Realistic Human Body Modeling

P. Fua, R. Plänkers, and D. Thalmann
Computer Graphics Lab (LIG)
EPFL
CH-1015 Lausanne, Switzerland

1 Introduction

Synthetic modeling of human bodies and the simulation of motion is a longstanding problem in animation and much work is involved before a near-realistic performance can be achieved. At present, it takes an experienced designer a very long time to build a complete and realistic model that closely resembles a specific person. Our ultimate goal is to automate the process: Eventually the whole task should be performed quickly by an operator who is not necessarily an experienced graphics designer. We should be able to invite a visitor to our laboratory, make him walk in front of a set of cameras, and quickly produce a realistic animation of himself.

We first present the models we use for body modeling. Our layered representation of the human body is very effective for animation purposes: Once the layered character is constructed, only the underlying skeleton need be scripted for animation.

We then show that the motion of this skeleton can be recovered from video sequences. We concentrate on a video-based approach because of its comparatively lower cost and better control of the dynamic nature of the process.

2 Models

In this section, we first describe the complete model that we use for animation purposes and then introduce a simplified model that we use to derive its position from image data.

2.1 Complete Animation Model

Generally, virtual humans bodies are structured as articulated bodies defined by a skeleton. When an animator specifies an animation sequence, he defines the motion using this skeleton.

A skeleton 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. Motion control methods (MCMs) specify how an actor is animated and may be characterized according to the type of information it privileges when animating the Virtual Human [Thalmann and Thalmann, 1991]. For example, in a keyframe system for an articulated body, the privileged information to be manipulated is the angle. In a forward dynamics-based system, the privileged information is a set of forces and torques; of course, in solving the dynamic equations, joint angles are also obtained in such a system, but they are considered as derived information. In fact, any MCM eventually has to deal with geometric information (typically joint angles), but only geometric MCMs explicitly privilege this information at the level of animation control. The nature of privileged information for the motion control of actors falls into three categories: geometric, physical and behavioral, giving rise to three corresponding categories of MCMs. Once the motion of the skeleton is designed, the realism of motion needs to be improved not only from the joint point-of-view, but also in relation to the deformations of bodies during animation. The body's inherent complexity makes things very difficult: A great many different materials that have no homogeneous behavior, from bones to muscles to fat tissues, come into play.

Since the overall appearance of a human body is very much influenced by its internal muscle structures, the layered model is the most promising for realistic human animation. 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.

Earlier Approaches: We first became involved in this field in 1987, when we proposed a polygonal surface model to animate Marilyn Monroe [Thalmann and Thalmann, 1987]. We also introduced the concept of Joint-dependent Local Deformation(JLD) operators to deform the skin surface, which are specific local deformation operators depending on the nature of the joints. The value of the operator itself is determined as a function of the angular values of the specific set of joints defining the operator.

Komatsu [1988] achieved the synthesis and transformation of a human skin model using the biquartic Bézier and Gregory patches. Surface bending, twisting and swelling is done by adjusting control points according to the joint angle.

Chadwick *et al.* [1989] have proposed a layered technique based on Free-Form Deformation to apply muscle effects to a skeleton. Gascuel *et al.* [1991] have used deformable cylinders with fixed axes to construct deformable articulated objects adapted to collision processing.

Forsey [1991] has extended the hierarchical B-spline technique to 3–D character animation. A hierarchical surface is attached to an underlying skeleton in such a way that the figure designer has control over the location and scope of the surface deformation. This allows the creation of both fine and broad-scale shape deformations in association with the underlying changes in the joint angles of the skeleton.

In the VISTEL system[1995], prototype figures of real humans are constructed by laser scanning a number of limbs and connecting the polygon based representations using implicit function blending.

The LEMAN system developed at LIG [Turner and Thalmann, 1993] uses a physically-based approach. The skin surface is implemented as a simulation of a continuous elastic surface.

Most published and existing methods represent the human body as a polygonal mesh, where the vertices information is usually acquired through digitizing equipment. The input data in this kind of model involves the tedious task of entering the many significant points or vertices that configure the surface. With the advent of laser scanning systems, meshes of extreme complexity are rapidly becoming common place. However, such meshes are notoriously expensive to store, transmit, edit, render, and are awkward to animate.

More recently there has been a growing interest in modeling the human body using "implicit surfaces." As their name suggests, the surface of an object is not modeled explicitly. Instead, equations are used to represent the surface. Complex surfaces can be modeled using relatively few primitives that can easily be animated.

Current Approach: Our model [Thalmann *et al.*, 1996] is depicted by Figure 2.1. It incorporates a highly effective multi-layered approach for constructing and animating realistic human bodies. Ellipsoidal metaballs 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 skin construction is made in a three step process. First, the implicit surface resulting from the combination of the metaballs influence is automatically sampled along cross-sections with a ray casting method [Shen and Thalmann, 1995, Thalmann *et al.*, 1996]. Second, the sampled points constitute control points of a B-spline patch for each body part (limbs, trunk, pelvis, neck). Third, a polygonal surface representation is constructed by tessellating those B-spline patches for seamless joining different skin pieces together and final rendering. The method, simple and intuitive, combines the advantages of implicit, parametric and polygonal surface representation, producing very realistic and robust body deformations. By applying smooth blending twice (metaball potential field blending and B-spline basis blending), the model's data size is significantly reduced.

2.2 Simplified Model of a Limb

To reduce the number of degrees of freedom (DOFs) and to be able to robustly estimate the skeleton's position, we replace the multiple me-

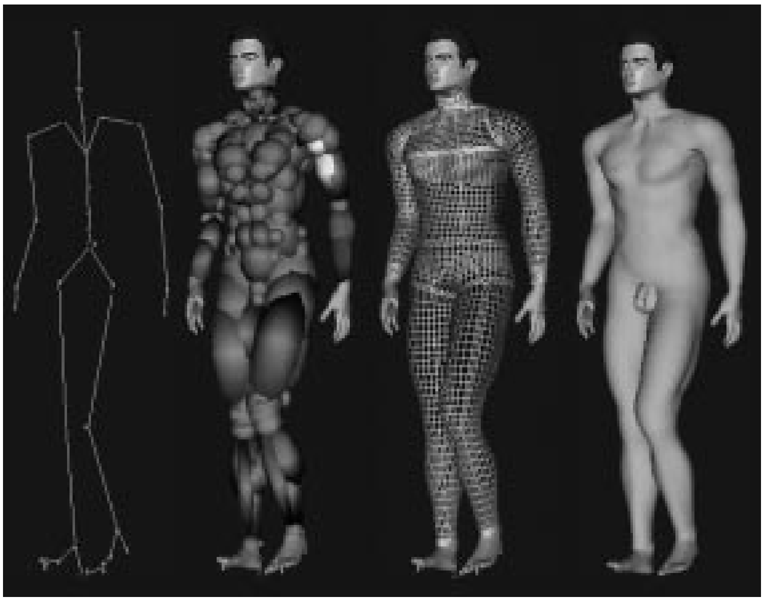
(a) (b) (c) (d)

Figure 1: The layered human body model: (a) Skeleton. (b) Ellipsoidal metaballs used to simulate muscles and fat tissue. (c) Polygonal surface refpresentation of the skin. (d) Shaded rendering.

taballs of Section 2 by one ellipsoid attached to each bone in the skeleton, as shown in Figure 2(a).

The ellipsoids attached to the skeleton have a fixed position and orientation with respect to their enclosing joints and are assumed to be cylindrically symmetric around the longest axis. Their center lies in the middle of the bone and their axis coincides with the axis of the reference joint's local coordinate system. The corresponding positions are depicted by Figure 2(b). The origin and the angles of each ellipsoid are calculated in an incremental manner, since the position and orientation of parts which are further down the hierarchy tree depend on the positions and orientations of all previous joints. For example the forearm depends on the upper arm which depends on the shoulder and so on until the root of the hierarchy is reached. Thanks to this incremental parameter calculation, the actual number of parameters for each body part differs.

We have chosen ellipsoids because, along with cylinders, they are the 3–D shapes with the least number of parameters (2: length and thickness plus the values of the skeleton's DOFs) that can be used to model human extremities. Ellipsoids, however, approximate more closely human extremities than cylinders. Furthermore, we rely on the rigid skeleton structure of Section 2 to constrain the length and connectivity of body parts.

3 Fitting the Models to Image Data

From a fitting point of view, the body model of section 2.2 embodies a rough knowledge about the shape of the body and can be used to constrain the search space. Our goal is to fix its degrees of freedom so that is conforms as faithfully as possible to the image data.

From motion sequences such as the one shown in Figure 3, we extract:

- **Range data:** Wherever a body part faces two or more of the cameras, its shape can be effectively derived from stereo.
- **Outlines:** Wherever a body part slants away from the camera, a silhouette edge appears in the images and can be used to derive 3–D information about the surface.

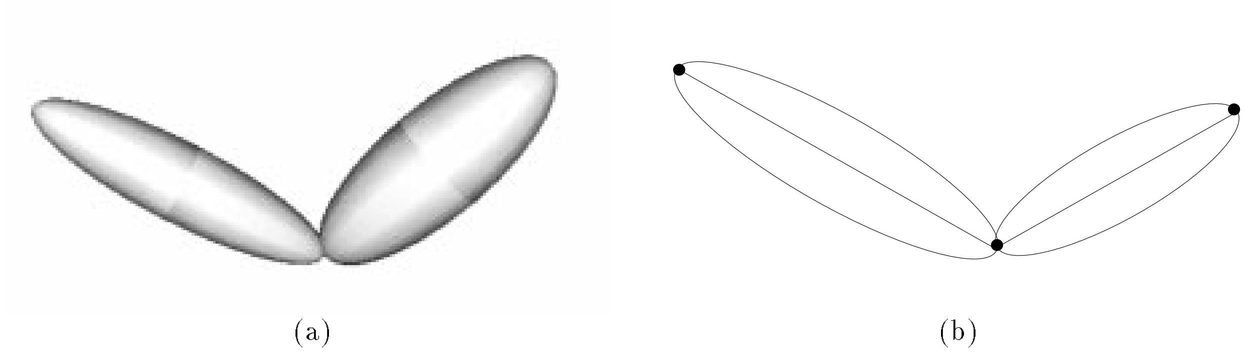

Figure 2: Simplified arm model. (a) Shaded view of the two ellipsoids representing the upperarm and the forearm. (b) Position of the two ellipsoids on the skeleton.

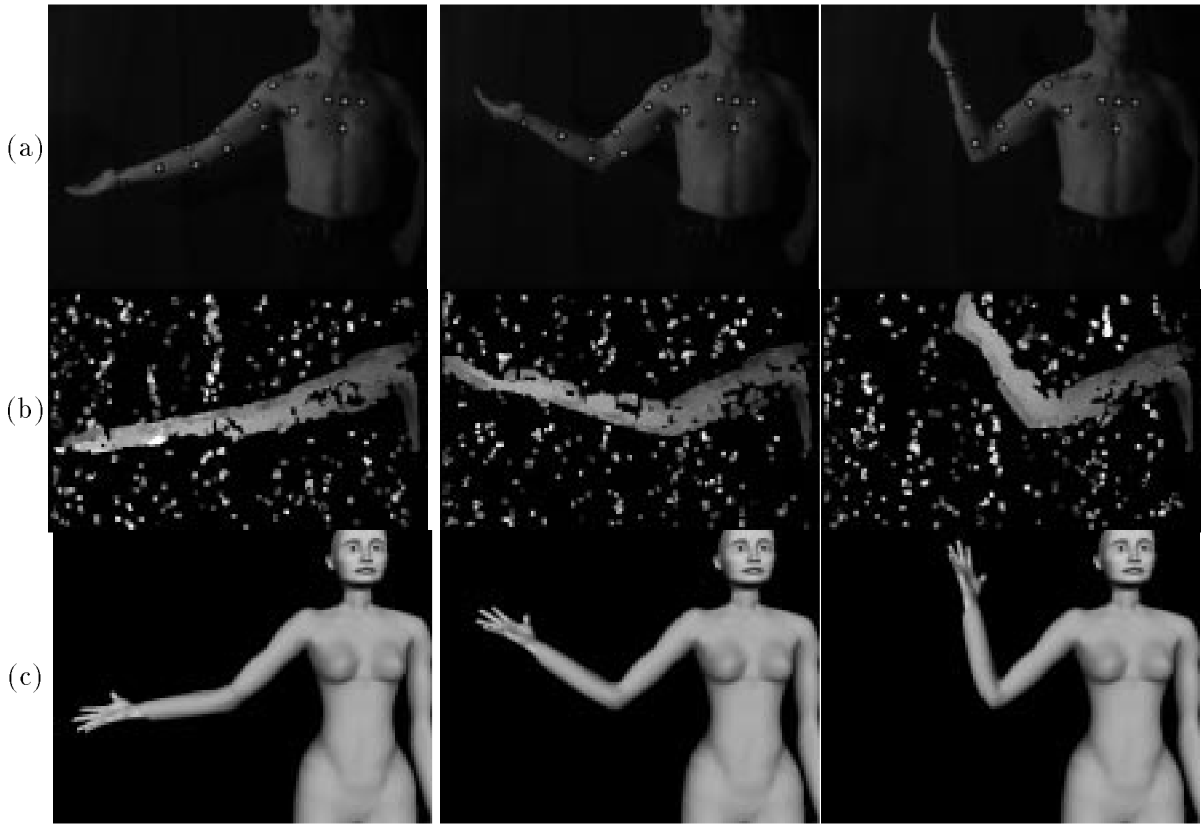

Figure 3: Arm motion sequence: (a) The three left images of three stereo pairs. (b) The corresponding disparity maps. (c) After motion recovery, a virtual human performs the same actions as the real person.

These two sources of information are therefore complementary: The former is unreliable where the surface slants away from the camera, which is precisely where silhouettes can be found. However, both these information sources are noisy and using the animation models not only to represent the data but also to guide the feature extraction process allows a substantial gain in performance.

We have implemented a least squares procedure that allows us to fit the simplified animation model of Section 2.2 and recover the motion as shown in Figure 3(c). For more details, we refer the interested reader to an earlier publication [Fua *et al.*, 1998]. In future work, we will use this knowledge to initialize the complete model and optimize its shape using the same input data.

4 Conclusion

We have presented a technique that allows us to fit a simplified animation model to noisy image data with very limited manual intervention. Because this model is closely related to the complete model we use to perform animation, these results can be used to initialize this complete model and further refine it using the same data.

The capability we intend to develop will be of great use in an area such as the generation of feature films for entertainment. Generating and animating sophisticated models requires a tremendous amount of manual labor. While this may be appropriate for big-budget one-off use, the mass market of television entertainment is much more cost-driven and would benefit greatly from using techniques such as those described above. Furthermore, there currently is an inherent limit to the complexity of the animation models: Realism requires complex models, that is, models that are controlled by large numbers of parameters. As this number increases, so does the difficulty of the task faced by the designer. Automating the process will help solve this problem and will allow an increase in realism while reducing the cost.

Acknowledgements

We wish to thank Prof. Grün for the many discussions we had with him on the topics discussed here and for sharing with us his insights about least-squares technology.

References

[Chadwick *et al.*, 1989] J.E. Chadwick, David R. Haumann, and Richard E. Parent. Construction for Deformable Animated Characters. In *Computer Graphics, SIGGRAPH Proceedings*, pages 243–252, 1989.

[Forsey, 1991] D.R. Forsey. A Surface Model for Skeleton-Based Character Animation. In *Eurographics*, 1991.

[Fua *et al.*, 1998] P. Fua, A. Grün, R. Plänkers, N. D'Appuzo, and D. Thalmann. Human Body Modeling and Motion Analysis from Video Sequences. In *International Symposium on Real-Time Imaging and Dynamic Analysis*, Hakodate, Japan, June 1998.

[Gascuel *et al.*, 1991] M.P. Gascuel, A. Verroust, and C. Puech. A Modeling System for Complex Deformable Bodies Suited to Animation and Collision Processing. *Journal of Visualization and Computer Animation*, 2:82–91, 1991.

[Komatsu, 1988] K. Komatsu. Human Skin Model Capable of Natural Shape Variation. *The Visual Computer*, 3(5):265–271, 1988.

[K.Singh *et al.*, 1995] K.Singh, J. Ohya, F. Kishino, and R. Parent. Human Figure Synthesis and Animation for Virtual Space Teleconferencing. In *International Symposium*, Research Triangle Park, North Carolina, 1995.

[Shen and Thalmann, 1995] J. Shen and D. Thalmann. Interactive shape design using metaballs and splines. In *Implicit Surfaces*, April 1995.

[Thalmann and Thalmann, 1987] N. Magnenat Thalmann and D. Thalmann. The Direction of Synthetic Actors in the Film Rendezvous a Montreal. *Computer Graphics and Applications*, 7(12):9–19, 1987.

[Thalmann and Thalmann, 1991] N. Magnenat Thalmann and D. Thalmann. Complex Models for Animating Synthetic Actors. *Computer Graphics and Applications*, pages 32–44, 1991.

[Thalmann *et al.*, 1996] D. Thalmann, J.Shen, and E. Chauvineau. Fast Realistic Human Body Deformations for Animation and VR Applications. In *Computer Graphics International*, Pohang,Korea, June 1996.

[Turner and Thalmann, 1993] R. Turner and D. Thalmann. Fast Realistic Human Body Deformations for Animation and VR Applications. In *Computer Graphics International*, pages 399–412, 1993.