Virtual Humans. However, we believe this is an important step towards a formal representation of Virtual Humans. The following are some of the main application scenarios where the ontology for Virtual Humans can play an essential role:

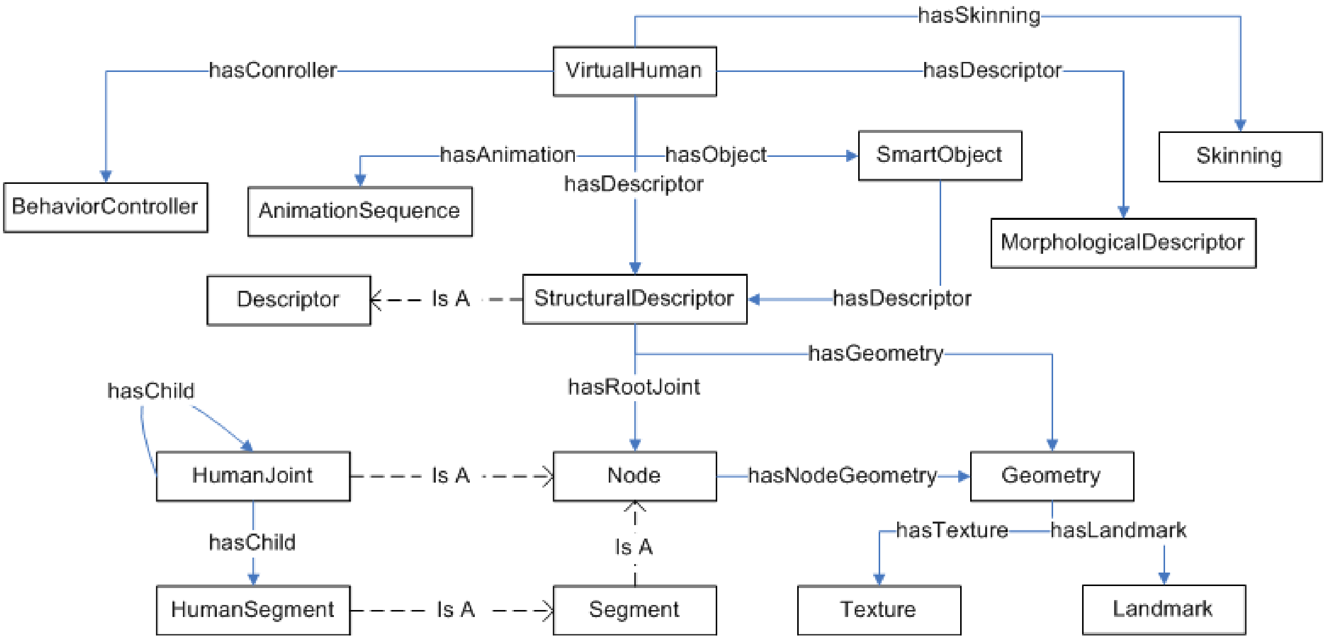

Figure 6. Main components of an Ontology for Virtual Humans.

Virtual Characters data repository: a search engine for retrieving VHs and Smart Objects with particular features/functionalities related to animation. The categories of competency questions that would correspond to this scenario are: Animation sequences, Animation algorithms, Interaction with objects and Features listing, to some extent (features linked to animation such as skeleton, geometry type).

Modeling data repository: a place where a modeler/animator could find VH shapes (whether full or partial bodies) and use them to model new VH, improve or reconstruct existing ones. Categories of competency questions involved: Model history, Features listing (when referring to geometric, anthropomorphic features), questions whose answer is a function of low/high level features (the ones dealing with similarity measures related to the anthropomorphic features).

Shape recognition/extraction/analysis: a knowledge base able to answer competency questions linked to low level features of the VH shape (landmarks, topological graphs, and so on). Main users would include researchers working on algorithms for recognizing features on a shape representing a virtual/real human. Data would be used on ergonomics studies, computer vision algorithms, etc.

In this paper we described some of the main issues to be solved in order to effectively model VHs. We have presented our advances on an ontology-based approach. This is a promising alternative for modeling and managing the knowledge related to Virtual Humans. Taking advantage of an ontology for VHs depends on a two-way process: labeling graphical representations with semantic information and being able to extract semantic information from graphical representations. This will be achieved through shape analysis and segmentation combined with anthropometric knowledge and large sets of acquired data. We are currently focusing our efforts on advancing the state of this research.

Acknowledgments

This work has been supported by the Swiss Federal Office for Education and Science in the framework of the European project IST-AIM@SHAPE Network of Excellence (http://www.aimatshape.net/).

References

[1] B. Allen, B. Curless, and Z. Popović. Exploring the space of human body shapes: Data-driven synthesis under anthropometric control. In *Proceedings of Conference on Digital Human Modeling for Design and Engineering. SEA International.*

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

[3] Z. B. Azouz, M. Rioux, C. Shu, and R. Lepage. Analysis of human shape variation using volumetric techniques. In *Proc. International Conference on Computer Animation and Social Agents*, pages 197 – 206. IEEE Press, July 2004.

[4] F. L. BookStein. Morphmetric Tools for Landmark Data: Geometry and Biology. Cambridge Press, 1997.

[5] F. Cordier, H. Seo, and N. Magnenat-Thalmann. Made-to-measure technologies for online clothing store. *IEEE Computer Graphics and Applications, special issue on Web Graphics*, 23(1):38–48, 2003.

[6] H. A. M. Daanen and G. J. van de Water. Whole body scanners. *Displays*, 19(3):111–120, November 1998.

[7] T. K. Dey. Sample based geometric modeling. CAMS/DIMACS volume on Computer-Aided Design and Manufacturing, eds Dutta, Janardan, Smid. To appear
http://www.cse.ohio-state.edu/ tamaldey/paper/sbgm/sbgm.pdf.

[8] T. K. Dey and S. Goswami. Tight cocone: a water-tight surface reconstructor. In *SM '03: Proceedings of the eighth ACM symposium on Solid modeling and applications*, pages 127–134. ACM Press, 2003.

[9] T. Gruber. The role of a common ontology in achieving sharable, reusable knowledge bases. In *Proceedings of the Second International Conference on Principles of Knowledge Representation and Reasoning*, pages 601–602, 1991.

[10] M. Gruninger and M. Fox. Methodology for the Design and Evaluation of Ontologies. In: Proceedings of the Workshop on Basic Ontological Issues in Knowledge Sharing, IJCAI-95, Montreal.

[11] M. Gutiérrez, D. Thalmann, and F. Vexo. Semantic virtual environments with adaptive multimodal interfaces. In *Proceedings of the 11th International Conference on Multimedia Modelling (MMM2005)*, pages 277–283, 2005.

[12] M. Gutiérrez, F. Vexo, and D. Thalmann. Semantics-based representation of virtual environments. In *International Journal of Computer Applications in Technology (IJCAT) Special issue - "Models and methods for representing and processing shape semantics" (to appear)*, 2005.

[13] H-Anim Working Group. ISO/IEC FCD 19774:200x, Humanoid Animation, Annex B, Feature points for the human body.
http://www.h-anim.org/.

[14] P. R. M. Jones and M. Rioux. Three-dimensional surface anthropometry: Applications to the human body. *Optics and Lasers in Engineering*, 28(2):89–117, September 1998.

[15] T. Ju. Robust repair of polygonal models. *ACM Trans. Graph.*, 23(3):888–895, 2004.

[16] Z. Kacic-Alesic, M. Nordenstam, and D. Bullock. A practical dynamics system. In *Proceedings of the 2003 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, pages 7–16. Eurographics Association, 2003.

[17] K. Kähler, J. Haber, H. Yamauchi, and H.-P. Seidel. Head shop: generating animated head models with anatomical structure. In *SCA '02: Proceedings of the 2002 ACM SIGGRAPH/Eurographics symposium on Computer animation*, pages 55–63. ACM Press, 2002.

[18] M. Kallmann and D. Thalmann. Direct 3d interaction with smart objects. In *VRST '99: Proceedings of the ACM symposium on Virtual reality software and technology*, pages 124–130. ACM Press, 1999.

[19] H. Ko and N. Badler. Animating human locomotion with inverse dynamics. In *IEEE Computer Graphics and Applications*, pages 50–59, 1996.

[20] L. Moccozet, F. Dellas, N. Magnenat-Thalmann, S. Biasotti, M. Mortara, B. Falcidieno, P. Min, and R. Veltkamp. Animatable human body model reconstruction from 3d scan data using templates. In *Proceedings of Workshop on Modelling and Motion Capture Techniques for Virtual Environments, CAPTECH'2004*, pages 73–79.

[21] M. Mortara, G. Patane;, M. Spagnuolo, B. Falcidieno, and J. Rossignac. Blowing bubbles for multi-scale analysis and decomposition of triangle meshes. *Algorithmica*, 38(1):227–248, 2003.

[22] H. Seo and N. Magnenat-Thalmann. An automatic modeling of human bodies from sizing parameters. In *SI3D '03: Proceedings of the 2003 symposium on Interactive 3D graphics*, pages 19–26. ACM Press, 2003.

[23] J. Shah and M. Mantyla. Parametric and Feature-Based CAD/CAM, John Wiley & Sons, inc. 1995, ISBN 0-471-00214-3.

[24] D. Thalmann. The virtual human as a multimodal interface. In *AVI '00: Proceedings of the working conference on Advanced visual interfaces*, pages 14–20. ACM Press, 2000.

[25] C. C. L. Wang. Parameterization and parametric design of mannequins. *Computer-Aided Design*, 37(1):83–98, January 2005.