Geometric Modelling of Moving Virtual Humans: A Survey of Various Practices

Soha Maad, Saida Bouakaz
LIRIS/UFR Informatique Bat 710
Universite Claude Bernard Lyon1
69622 Villeurbanne cedex, France
smaad@bat710.uni-lyon1.fr, bouakaz@bat710.uni-lyon1.fr

Abstract.

This paper surveys various approaches for modelling moving virtual humans for uses in various contexts. Three categories of human body models are identified: 1) the image-based reconstructed models; 2) the synthesized models; and 3) the synthesized and reconstructed models. Models in each of these categories are overviewed and discussed with reference to the context of use, the geometry adopted, the image analysis technique, and the overall performance of the approach. This paper is prepared within the context of the CYBERII project. It serves in drawing an agenda for the development of a corresponding prototype application. The paper is divided into five sections. The first section describes the background motivations and initiatives for conducting this survey and provides a quick overview of the surveyed literature. The second section overviews the prevalent standards for modelling virtual humans, these include the H-Anim and MPEG standards. The third section provides the results of the survey of various virtual human models. The fifth section draws an agenda for the CYBERII project in light of the results of the conducted survey. The paper concludes with a brief summary and a framing of the major findings.

Keywords. rendering, augmented reality, moving virtual human, motion capture, pose estimation, silhouette based modelling, voxel data, MPEG, H-Anim.

1. BACKGROUND

This paper is prepared within the context of the CYBERII project[1]. The CYBERII project aims at simulating, in real-time, the presence of a person (e.g. a TV presenter or a teacher) in a virtual environment. This simulation consists mainly in visualizing the combined scenario, and possibly in providing tools for interaction between the real person, the virtual environment, and the observer (e.g. TV viewer or pupil). The main overall technical requirements are thus a highly realistic visualization, which works in real time. The CYBERII project consists of several parts: 1) acquisition; 2) modelling; 3) rendering; and 4) data management and dstributed parallel computing.

---

1 Http://www710.univ-lyon1.fr/~smaad/CYBERII/

Major considerations in the development of the CYBERII project involves meeting the following requirements:

i. Real Time Operation: The various parts of the CYBERII project handle large amounts of data and many of the computations required for modelling and rendering are rather complex. This calls for refined design and distributed computing.
ii. Realistic Rendering: The augmented scene must seem natural. This poses various requirements on and considerations for the modelling activity.
iii. Non Constrained Motion Capture: The filmed person will not have to wear any specific clothing or sensors.
iv. Compliance with prevalent standards for modelling and design.

Focusing on the modelling part of the CYBERII project, and in particular the modelling and tracking of the moving human, it is important to identify an appropriate geometric representation for the moving virtual human that can support the development and meets its advanced constraints.

The choice of an appropriate model of the moving virtual human is not an easy task due to the following challenges:

i. the diversity of the available approaches for modelling and tracking moving humans
ii. the lack of unified practices in modelling moving virtual humans
iii. the lack of surveys that provides quantitative and qualitative assessments of various approaches for modelling moving virtual humans used in various contexts of applications.
iv. The lack of a complete description of various approaches for modelling moving virtual humans. This type of work is usually conducted at research and industry levels and no complete descriptions of the methodologies and implementations can be accessed or is being appropriately and perfectly depicted.

The above enumerated challenges prompt the need to conduct a survey of various approaches for modelling virtual humans. Given the dimension of the work involved, a special purpose survey is conducted focusing on particular applications with related interests to the CYBERII project. This survey is by no mean a complete one and admits further classification, descriptive details, and enlargement of scope of application. The direct objective of the survey is to support a work plan for the design and implementation involved in the modelling activity of the CYBERII project.

2. STANDARDS FOR MODELLING VIRTUAL HUMANS

This section overviews the two prevalent standards for modelling and animating virtual humans: the H-Anim and the MPEG standard. It overviews also the Virtual Human Markup Language (VHML).

2.1 The H-Anim Standard

The H-Anim 1.1 standard[2] describes the human body in terms of a number of *segments* (forearm, hand, and foot) connected to each other by *joints* (elbow, wrist and ankle). Animating the humanoid involves having access to the joints and subsequently altering the joint angles. Retrieving information about joint limits and segment masses is essential to ensure the right animation. A segment is defined as a mesh of polygons with vertices (Sites) that can be altered in group or individually as needed. As such an H-Anim file contains a set of *Joint* nodes that contains *Segment* nodes that in turn contain *Site* nodes. The standard description is VRML97 compliant. Nesting is represented by a parent-child relationship.

---

[2] www.h-anim.org

A hierarchical description of the human body, according to the H-Anim standard, is depicted in the following figure.

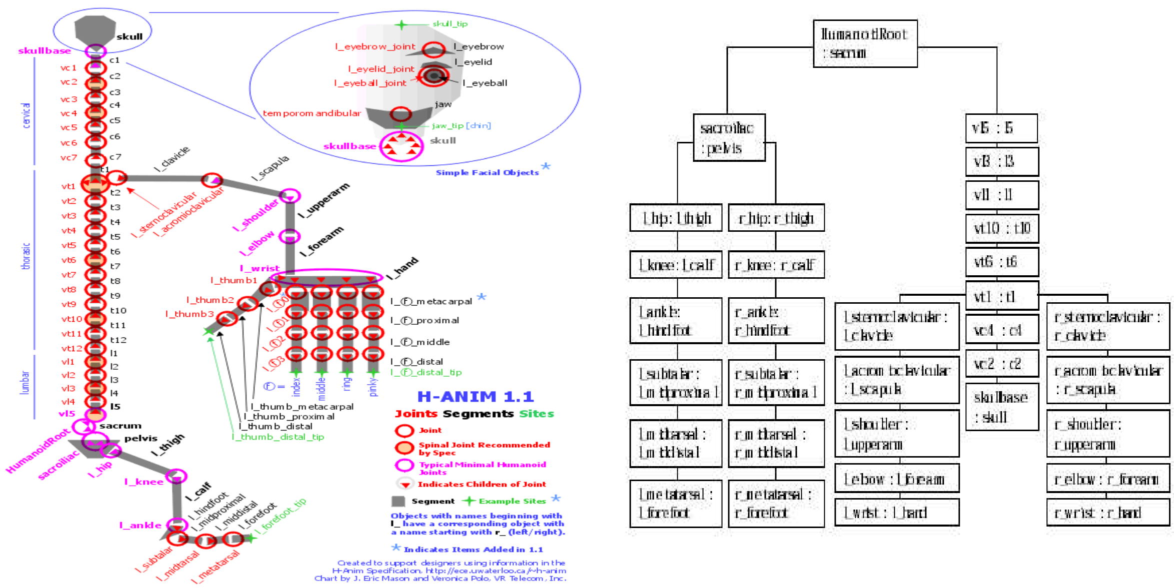

Figure 1. The H-Anim description of a human body:
the hierarchical model and its counterpart tree-like structure

As depicted in the above figure, the body is typically built as a series of nested Joints, each of which may have a Segment associated with it. The root node has associated with it the ***Sacrum*** segment. This root node has two joints as children: the joints ***Lumbar 5*** and the Sacroiliac. In turn the Sacroiliac joint has two joints as children: the ***left hip*** and the ***right hip*** joints. Looking at the head we can see that the **skullbase** joint has seven joint children attached to it to describe the head: the left and right ***eyebrows,*** the left and right ***eyelids***, the left and right ***eyballs***, and the ***temporum andibular***. To each node is associated a segment. For instance, the sacroiliac node has associated to it the pelvis segment.

2.2 The MPEG Standard

MPEG-4[3] [1] defines the FBA (Facial and Body Animation) object to describe the geometry of the virtual character and animate it [2].

[3] MPEG-4 is an ISO/IEC standard developed by MPEG (Moving Picture Experts Group). It provides standard and unified ways to: represent units of audiovisual contents, called media objects ; describe the composition of these objects to create compound media objects that form audio visual scenes; multiplex and synchronize the data associated with media objects, to transport them over network channels, providing appropriate QoS; and - interact with the audiovisual scene on the receiver s end.

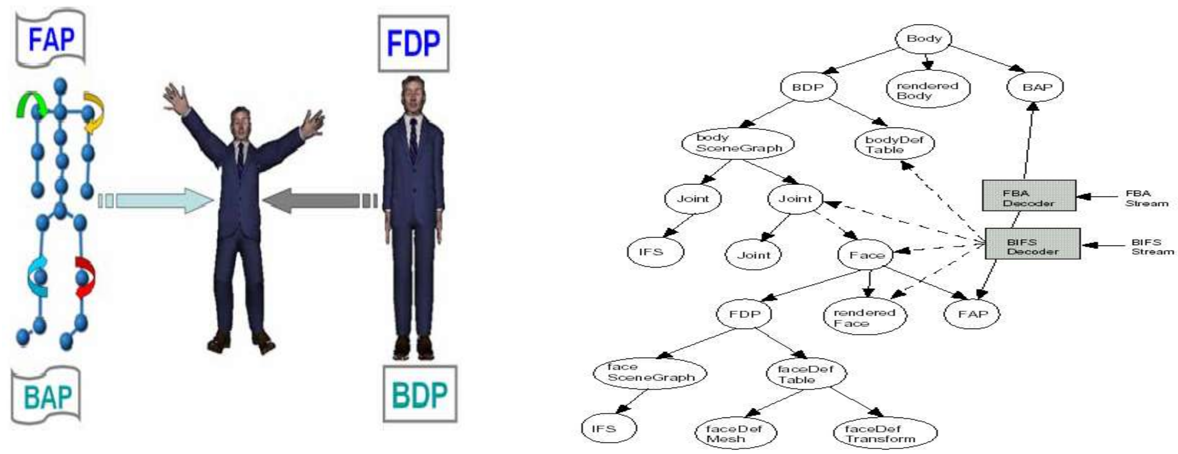

Figure 2. Illustration and schematic view of the FBA object

The FBA specification defines two sets of parameters: The first one specifies the geometry of the FBA model: FDPs (Face Definition Parameters) and BDPs (Body Definition Parameters). The second set defines the animation of the face and body: FAPs (Face Animation Parameters) and BAPs (Body Animation Parameters). The MPEG-4 BDPs used to define the body of the virtual human, are directly related to the VRML/Web3D H-Anim 1.1 specification.
The MPEG-4 BAP (Body Animation Parameters) are used for the synthesis of body movements. In order for the tool to animate the virtual human, it obtains access to the H-Anim joints and alters the orientation angles (yaw, roll, pitch) to match those defined in a BAP stream. A BAP stream can contain up to 296 parameters describing the topology of the body skeleton.
The FBA scene graph [2] has a Body node as its root. The Body node contains three children, BDP, BAP and renderedBody. The BDP node encapsulates the downloaded geometry of the body, together with the skeleton joint hierarchy, surface geometry and the deformation tables. It has two children: bodySceneGraph and a bodyDefTable list. The bodySceneGraph contains a hierarchy of body joints, together with the attached surface and texture nodes.

2.3 The VHML

The Virtual Human Markup Language[4] is designed to accommodate the various aspects of Human-Computer Interaction with regards to Facial Animation, Body Animation, Dialogue Manager interaction, Text to Speech production, Emotional Representation plus Hyper and Multi Media information.

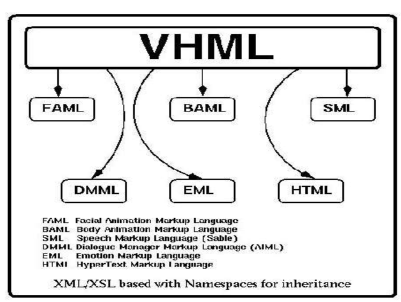

Figure 3. The various component of the Virtual Human Markup Language

---

4 http://www.vhml.org/

3. THE SURVEYED MODELS

This section classifies the human body models in three categories and overviews various models in each of these categories. These categories are: 1) the image-based reconstructed models; 2) the synthesized models; and 3) the synthesized and reconstructed models.

Several considerations can be addressed in modeling virtual humans these include:
- The level of automation of the modelling activity (manual vs automatic initialization and generation of the model)
- The degree of realism in the rendering
- The timeliness of the rendering
- The portability of the reconstruction and rendering framework
- The potential for interactivity with the virtual human model

These considerations are addressed to various extent in the modelling approaches overviewed in this survey.

Whereas other surveys such as the one described in [3] focused on the modelling and synthesis of realistic human motion, the current survey focuses on the geometry of the models and their potential use in combination with an associated image analysis technique. For instance, the survey conducted in [3] classifies the techniques for modelling and synthesis of realistic human motion into three main groups: kinematic animation, dynamic simulation and motion capture.

3.1 Image-based reconstructed models

This section overviews two techniques for the geometric reconstruction of a human body model out of the analysis of captured images of real humans in motion. The first technique is based on a parametric motion segmentation approach, while the second technique is based on a voxel based reconstruction.

3.1.1 Motion segmentation

The context of use

With the aim of meeting the challenges of automatic acquisition and initialization of articulated models from monocular video without any prior knowledge of shape and kinematic structure, Krahnstoever (et al, 2003) proposed a framework for 3D human reconstruction and tracking. This framework was developed for use in the context of human computer interaction where articulated shape models have to be acquired from unknown users for subsequent limb tracking.

The image analysis technique

The framework proposed in [4] involves the use of a Bayesian motion segmentation to extract and initialize articulated models from visual data. Image sequences are decomposed into rigid components that can undergo parametric motion. The relative motion of these components is used to obtain joint information. The resulting components are assembled into an articulated kinematic model which is then used for visual tracking eliminating the need for manual initialization or adaptation. The efficacy of the method is demonstrated on synthetic as well as natural image sequences. The following figure illustrates the results of the motion segmentation analysis technique in recovering the 3D geometry of a human body.

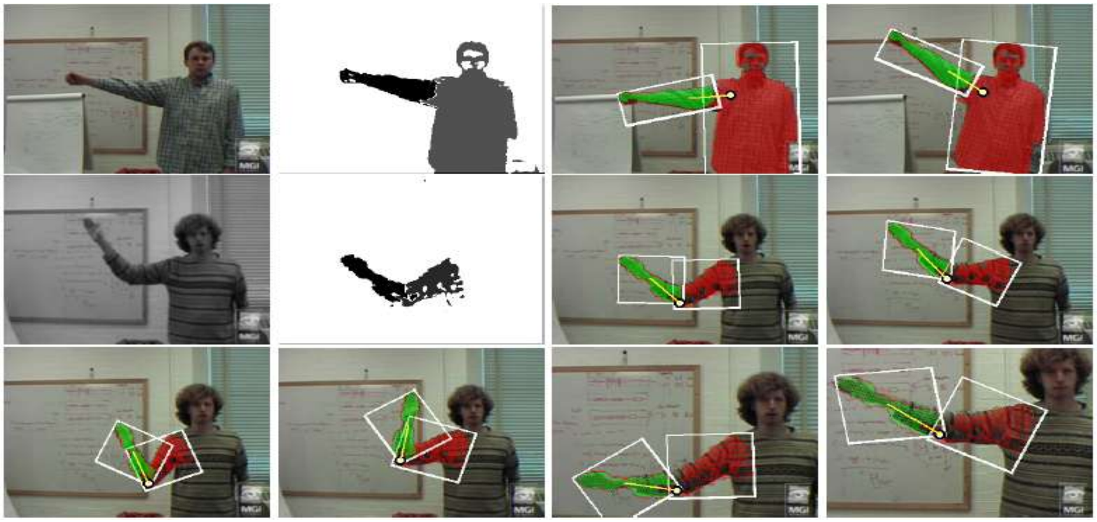

Figure 4. Motion Segmentation approach for 3D human reconstruction

Discussion

The developer of the approach described in [4] acknowledges the limitation of his approach in competing with detailed hand crafted models. He argued that it just offers the potential of gaining further insight into the domain of articulated motion capture, analysis and synthesis.

3.1.2 Voxel based reconstruction

The context of use

With the aim of meeting the requirements of the highly challenging TV industry applications (on-line presentations), of the game industry applications (where the player is rendered into the game), and of virtual education (where the student visits a virtual place), an attempt for real-time reconstruction and insertion into virtual world of human actors was presented in [5].

The image analysis technique

The 3D reconstruction technique of the virtual human adopted in [5] involves the following steps: 1) multi-camera image capturing from different viewpoints; 2) silhouette extraction for images taken by the various cameras from different viewpoints; 3) the use of the visual hull to estimate the shape of the human actor from the silhouettes corresponding to the multi camera viewpoints; 4) voxel reconstruction by projecting the silhouettes on an n plane depth; 5) the logical anding of the various silhouettes corresponding to the various viewpoints in one projection plane; 6) the use of the marching cube algorithm to produce a triangulated surface from the voxel data.

Figure 5. The silhouette reconstruction corresponding to the four camera viewpoints and thefinal rendering from applying marching cube algorithm on voxel data

Discussion

Whereas the paper has demonstrated the feasibility of capturing and inserting in real time an actor in a virtual world without any prior consideration or initialization (such as the use of intrusive trackers, environment or material specification, or an a priori shape estimation), the realism of the visual rendering of the virtual human is still below the expected requirements of the targeted industries' applications. Hence the need to support the approach presented in [5] with an appropriate geometric modelling of the virtual human actor.

3.2 Synthesized models

In this section we overview two approaches for synthesizing a virtual human. The first approach is used in animating human motion and the second approach has a general use in applications that needs realistic rendering such as virtual art and virtual museum.

3.2.1 A synthesized model for animating human motion

The context

The model overviewed in this section was used to animate men and women performing three dynamic athletic behaviors: running, bicycling, and vaulting. The human models for men and women are used in conjunction with control algorithms to generate visually realistic human athletics [6].

The geometry

The human models used in [6] to animate the dynamic behaviors were constructed from rigid links connected by rotary joints with one, two or three degrees of freedom. The dynamic models were derived from the graphical models shown in Figure 6 by computing the mass and moment of inertia of each body part using algorithms for computing the moment of inertia of a polygonal object of uniform density and density data measured from cadavers. The models has several controlled degrees of freedom. Each internal joint of the model has a very simple muscle model, a torque source, that allows a control algorithm for motion to apply a torque between the two links that form the joint. Using this model to animate the dynamic athletics behaviors of men and women, the following degrees of freedom are assigned to the model depending on the type of athletic behavior: the gymnast has 15 body segments and a total of 30 controlled degrees of freedom; the runner has 17 body segments and 30 controlled degrees of freedom (two-part feet with a one degree of freedom joint at the ball of the foot and only one degree of freedom at the ankle); the bicyclist has 15 body segments and 22 controlled degrees of freedom (only one degree of freedom at the neck, hips, and ankles). In the figure shown below, the directions of the arrows indicates the positive direction of rotation for each degree of freedom.

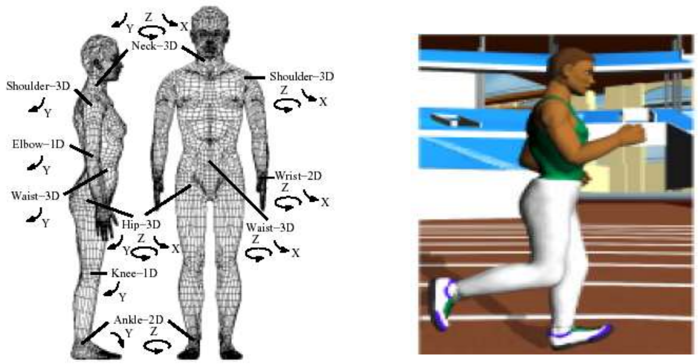

Figure 6. A Polygonal Model used in conjunction with control algorithms to produce animated human athletics

Discussion

Realism in rendering a moving virtual human is not only a matter of exact shape modelling but also of appropriate motion modelling. Research in three fields is relevant to the problem of animating human motion: robotics, biomechanics, and computer graphics.

As argued in [6] people are skilled at perceiving the subtle details of human motion. We can, for example, often identify friends by the style of their walk when they are still too far away to be recognizable otherwise. As such, realistic human modelling depends on three factors: the use of a correct kinematics and dynamics modelling algorithm; the design of a natural looking virtual human actor; and the ease of application of the kinematics and dynamics modelling algorithm to animate the virtual actor.

3.2.2 A Synthesized model for realistic rendering

The context of use

The model overviewed in this section was used in [7] for real time character modeling and animation. The ultimate objective in [7] was to improve the realism of the rendered virtual actor by leveraging the combination of various geometric representations to model the human body.

The geometry

The model used in [7] is depicted in the following figure that shows the merge of two types of geometric representations for modelling virtual humans: implicit surfaces and triangle meshes. Implicit surfaces are used to represent organic shapes, while triangle meshes are versatile and can model any kind of shape at different levels of detail. Moreover, triangle meshes can be rendered in real time using the graphic hardware. As argued in [7], mixing the two types of geometric representation can potentially accelerate and improve collisions, and the skinning of skeletal models.

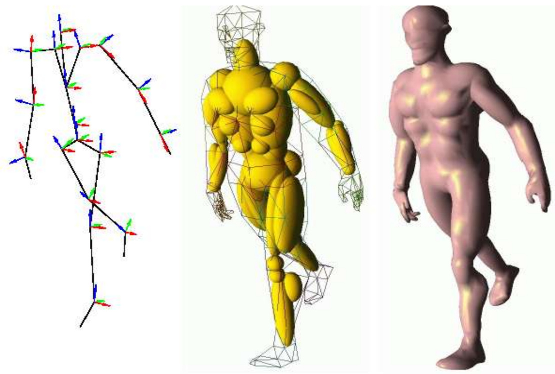

Figure 7. Illustration of the approach for mixing triangle meshes and implicit surfaces for realistic human modelling

Discussion

Several factors should be taken into consideration to create realistic human models. One important factor is the real looking appearance and texture of the human body. A realistic rendering may benefit from the combination of several type of geometric representations of a human body. For instance, the approach in [7] benefits from mixing two types of geometric representations for modelling virtual human actors: triangle meshes and implicit surfaces. As argued in [7] triangle meshes provide a fast and versatile geometric representation of surfaces that can be rendered in real time with existing hardware. Implicit surfaces lend themselves for the creation of smooth organic shapes. Combining the two geometric representations would results in an added value for the final rendering. This gives insight on exploring the combination of various geometric representations to model virtual human actors.

3.3 Synthesized and reconstructed models

This section describes four types of human body models that were used in combination with some image analysis techniques to provide a reconstruction and motion capture of real humans in movement. The models surveyed in this category of synthesized and reconstructed models include: the layered model, the twist based model, the skeleton based model, and the self occluding model.

The survey describes the context of use, the model geometry, the underlying standard, the associated image analysis technique and it provides a short discussion about its efficiency of use.

3.3.1 The layered model

The context-of-use

The layered human body model was used in [8] and [9] for tracking and modeling people from video sequences for the purpose of applications such as animation, surveillance and sports medicine.

The model geometry

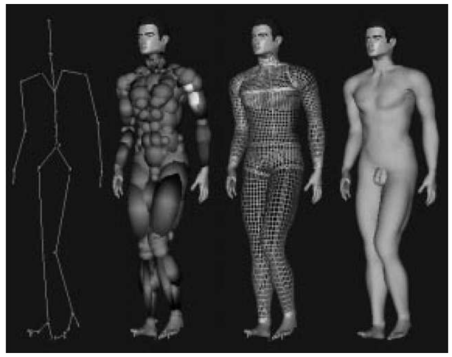

Figure 8. The layered human body model

The human body model used in [8] and [9] is depicted above. It incorporates a multi layered approach for constructing and animating realistic human bodies. The first layer is a skeleton that is a connected set of segments, corresponding to limbs and joints. A joint is the intersection of two segments, which means it is a skeleton point where the limb linked to that point may move. Smooth Implicit surfaces, so-called ellipsoidal meta-balls, form the second layer. They are used to simulate the gross behavior of bone, muscle, and fat tissue; they are attached to the skeleton and arranged in an anatomically-based approximation. The third layer, a polygonal skin surface, is constructed via Bspline patches over control points generated by a ray casting method .

The underlying standard

The layered model is conform to the H-Anim standards as it can be viewed as consisting of joints, segments, and sites.

The associated image analysis technique

As used in [8] and [9], the model provided an a priori information about the shape, and the allowable motions of the human body. A least squares framework was used to fit the body model to tracked images from different camera positions. To initialize the process, the user simply clicks on the approximate location of a few keypoints in one image pair. The recovered shape and motion parameters can then be used to reconstruct the original motion, to display it from a different viewpoint or to make other animation models mimic the subject's actions.

Discussion

As claimed in [8] and [9], the key advantage of the layered methodology is that once the layered character is constructed, only the underlying skeleton need be scripted for animation; consistent yet expressive shape deformations are generated automatically. The image analysis technique is not fully automated and requires an initial human intervention to initialize the matching process.

3.3.2 The Twist-based model

The context-of-use

The twist-based model was used in [10] for motion capture techniques with the aim of finding widespread applications grouped in three classes. The first class of application involves interaction with a virtual world or driving an animated avatar in a video game or for computer graphics character animation. The second class of application involves classifying and recognizing people, gestures or motions (surveillance systems, intelligent environments), and developing advanced user interfaces (sign language translation, gesture driven control, gait, or pose recognition). The third class of applications include personalized sports training, choreography, and clinical studies of orthopedic patients.

The model geometry

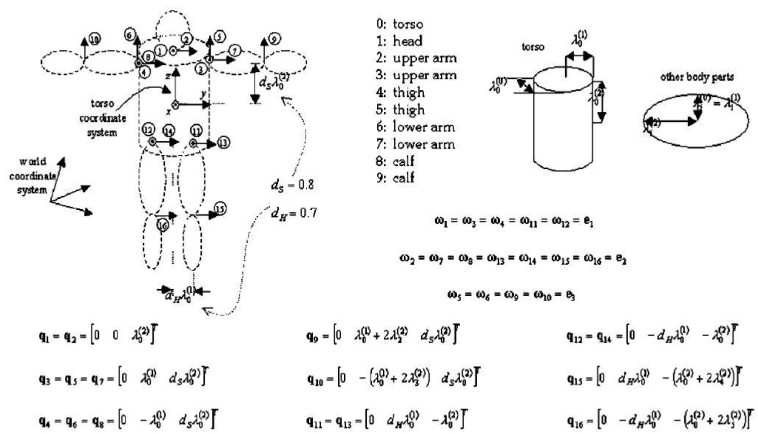

Figure 9. The twist-based model

The human body model used in [10] is described using the twists framework developed in the robotics community (Murray et al., 1993). It consists of ellipsoids and cylinders and is described using the twists framework resulting in a non-redundant set of model parameters.
The twist based model is conceived as an articulated body model that consists of five open kinematic chains: torso-head, torso-left arm, torso-right arm, torso-left leg and torsoright leg. Sixteen axes of rotation are modeled in different joints. Two in the neck, three in each shoulder, two in each hip and one in each elbow and knee. The torso-centered coordinate system is taken as the reference and a range of allowed values is set for each angle. For example, the rotation in the knee can go from 0 to 180 degrees the knee cannot bend forward. The rotations about these axes (relative to the torso) are modeled using exponential maps.

The underlying standard

The model is conform to the H-Anim standard as it can be viewed as consisting of segments linked by joints.

The associated image analysis technique

Model acquisition starts with a simple body part localization procedure based on template fitting and growing, which uses prior knowledge of average body part shapes and dimensions. The initial model is then refined using a Bayesian network that imposes human body proportions onto the body part size estimates. The tracker is an extended Kalman filter that estimates model parameters based on the measurements made on labeled voxel data. A voxel labeling procedure that handles large frame-to-frame displacements was designed resulting in very robust tracking performance.

Discussion
According to the evaluation conducted in [10], it is shown that the twist-based model combined with the proposed image analysis technique perform very reliably on sequences that include different types of motion such as walking, sitting, dancing, running and jumping and people of very different body sizes. However, the model creation process and initialization can not be considered as a fully automated process.

3.3.3 Skeleton based aligned model

The context-of-use
The skeleton based aligned model is used in [12] with the aim of developing highly accurate avatars of humans with a high level of realism for applications in engineering and entertainment, including areas such as computer animated movies, computer game development of interactive virtual environments and tele-presence.

The model geometry

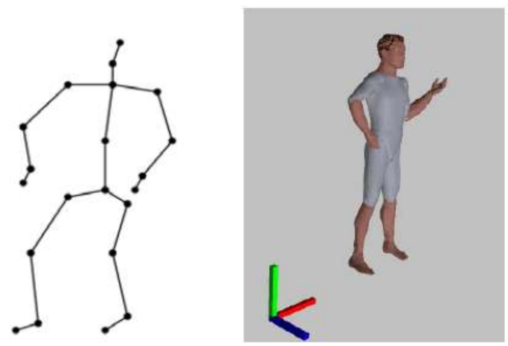

Figure 10. The skeleton model and the correspondingly aligned model

The skeleton based aligned model used in [12] consists of two parts, a skeleton and a reference model, that are aligned together following image analysis techniques. The skeleton description is simple being composed of segments and nodes arranged to fit an articulated human model.

The underlying standard
The model is also conform to the H-Anim standard as the underlying skeleton description is being composed of joints and segments.

The associated image analysis technique
The steps for automatic creation of 3D avatars, as described in [12] include: pose estimation, skeleton fitting, body part segmentation, geometry construction and coloring. The resulting avatar can be animated and included into interactive environments.
The body pose estimation begins with the calculation of the visual hull from the available input images and camera parameters from the acquisition stage. Next, the process of skeleton fitting to the space occupied by the input subject is performed. For this purpose, the volume is divided into slices parallel to the XZ, YZ and XY planes. Using a fast algorithm for 2D image moment calculation, a set of parameters such as the area, centroid, orientation and eccentricity are computed and used to segment the reconstructed volume into different body parts. A set of lines is then adjusted by least squares fit to the cloud of centroids that belong to the underlying skeleton. A general synthetic human model used as a reference in the body geometry estimation stage is then aligned to match the pose of the captured subject. In addition, a set of key features that allows

dividing the input body images in body parts are extracted from the boundaries between the homogeneously connected 3D regions obtained from the visual hull. The body geometry estimation stage obtains a set of images from the aligned synthetic 3D reference model analogous to the one available from the captured subject. Next, a silhouette-finding algorithm is applied to all input images. In combination with the body parts division information extracted in the body pose estimation stage, a 2D mapping function is established between similar body parts of each <synthetic model, captured subject> pair of input silhouettes. This function allows establishing for each view a one-to-one correspondence between the known 3D points in the aligned synthetic reference model and the 3D points of the subject under reconstruction. A set of 2D displacements are then generated for each view and transformed into 3D displacements while merging the available views. Finally, the recovered avatar is colored using the color information available from the captured input images.

Discussion

As claimed in [12] the use of the skeleton based model and the above described technique can generate animated 3D avatars in near-real time with very limited user intervention. Whereas emphasis is on the realistic vision of the 3D virtual human avatar, no concrete measurements were obtained on the ability to track human motion.

4.3.4 Self occluding model

The context-of-use

The self occluding model described in [13] was used in monocular 3D body tracking.

The model geometry

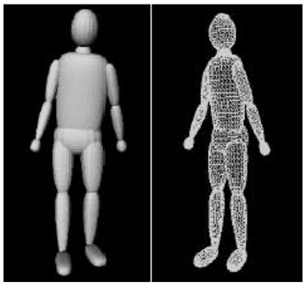

Figure 11. The self occluding model

The self occluding model used in [13] consists of kinematic skeletons of articulated joints controlled by angular joint parameters covered by flesh built from superquadric ellipsoids with additional tapering and bending parameters. A typical model has about 30 joint parameters xa; 8 internal proportion parameters xi encoding the positions of the hip, clavicle and skull tip joints; and 9 deformable shape parameters for each body part, gathered into a vector xd. A complete model can be encoded as a single large parameter vector x = (xa; xd; xi).

The underlying standard

The self occluding model seems to be conform to the H-Anim standard as it can be conceived as composed of joints and segments.

The associated image analysis technique

Monocular 3D human body tracking using the self occluding model described in [13] is based on optimizing a robust model-image matching cost metric combining robustly extracted edges, flow and motion boundaries, subject to 3D joint limits, nonself- intersection constraints, and model priors. Optimization is performed using Covariance Scaled Sampling, and a high-dimensional search strategy.

Discussion

The use of the self occluding model promised good results for monocular images. However, other type and styles of captured image data are not tested.

4.4 A Comparative Study

This paper surveyed 8 approaches for modelling virtual human actors. The models were classified in three categories: 1) the image-based reconstructed models; 2) the synthesized models; and 3) the synthesized and reconstructed models. It is difficult to establish common benchmarks for comparing the various surveyed approaches for modelling virtual humans as each approach attempts to meet the challenges of modelling moving virtual humans from a given perspective. For instance, the image-based reconstructed models, surveyed in this paper, takes into consideration the real time constraint and the fully automated modelling activity while declining the importance of the realism of the rendering of the virtual human character. The synthesized models do not take into a major account the problem of tracking a real virtual human. The synthesized and reconstructed models are rather sophisticated and may fail to meet a real time constraint due to the complication of the modelling activity hence the escalating level of computation. Another inconvenience of the latter category of the surveyed models is that these models may suffer from the need for human intervention in the initialization process thus jeopardizing the full automation of the modelling activity.

It is difficult to identify a "**best working model**" as every model surveyed in this paper serves a specific objective and aims at a particular targeted enhancement. There exists no single model that can meet all constraints including the real time, the realistic rendering, the high level of interactivity, the portability, and the full automation. Further research and practical work is needed to support this survey with concrete results and firm decisions.

4. FUTURE WORKPLAN

It is obvious that the survey presented in this paper is rather an informative one and cannot deliver actual and concrete metrics for the assessment of the use of various geometries for human models combined with various image analysis techniques. However, this survey motivates the following work plan for undertaking the modelling part of the CYBERII project. This work plan involves the development of an exploratory environment for various moving human models associated with various image analysis techniques to obtain a realistic insertion of a moving human in a virtual world. This exploratory environment involves:

i. ***A medium for user interaction.*** This consists of a menu bar with various options for setting the desired functionality (choice of human model, image analysis techniques).

ii. ***A scrolling image display part.*** This part display the sequences of captured images corresponding to the various positions of the cameras used to film the moving human.

iii. ***A renderer for the virtual human in motion.*** The virtual human model could be initially predefined with floating parameters to be fixed later. The model could alternatively evolve from the results of the image analysis without any a priori specification or parametrization. An option for editing a model script is also available provided that the script follow a predefined format understandable by the renderer.

iv. ***A renderer for the scene surrounding the moving human.*** This could be synthesized or reconstructed from captured images.

v. ***A textual image analysis display.*** This shows the status / results of the analysis of the captured images following the application of the chosen image analysis technique.

vi. ***An image display.*** This provides a visual display of the result of the applied image processing techniques on selected captured images

The main objective of the proposed exploratory environment is the recording of metrics assessing the extent of which the real time development constraint can be met and to support the survey conducted in this paper with more concrete results.
This exploratory environment can be best implemented using a portable and modular object oriented programming language with 3D and advanced imaging application interfaces such as JAVA.

A supporting basis for the proposed exploratory environment is the development of the XML based 3D Moving Human (X-3DMH) description. The X-3DMH script would be rendered dynamically in a 3D viewer. This dynamic rendering is updated following the results of the analysis of captured images of a real human. The X-3DMH script is not confined solely to the description of the moving human skeleton and geometry but includes as well the description of the characteristics of the surrounding scene (light, texture, and material). As such the script would consist of 3 nodes:

1. the Human node conform to the H-anim 1.1 Specification for a standard humanoid
2. the texture node describing the texture applied to the humanoid
3. the environment node describing the lighting of the surrounding scene

The general form of the script would look like:

```
<X3DMH>
        <HUMAN>
                <JOINT>
                        &lt;SEGMENT&gt;
                                <SITE>
                                <\SITE>
                        &lt;\SEGMENT&gt;
                <\JOINT>
        <\HUMAN>
        <TEXTURE>
            .
            .
        <\TEXTURE>
        <ENVIRONMENT>
            .
            .
        <\ENVIRONMENT>
<\X3DMH>
```

5. CONCLUSION

This paper provided a survey of various models of moving virtual humans used in the context of augmented reality applications. The survey was preceded by a background introduction to its motivating initiatives and an overview of prevalent standards for modelling and animating virtual humans. The models were classified in three broad categories. Eight models lying in one of these categories were considered and analyzed. A work plan for the CYBERII modelling part is drawn in light of the results of the survey. This work plan involves the development of an exploratory environment of various human motion models that can be combined with various image analysis techniques. The aim is to identify a ***best working human body model*** that can be used in combination with an ***efficient image analysis technique*** based on ***well established metrics*** that can assess the level of attainability of the objectives of the CYBERII project, namely: realistic and real time rendering.

ACKNOWLEDGEMENTS

The author is indebted to many members of the LIRIS lab for relevant ideas, in particular to Bernard Peroche, Jean Claude Ihiel and Irwan Guillou for valuable discussions and comments on the CYBERII project. The author would like to thank Yannick Perret for providing information about his special purpose modelling language for description of models used for multi-views tracking in video sequences. The work of Yannick Perret was quite inspiring for the technical development of the CYBERII project mentioned in this paper. The author thanks INRIA for sponsoring her postdoctoral work at LIRIS Lab.

REFERENCES

[1] Daniel Thalmann and Frederic Vexo, MPEG-4 Character Animation, Virtual Reality Laboratory, Swiss Federal Institute of technology CH -1015 Lausanne, Switzerland
[2] Tolga K. Capin, Eric Petajan, Joern Ostermann: Efficient Modeling of Virtual Humans in MPEG-4. IEEE International Conference on Multimedia and Expo (II) 2000: 1103-
[3] Guowei Wu, A Survey of Modelling and Synthesis of Realistic Human Motions, Department of Computing Science University of Alberta Edmonton, Alberta, December 2003
[4] N. Krahnstoever , M. Yeasin , R. Sharma, Automatic acquisition and initialization of articulated models, Machine Vision and Applications, v.14 n.4, p.218-228, September 2003
[5] Jean-Marc Hasenfratz, Marc Lapierre, Jean-Dominique Gascuel, Edmond Boyer, Real-Time Capture, Reconstruction and Insertion into Virtual World of Human Actors VVG-2003, Vision, Video and Graphics , 2003.
[6] Jessica K. Hodgins Wayne L.Wooten David C. Brogan James F. O Brien, Animating Human Athletics, Proceedings of the 22nd annual conference on Computer graphics and interactive techniques, 1995.
[7] Antoine LECLERCQ, Samir AKKOUCHE, Eric GALIN, Mixing Triangle Meshes and Implicit Surfaces in Character Animation, Computer Animation and Simulation 2001Proceeding , sep. 2001
[8] P. Fua, R. Plankers, and D. Thalmann, Realistic Human Body Modeling. In Fifth International Symposium onthe 3-D Analysis of Human Movement, Chattanooga, TN, July

1998.
[9] R. Plankers and P. Fua, Tracking and Modeling People in Video Sequences , International Journal of Computer Vision and Image Understanding, 81(3):285--302, March 2001
[10] IVANA MIKI , MOHAN TRIVEDI, EDWARD HUNTER, PAMELA COSMAN, Human Body Model Acquisition and Tracking Using Voxel Data, International Journal of Computer Vision 53(3), 199 223, 2003.
[11] Murray, R., Li, Z., and Sastry, S. 1993. A mathematical introduction to robotic manipulation, CRC Press.
[12] M. Villa-Uriol, M. Sainz, F. Kuester, N. Bagherzadeh, Automatic creation of three-dimensional avatars, Electronic Imaging 2003, California, USA 2003.
[13] C. Sminchisescu and B. Triggs. Covariance Scaled Sampling for Monocular 3D Body Tracking CVPR'01.