Optical Engineering

SPIEDigitalLibrary.org/oe

Automatic human body modeling for vision-based motion capture system using B-spline parameterization of the silhouette

Antoni Jaume-i-Capó
Javier Varona
Manuel González-Hidalgo
Ramon Mas
Francisco J. Perales

SPIE

Automatic human body modeling for vision-based motion capture system using B-spline parameterization of the silhouette

Antoni Jaume-i-Capó, Javier Varona, Manuel González-Hidalgo, Ramon Mas, and Francisco J. Perales
Universitat de les Illes Balears, Departament de Ciències Matemàtiques i Informàtica, Unitat de Gràfics, Visió, i Intelligència Artificial, Spain
E-mail: antoni.jaume@uib.es

Abstract. Human motion capture has a wide variety of applications, and in vision-based motion capture systems a major issue is the human body model and its initialization. We present a computer vision algorithm for building a human body model skeleton in an automatic way. The algorithm is based on the analysis of the human shape. We decompose the body into its main parts by computing the curvature of a B-spline parameterization of the human contour. This algorithm has been applied in a context where the user is standing in front of a camera stereo pair. The process is completed after the user assumes a predefined initial posture so as to identify the main joints and construct the human model. Using this model, the initialization problem of a vision-based markerless motion capture system of the human body is solved. © *2012 Society of Photo-Optical Instrumentation Engineers (SPIE).* [DOI: 10.1117/1.OE.51.2.020501]

Subject terms: motion capture; human body modeling; shape analysis; B-spline parameterization.

Paper 111149L received Sep. 29, 2011; revised manuscript received Nov. 29, 2011; accepted for publication Dec. 16, 2011; published online Mar. 7, 2012.

1 Introduction

Presently, human motion capture has a wide variety of applications such as 3D interaction in virtual worlds, performance animation in computer graphics, and motion analysis in clinical studies. Human motion capture is achieved by commercially available motion capture equipment, but this is far too expensive for common use and requires special clothing with retroreflective markers.[1] Markerless motion capture systems are cheaper and do not require any special clothing. They are based on computer vision techniques[2–4] and they are therefore termed vision-based motion capture systems. However, results are less accurate although sufficient for applications such as 3D interaction in virtual worlds.

In vision-based motion capture the human body model is a major issue. The model must be sufficiently accurate for representing motion by means of body postures but also simple enough to make it workable and to obtain real-time feedback from the application. Usually, the model is built previously from the user's images.[5] Common techniques for modeling are visual hulls.[6–8] Formerly built models have a realistic appearance but are too accurate for use in real-time applications. It is more interesting to build less accurate models that represent user's motions to an acceptable degree. Then we try to model the user's kinematical structure.[9]

In addition, vision-based approaches are based on the time coherence of user's motions. This implies that the user's previous postures and initial posture must be known. That is, the body model has to be initialized at the first frame. Initialization is defined by tracing the 3D position of the user's joints during the initial frame. This issue is currently being overcome by manual annotation,[10,11] a common practice in vision-based work.

In this paper we present an automatic model initialization for a vision-based motion capture system. Our algorithm is based on the analysis of the user's body shape projected onto a stereo image, that is, the image silhouette. The key idea is to cut each silhouette into different body parts, assuming the user stands with a predefined posture. From these cuts, the user's joints' 3D position can be subsequently estimated. Thus the kinematical human model of the user can also be built.

In Sec. 2 our approach to obtaining the user's joints' positions is described. An explanation of the parameterization of the user's silhouettes and the automatic placement of cuts is included in this section. Section 3 is an overview of the application where the described algorithm is used. Finally, the conclusions and direction of future work are discussed in Sec. 4.

2 Automatic Human Body Modeling

As stated previously, this work is based on the analysis of the human body shape, referring to shape as the human 2D silhouette projected onto the image (see Fig. 1). The approximation is achieved by decomposing the human shape into different parts by means of cuts. Therefore, if the user stands in a predefined posture, it is possible to assume that the joints are placed following a known order and then these joints can be correctly labeled to build the body model.

First, the body model used in the process and the required initial posture are described. Then the human body shape decomposition method is explained. Finally, a detailed explanation is given on how to find and label the joints used to build the model.

2.1 *The Simplified Articulated Body Model*

The selection of the human model has to be done according to the kind of information that is to be extracted from the image data, i.e., human body models used for synthesis applications may need to be more accurate and realistic than human body models used in motion capture applications. Our model is used for motion capture purposes,[9] specifically for 3D interaction in virtual worlds. Therefore, it has to be accurate enough for representing motions by means of body postures but simple enough to make the problem easily solvable and obtain real-time feedback from the application.

As a result, the body model used here is an articulated model with nine joints (Fig. 2), where every joint has three degrees of freedom and the links between the segments are rigid. By using this model, our objective is to estimate the

0091-3286/2012/$25.00 © 2012 SPIE

Fig. 1 The human body shape.

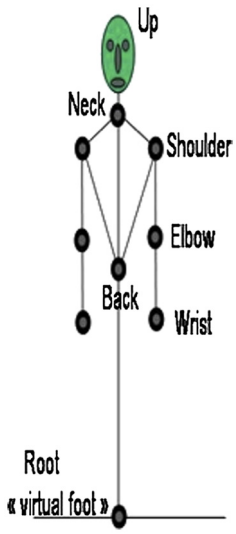

Fig. 2 Articulated body model.

anthropometric measurements of the body while the user is maintaining a predefined posture at the initialization stage. The anthropometric measurements that we wish to estimate are the limbs of the model in Fig. 2 forearms (distance from wrist to elbow), arms (distance from shoulder to elbow), trunk (distance from back to neck), upper back (distance from back to shoulders), and the virtual leg (distance from virtual foot to back).

2.2 *Human Body Shape Decomposition*

Prior to locating the body parts it is necessary to separate the human body shape from the background scene. Two methods are used to carry out this task: a chroma-key approach[12] and background subtraction.[13] It is true that the chroma-key approach[12] ensures real-time processing, but it is also true[13] that it allows for background subtraction in real time as long as the environment is a controlled one. It is necessary to emphasize that the automatic initialization system will work provided that the background can be deleted, irrespective of the algorithm used.

Once the human shape has been obtained, the silhouette has to be cut into different parts so as to find the joints' positions. According to human intuition about parts, segmentation occurs at negative minimum curvature (NMC) of the silhouette, leading to Hoffman's and Richards's minima rule: "For any silhouette, all negative minima curvature of its bounding curve are boundaries between parts."[14] Therefore, it is necessary to find the NMC points of the shape. It is possible to compute the shape curvature directly by means of finite differences in the discrete shape. However, this local computation of curvature is not robust on images. Hence, the human body shape is parameterized using a B-spline interpolation.

A p-th degree B-spline is a parametric curve, where each component is a linear combination of p-th degree basis curves. A set of contour points $\{Qk\}$, $k = 0, \ldots n$ is obtained from the image and they are interpolated with a cubic B-spline.[15] Figure 3 shows how the B-spline interpolation smoothes the human silhouette.

Using the B-spline shape parameterization, it is possible to analytically compute the partial derivatives up to order 2 to obtain the curvature values along the shape.

However, the minima rule only constrains cuts to pass through the NMC points it does not guide the selection of cuts themselves. On one hand, Singh et al. noted that when the boundary points can be joined in more than one way to decompose a silhouette, human vision prefers the partitioning scheme that uses the shortest cuts.[16] This leads to the short-cut rule, which requires a cut:

1. to be in a straight line.
2. to cross an axis of local symmetry.
3. to join two points on the outline of a silhouette such that at least one of the two points has negative curvature,
4. to be the shortest one if there are several possible competing cuts.

On the other hand, if the user's posture is known, it is possible to predict where the cuts are. In order to easily obtain the main cuts, the user must stand in a predefined posture adequate for finding all the joints of the body model (see Fig. 3). If the user assumes the same posture shown in Fig. 3, facing the stereoscopic system, the system will be able to perform the initialization correctly since it will be able to analyze all the user's joints adequately.

2.3 *Initialization of the Body Model*

Studying the initial posture in Fig. 3, it can be found that negative and positive minimum curvature points lay near the joints that we aim to find. Therefore, according to the short-cut rule and taking into account the user's initial posture, we propose the next rules to decompose the human shape:

Fig. 3 B-spline interpolation, initial posture and obtained cuts.

1. The Back is found at the negative minimum curvature point with the lowest y component.
2. The Neck is placed at the middle point of the cut between the two negative minimum curvature points with the highest y component.
3. The body's principal axis is built with the Neck and Back points. This axis divides the human body shape into two parts.
4. Shoulders are located at the middle point of the cut between the negative minimum curvature point with the highest y of the left/right side, and the negative minimum curvature point with lowest y component of the left/right side, excluding the Back point.
5. Elbows are placed at the middle point of the cut between the negative minimum curvature point with highest/lowest x component and positive maximum curvature with highest/lowest x component,

where x and y refer to the horizontal and vertical image coordinates respectively. Figure 3 also shows the human body model that results from the obtained cuts.

Subsequently, the positions of the shoulders and elbows joints are estimated as the cut middle point. Once the joints' 2D positions in each image of the stereo pair are established, their 3D positions can be estimated using the midpoint triangulation method. With this method, the 3D position is computed projecting each joint's 2D position onto each image to infinity and computing the 3D coordinates as the nearest point to these two lines.[17]

Finally, wrist joints have to be estimated to complete the body model. However, they cannot be detected using the proposed shape analysis method. Due to this, the color cue is used to find the hands on the images.[18] To locate wrists, the hands are approximated by means of ellipses. Using the previously located 3D positions of the elbows, the intersection between a 2D line defined by the elbow and center of the hand positions and the 2D ellipse hand approximation is located in the image.

3 Vision-Based Motion Capture

The described method of building a human body model has been used at the initialization stage of a motion capture system.[9] The main advantage of the proposed system is to avoid the use of invasive devices on the user. Besides, the whole process must be done in real time because the system is used in virtual environments where the interaction must be done under very strict deadline times if good feedback rates are to be achieved.

This approach combines video sequence analysis, visual 3D tracking, and geometric reasoning to deliver the user's motions in real time. As a consequence, the end user can make large upper body movements in a natural way in a 3D scene. For example, the user can explore and manipulate complex 3D shapes by interactively selecting the desired entities and intuitively modifying their location or any other attributes in real time. This technology could be used to implement a wide spectrum of applications where an individual user could share, evaluate key features, and edit virtual scenes between several distributed users. One key novelty of the present work is the possibility to interact in real time, in 3D, through the current body posture of the user in the client application.

Presently, the system is able to process not only the 3D position of the end effectors but also a set of human gesture signs, hence offering a richer perceptual user interface.

4 Conclusions

This paper presents an algorithm for initializing a human body model in an automatic way, that is, estimating the user's anthropometric measurements. This model is adequate for motion capture purposes. The algorithm is mainly based on shape analysis and human body silhouette decomposition. In order to automatically model the user's body, a set of rules are defined that, when followed, produce the desired result if the user stands in a predefined posture.

Acknowledgments

This work is partially supported by the projects MAEC-AECID A/030033/10, MAEC-AECID A2/037538/11, I3 program, TIN2007-67993 and TIN2010-16576 of the Spanish Government, with FEDER support.

References

1. Vicon Systems, http://www.vicon.com (2011).
2. T. B. Moeslund and E. Granum, "A survey of computer vision-based human motion capture," *Comput. Vis. Imag. Understand.* **81**(3), 231–268 (2001).
3. T. B. Moeslund, A. Hilton, and V. Kruger, "A survey of advances in vision-based human motion capture and analysis," *Comput. Vis. Imag. Understand.* **104**(2–3), 90–126 (2006).
4. L. Wang, W. Hu, and T. Tan, "Recent developments in human motion analysis," *Pattern Recogn.* **36**(3), 585–601 (2003).
5. F. Remondino, "3-D reconstruction of static human body shape from image sequence," *Comput. Vis. Imag. Understand.* **93**(1), 65–85 (2004).
6. J. Carranza et al, "Free-viewpoint video of human actors," *Proc. of ADM SIGGRAPH* **22**(3), 569–577 (2003).
7. G. Cheung, S. Baker, and T. Kanade, "Shape-from-silhouette of articulated objects and its use for human body kinematics estimation and motion capture," *Proc. IEEE CVPR* **1** I-77–I-84 (2003).
8. A. Hilton, M. Kalkavouras, and G. Collins, "3D studio production of animated actor model," *Vis. Imag. Signal Process., IEEE Proceedings* **152**(4), 481–490 (2005).
9. R. Boulic et al, "Evaluation of on-line analytic and numeric inverse kinematics approaches driven by partial vision input," *Virtual Reality* **10**(1), 48–61, ISSN: 1359–4338 (2006).
10. C. Bregler, J. Malik, and K. Pullen, "Twist based acquisition and tracking of animal and human kinematics," *Int. J. Comput. Vis.* **56**(3), 179–194 (2004).
11. J. Deutscher and I. Reid, "Articulated body motion capture by stochastic search," *Int. J. Comput. Vis.* **61**(2), 185–205 (2005).
12. A. R. Smith and J. F. Blinn, "Blue screen matting," *Proc. SIGGRAPH* **96**, 259–268, ACM SIGGRAPH, Addison-Wesley, New York, NY, USA (1996).
13. T. Horprasert, D. Harwood, and L. S. Davis, "A statistical approach for real-time robust background subtraction and shadow detection," *Computer* **99**(Proceeding conference 1999), 1–19 (1999).
14. D. Hoffman and W. Richards, "Parts of recognition," *Cognition* **18**(1–3), 65–96 (1984).
15. L. Piegl and W. Tiller, "The NURBS Book," Springer, ISBN: 3-540-61545-8 (1997)
16. M. Singh, G. D. Seyranian, and D. D. Hoffman, "Parsing silhouettes: the short-cut rule," *Percept. Psychophys.* **61**(4), 636–660 (1999).
17. E. Trucco and A. Verri, *Introductory Techniques for 3D Computer Vision*, Prentice-Hall (1998).
18. J. Varona, J. M. Buades, and F. J. Perales, "Hands and face tracking for VR applications," *Comput. Graph.* **29**(2), 179–187 (2005).