Continuous Capture of Skin Deformation

Peter Sand* Leonard McMillan† Jovan Popović

Laboratory for Computer Science
Massachusetts Institute of Technology
†University of North Carolina, Chapel Hill

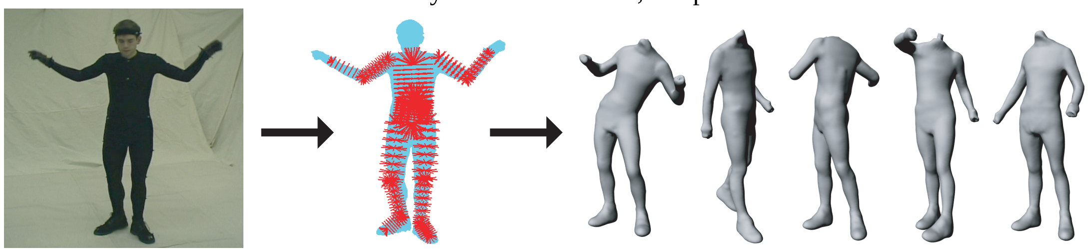

Figure 1: We extract silhouettes from video sequences to build a deformable skin model that can be animated with new motion.

Abstract

We describe a method for the acquisition of deformable human geometry from silhouettes. Our technique uses a commercial tracking system to determine the motion of the skeleton, then estimates geometry for each bone using constraints provided by the silhouettes from one or more cameras. These silhouettes do not give a complete characterization of the geometry for a particular point in time, but when the subject moves, many observations of the same local geometries allow the construction of a complete model. Our reconstruction algorithm provides a simple mechanism for solving the problems of view aggregation, occlusion handling, hole filling, noise removal, and deformation modeling. The resulting model is parameterized to synthesize geometry for new poses of the skeleton. We demonstrate this capability by rendering the geometry for motion sequences that were not included in the original datasets.

CR Categories: I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Animation; I.4.8 [Image Processing and Computer Vision]: Scene Analysis—Shape

Keywords: motion capture, skin modeling, human animation

1 Introduction

A digital replica of a moving human body has applications in video games, teleconferencing, automated news shows, and filmmaking. For example, the physical appearance of a celebrity actor could be recorded and later animated with acrobatic motions controlled by an animator or performed by a stunt double in a motion-capture suit. In current filmmaking, this application requires extensive manual labor to position and adjust skin around each bone and muscle. In some cases, months are spent matching a virtual character to an existing actor [Stokdyk et al. 2002].

Our goal is to build a skin model that replicates the skin deformations of a particular person. The technique described in this paper builds this model automatically from video of the subject and motion data that describes how the subject's skeleton moves throughout the video recording. To build the model from this data, we exploit the idea that video of a moving person provides many observations of the same surface. A single set of silhouettes (even from several viewpoints) provides a highly incomplete characterization of the geometry. By having the subject move through many different poses, local configurations of the body parts are repeated, allowing the construction of a complete model.

Our main contribution is a method of gathering silhouette observations such that a simple reconstruction algorithm can create a complete deformable model, parameterized in a way that is useful for animation. We do not contribute new techniques in the areas of skin representation and skin interpolation, but in ways of quickly acquiring skin data. By using the right combination of prior tools, we substantially simplify the problem of generating a 3D model from moving silhouettes.

Our skin model, described in Section 3, represents a complex articulated figure using a collection of elongated deformable primitives. Our acquisition algorithm, described in Section 4, uses the silhouettes to provide constraints on the possible body geometry. The reconstruction algorithm, described in Section 5, uses these constraints to find a model of the skin deformations, parameterized with the motion of the skeleton. This parameterization allows animation of the skin with new motion data.

2 Related Work

The most general 3D reconstruction systems attempt to build a model of the scene at each successive time frame, allowing the acquisition of moving objects. These systems use vision methods such as binocular stereo [Nebel et al. 2001] and voxel coloring [Vedula et al. 2002]. For certain kinds of scenes, the geometry can be reasonably represented using a *visual hull*: the space carved about by silhouettes from a set of viewpoints [Matusik et al. 2000; Würmlin et al. 2002].

Some of these methods make frame-to-frame comparisons of the geometry [Würmlin et al. 2002; Vedula et al. 2002], but they do not accumulate observations to improve the geometry. The strength of gathering information from temporally distinct views is illustrated in recent work in real-time model acquisition, in which a rigid ob-

*77 Massachusetts Avenue, Cambridge, MA 02139

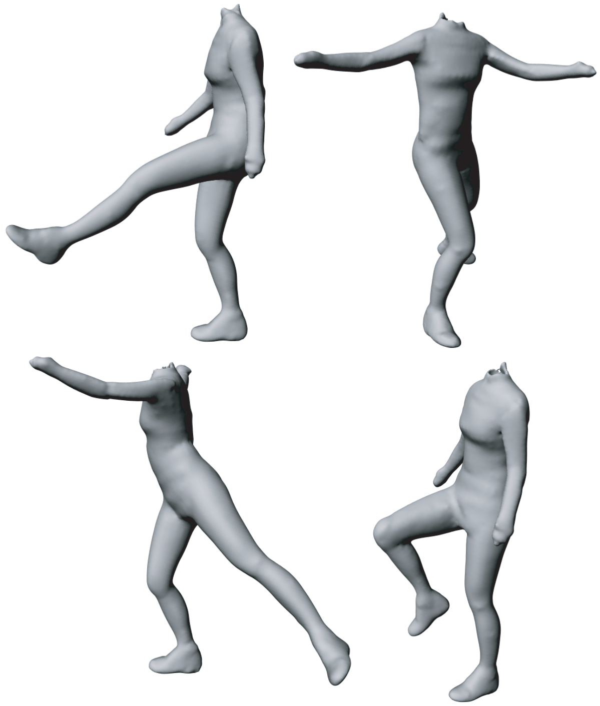

Figure 13: This model was generated from a female subject using 5 minutes of motion and silhouettes from three viewpoints. In the future we would like to capture a wide variety of people and interpolate their geometries to synthesize new deformable models.

Acknowledgements

Special thanks to Frédo Durand, Seth Teller, William Freeman, Farid Jahanmir, Barb Cutler, Tom Buehler, Adnan Sulejmanpašić, Daniel Vlasic, Eric Chan, and Eugene Hsu. We are grateful to Vicon Motion Systems for their permission to use the beta release of the Vicon iQ software. This research was sponsored in part by the NTT Corporation.

References

ALLEN, B., CURLESS, B., AND POPOVIC, Z. 2002. Articulated body deformation from range scan data. In *Proceedings of the 29th annual conference on Computer graphics and interactive techniques (SIGGRAPH)*, 612–619.

BLANZ, V., AND VETTER, T. 1999. A morphable model for the synthesis of 3d faces. In *Proceedings of the 26th annual conference on Computer graphics and interactive techniques (SIGGRAPH)*, 187–194.

BRAND, M. 2001. Morphable 3D models from video. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR)*, II:456–463.

BREGLER, C., HERTZMANN, A., AND BIERMANN, H. 2000. Recovering non-rigid 3D shape from image streams. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR)*, II:690–696.

BROOMHEAD, D., AND LOWE, D. 1988. Multivariable functional interpolation and adaptive networks. *Complex Systems 2*, 3, 321–355.

CARR, J. C., BEATSON, R. K., CHERRIE, J. B., MITCHELL, T. J., FRIGHT, W. R., MCCALLUM, B. C., AND EVANS, T. R. 2001. Reconstruction and representation of 3d objects with radial basis functions. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques (SIGGRAPH)*, 67–76.

COOTES, T. F., EDWARDS, G. J., AND TAYLOR, C. J. 1998. Active appearance models. In *Proceedings of the Fifth European Conference on Computer Vision (ECCV)*, 484–498.

GU, X., GORTLER, S. J., HOPPE, H., MCMILLAN, L., BROWN, B. J., AND STONE, A. D. 1999. Silhouette mapping. Tech. Rep. TR-1-99, Harvard.

GUENTER, B., GRIMM, C., WOOD, D., MALVAR, H., AND PIGHIN, F. 1998. Making faces. In *Proceedings of the 25th annual conference on Computer graphics and interactive techniques (SIGGRAPH)*, 55–66.

KAKADIARIS, I. A., AND METAXAS, D. 1993. 3d human body model acquisition from multiple views. In *Proceedings of the 5th IEEE International Conference on Computer Vision (ICCV)*, 618–623.

KAYDARA. 2001. *FiLMBOX Reference Guide*. Kaydara Inc., Montréal, Québec.

LEWIS, J. P., CORDNER, M., AND FONG, N. 2000. Pose space deformation: a unified approach to shape interpolation and skeleton-driven deformation. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques (SIGGRAPH)*, 165–172.

MATUSIK, W., BUEHLER, C., RASKAR, R., GORTLER, S. J., AND MCMILLAN, L. 2000. Image-based visual hulls. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques (SIGGRAPH)*, 369–374.

MIKIĆ, I., TRIVEDI, M., HUNTER, E., AND COSMAN, P. Human body model acquisition and tracking using voxel data. *International Journal of Computer Vision*. In Press.

NEBEL, J.-C., RODRIGUEZ-MIGUEL, F. J., AND COCKSHOTT, W. P. 2001. Stroboscopic stereo rangefinder. In *Proceedings of the Third International Conference on 3D Imaging and Modeling*, 59–64.

NELDER, J. A., AND MEAD, R. 1965. A simplex method for function minimization. *Computer Journal 7*, 4, 308–313.

NEVATIA, R., AND BINFORD, T. O. 1977. Description and recognition of curved objects. *Artificial Intelligence 8*, 1, 77–98.

PLÄNKERS, R., AND FUA, P. 2001. Articulated soft objects for video-based body modeling. In *Proceedings of the 8th IEEE International Conference on Computer Vision (ICCV)*, I:394–401.

RUSINKIEWICZ, S., HALL-HOLT, O., AND LEVOY, M. 2002. Real-time 3d model acquisition. In *Proceedings of the 29th annual conference on Computer graphics and interactive techniques (SIGGRAPH)*, 438–446.

SLOAN, P.-P. J., CHARLES F. ROSE, I., AND COHEN, M. F. 2001. Shape by example. In *Proceedings of the 2001 symposium on Interactive 3D Graphics*, 135–143.

STOKDYK, S., HAHN, K., NOFZ, P., AND ANDERSON, G., 2002. Spider-man: Behind the mask. Special Session of SIGGRAPH 2002.

SZELISKI, R., AND TONNESEN, D. 1992. Surface modeling with oriented particle systems. In *Proceedings of the 19th annual conference on Computer graphics and interactive techniques (SIGGRAPH)*, 185–194.

THEOBALT, C., MAGNOR, M., SCHUELER, P., AND SEIDEL, H.-P. 2002. Combining 2d feature tracking and volume reconstruction for online video-based human motion capture. In *Proceedings of the 10th Pacific Conference on Computer Graphics and Applications*, 96–103.

TORRESANI, L., AND BREGLER, C. 2002. Space-time tracking. In *Proceedings of the 7th European Conference on Computer Vision (ECCV)*, 801–812.

VEDULA, S., BAKER, S., AND KANADE, T. 2002. Spatio-temporal view interpolation. In *Proceedings of the 13th ACM Eurographics Workshop on Rendering*, 65–76.

VICON. 2003. *Vicon iQ Reference Manual*. Vicon Motion Systems Inc., Lake Forest, CA.

WANG, X. C., AND PHILLIPS, C. 2002. Multi-weight enveloping: least-squares approximation techniques for skin animation. In *Proceedings of the ACM SIGGRAPH symposium on Computer animation*, 129–138.

WÜRMLIN, S., LAMBORAY, E., STAADT, O. G., AND GROSS., M. H. 2002. 3d video recorder. In *Proceedings of the 10th Pacific Conference on Computer Graphics and Applications*, 325–334.