Automatic Human Body Modeling for Vision-Based Motion Capture

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

Unitat de Gràfics i Visió, Dept. Matemàtiques i Informàtica,

Ed. A. Turmeda, Universitat de les Illes Balears (UIB),

07122 Palma de Mallorca, Spain

antoni.jaume@uib.es

http://dmi.uib.es/~ajaume/

ABSTRACT

In this paper we present a computer vision algorithm for building a human body model skeleton in an automatic way. The algorithm is based on analyzing 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 performs a predefined initial posture for identifying her principal joints, and building the human model. Using this model, we solve the initialization problem of a vision-based markerless motion capture system of the human body.

Keywords

Motion capture, human body modeling, shape analysis.

1. INTRODUCTION

Nowadays, human motion capture has a wide variety of applications such as 3D interaction in virtual worlds, performance animation in computer graphics, or motion analysis in clinical studies. This problem has been solved by commercially available motion capture equipment, but this solution is far too

expensive for common use and it requires special clothing with retro-reflective markers [Vic05a]. Markerless motion capture systems are cheaper and they do not require any special clothing. They are based on computer vision techniques [Moe01a, Wan03a] and therefore they are named as vision-based motion capture systems. However, their results are less accurate but enough for applications such as 3D interaction in virtual worlds.

In vision-based motion capture a main issue is the human body model. This model must be accurate for representing motions by means of body postures but also simple to make the problem treatable and to obtain a real-time feedback of the application. Usually, this model is built previously and can be modeled from user's images [Rem04a]. Common based techniques for modeling are visual-hulls [Car03a, Che03a]. All these previous models have a realistic appearance but they are too accurate for their use in real-time applications. We are most interested in build less accurate models that

represents well enough users' motions. Then, we try to model the user's kinematical structure [Hil05a].

In addition, vision-based approaches are based on the temporal coherence of user's motions. This fact implies to know the user previous postures and his initial posture. That is, the body model has to be initialized at the first frame. We define this initialization as finding the joints 3D position of the user in the first frame. Currently, a common practice of vision-based works is overcome this problem by manual annotation [Bre04a, Deu05a].

In this paper we present an automatic initialization for a vision-based motion capture system. Our algorithm is based on analyzing user's body shape projected into the images, that is, his image silhouettes. The key idea is to cut each silhouette in different body parts assuming that the user stands in a predefined posture. Subsequently, from these cuts we can estimate the user joints 3D position. Therefore, we can also build the kinematical human model of the user.

The remainder of this paper is organized as follows. Section 2 gives our approach to obtain user's joint positions. In this section, we explain how to parameterize the user's silhouette and to automatically find the cuts. Next, section 3 overviews the application where we use the described algorithm. Finally, the conclusion and future work are described in section 4.
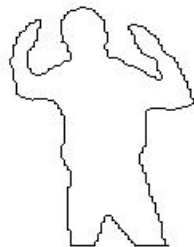

Figure 1: The human body shape.

2. FINDING USER'S BODY JOINTS

As started before, this work is based on the analysis of the human body shape. In this work, when we talk about *shape* we refer to the human 2D silhouette projected on the image, see Fig. 1. Our approximation is to decompose 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, then we can correctly label these joints to build user's body model.

First, we describe the body model used in the process and the required initial posture. Next, the human body shape decomposition is described. Finally, we explain how to find and label the joints to build the user's body 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 we want to extract 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, specifically for 3D interaction in virtual worlds. Therefore, the human model has to be accurate enough for representing motions by means of body postures but simple enough to make the problem treatable and to obtain a real-time feedback of the application.

As result, the body model used in this work is an articulated model with 9 joints (Fig. 2), where every joint have 3 degrees of freedom and the links between the segments are rigid:

1. Virtual foot: roots the body to the floor.
2. Back: corresponds to the beginning of the spine.
3. Neck.
4. 2 x Shoulders.
5. 2 x Elbows.
6. 2 x Wrists.

Using this model, our goal is to describe the configuration of the human body while user is performing a predefined posture in the initialization stage.
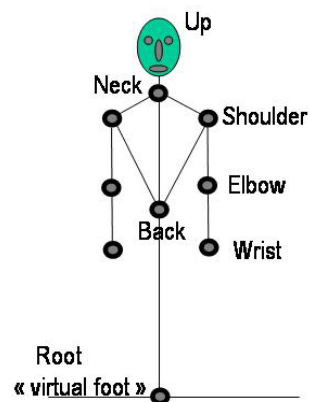

Figure 2: Articulated body model.

2.2 Human body decomposition

Before finding body parts it is necessary to segment the human body shape from the scene background. We have used two methods to do this task: a chroma-key approach [Smi96a] and by means of background subtraction [Hor00a]. Currently, we use the chroma-key approach due to the real-time nature of the final application of our system. However, the algorithm for human body modeling that we present performs equally well without chroma-key. The shapes found using both methods can be seen in Fig, 3.

Once the human shape has been obtained we need to cut the silhouette into different parts to find the joints. According to human intuition about parts, segmentation into parts occurs at negative minimum curvature (NMC) of the silhouette, leading to Hoffman and Richards's minima rule: "For any silhouette, all negative minima curvature of its bounding curve are boundaries between parts" [Hof84a].
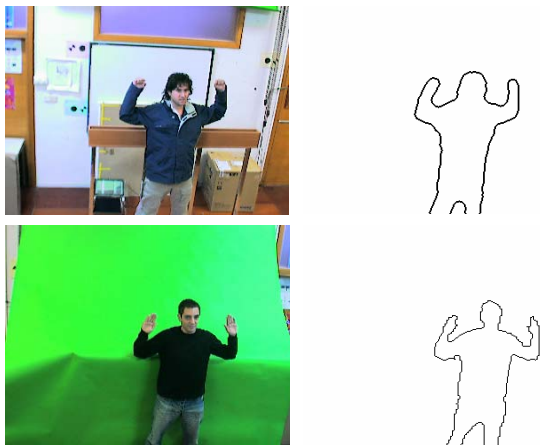

Figure 3: Human body segmentation.

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 over the discrete shape. However, this local computation of curvature is not robust in images. Hence, we parameterize the human body shape using a B-Spline interpolation.

A *pth* degree B-Spline is a parametric curve, where each component is a linear combination of *pth* degree basis curves. From the image we obtain a set of contour points $\{Q_k\}$, k=0,...n and we interpolate these points with a cubic B-Spline [Pie97a]. It can be seen in Fig 4 how the B-spline interpolation smoothes the human silhouette.

Figure 4: B-Spline interpolation of 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. In Fig. 5 the maximum and minimum of the curvature values are shown.


Figure 5: Minimum (in white) and maximum (in green) of the curvature.

However, the minima rule only constrains *cuts* to pass through the NMC points, but 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 which uses the shortest *cuts* [Sin99a]. This leads to the short-cut rule which requires a cut:

1. Be straight line.

2. Cross an axis of local symmetry.

3. Join two points on the outline of a silhouette, such that at least one of the two points has negative curvature.

4. Be the shortest one if there are several possible competing cuts.

On the other hand, if we know the user's posture it is possible to predict where the *cuts* are. In order to easily obtain the main *cuts*, it is required that the user stands in a predefined posture adequate for finding all the joints of our body model, see Fig. 6.
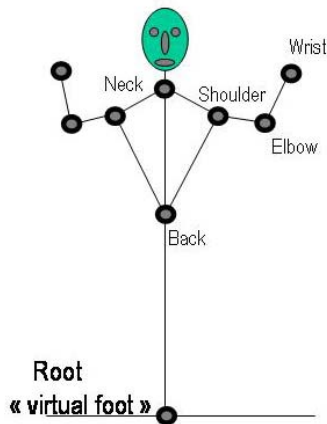
Studying the initial posture of Fig. 6, we find that negative and positive minimum curvature points lay near the joints that we aim to find. This fact is clearly visible in Fig. 5. Then, according to the short-cut rule and taking into account the user's initial posture we propose the next rules to decompose the human shape:

1. The Back is found at the negative minimum curvature point with the lowest y component.

2. Neck is placed at the middle point of the *cut* between the two negative minimum curvature points with the highest y component.

3. Build the body principal axis with the Neck and Back points. This axis divides the human body shape in 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 refers to the horizontal and vertical image coordinates respectively. Applying these rules we obtain the user's body shape decomposition as it is showed in Fig. 7. Fig. 8 shows the human body model from the obtained *cuts*.

Subsequently, we estimate the positions of the shoulders and elbows joints as the *cut* middle point. Once we have the joints 2D positions in each image of the stereo pair, we can estimate their 3D position using the mid-point triangulation method. With this method, the 3D position is computed projecting each joint 2D position on each image to infinity and computing their 3D coordinates as the nearest point to these two lines [Tru98a].

Figure 7. Obtained *cuts*.

Figure 8. Human Body Models of different users.

Finally, wrist joints need to be estimated to complete our body model. However, they can not detect using the proposed shape analysis method. In this case we use the color cue to find in the images the hands [Var05a]. To locate wrists we approximate the hands by means of ellipses. Using the 3D positions of elbows previously located, we search in the image for the intersection between a 2D line, defined by the elbow and center of the hand positions, and the 2D ellipse hand approximation.

3. VISION-BASED MOTION CAPTURE

The described method to build a human body model has been used at the initialization stage of a motion capture system [Bou05a]. The main advantage of the proposed system is to avoid invasive devices on the user. Besides, the whole process must be done in real-time because the system is used for interacting with virtual environments, where the interaction must be done under very strict deadline times to achieve good feedback rates.

This approach combines video sequence analysis, visual 3D tracking and geometric reasoning; to deliver the user's motions in real-time. This brings to the end user to make large upper body movements in a natural way in a 3D scene. For example, 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 3D position of the end effectors, but also a set of human gesture signs, hence offering a richer perceptual user interface. Results obtained using this system is shown in Fig. 9.

4. CONCLUSIONS

In this paper we have presented an algorithm for initializing a human body model in an automatic way. 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 we

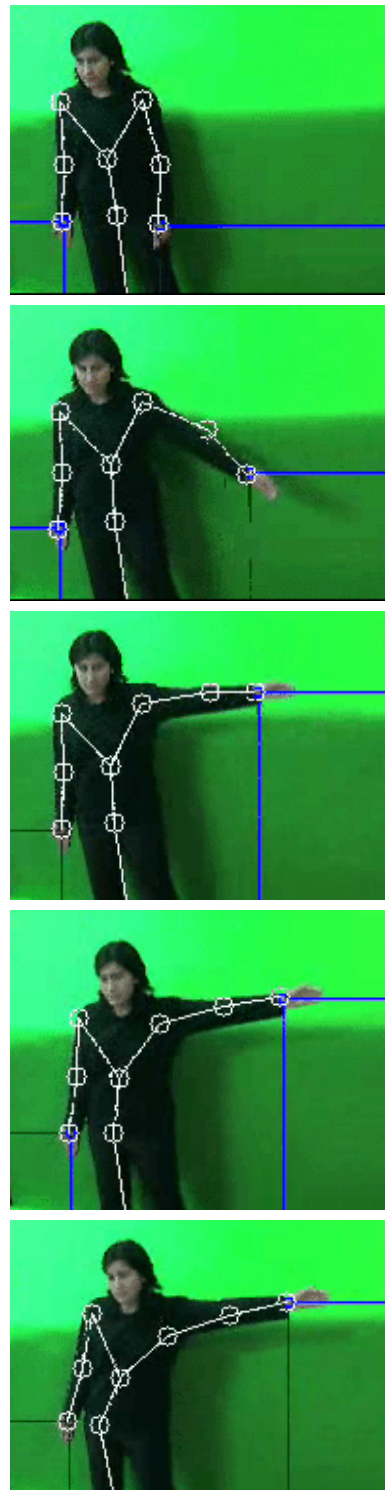have defined a set of rules that performs well if the user stands in a predefined posture.

Figure 9: Using the body model in vision-based motion capture.

5. ACKNOWLEDGMENTS

6. REFERENCES

[Bou05a] R. Boulic, J.Varona, L. Unzueta, M. Peinado, A. Suescun, F. Perales, "Real-Time IK Body Movement Recovery from Partial Vision Input", Proceedings of the 2nd International ENACTIVE Conference, Genoa, 2005.

[Bre04a] C. Bregler, J. Malik, K. Pullen, K., "Twist based acquisition and tracking of animal and human kinematics", International Journal of Computer Vision, 56(3):179-194.

[Car03a] J. Carranza, C. Theobalt, M. Magnor, H. Seidel, "Free-Viewpoint video of Human Actors", Proceedings of ADM SIGGRAPH 2003:569-577.

[Cha05a] J. Chai, J.K. Hodgins, "Performance animation from low-dimensional control signals", Proceedings of SIGGRAPH 2005, ACM Trans. Graph. 24(3):686-696, 2005.

[Che03a] G. Cheung, S. Baker, T. Kanade, "Shape-From-Silhouette of articulated objects an its use for human bosy kinematics estimation and motion capture", Proceedings of IEEE CVPR 2003.

[Deu05a] J. Deutscher, I. Reid, "Articulated Body Motion Capture by Stochastic Search", International Journal of Computer Vision, 61(2):185--205, 2005.

[Hor00a] T. Horprasert, D. Harwood, L. S. Davis, "A robust background subtraction and shadow detection", Proc. ACCV'2000, Taipei, Taiwan, 2000.

[Hil05a] A. Hilton, M. Kalkavouras, G. Collins, "3D studio production of animated actor model", Vision, Image and Signal Processing, IEEE Proceedings-Volume 152, Issue 4, 5 Aug. 2005 Page(s):481 - 490.

[Hof84a] D. Hoffman, W. Richards, "Parts of recognition", Cognition 18: 65-96, 1984.

[Lee04a] M.W. Lee, I. Cohen, "Human Upper Body Pose Estimation in Static Images", Proc. ECCV'2004, LNCS 3022: 126--138, 2004.

[Moe01a] T.B. Moeslund, E. Granum, "A Survey of Computer Vision-Based Human Motion Capture", Computer Vision and Image Understanding 2001, 81(3):231-268

[Pie97a] L. Piegl, W. Tiller, "The NURBS Book ", Springer, ISBN 3-540-61545-8, 1997.

[Rem04a] F. Remondino, "3-D reconstruction of static human body shape from image sequence", Computer Vision and Image Understanding, 93(1):65--85, 2004.

[Sin99a] M. Singh, G. D. Seyranian, D. D. Hoffman, "Parsing silhouettes: The short-cut rule", Percetion and Psychophysics, 61:636-660, 1999.

[Smi96a] A.R. Smith and J.F. Blinn, "Blue screen matting," Proc. SIGGRAPH 96, pp: 259-268, ACM SIGGRAPH, Addison-Wesley, 1996.

[Tru98a] E. Trucco, A. Verri, "Introductory Techniques for 3D Computer Vision", Prentic-Hall, 1998.

[Var05a] J. Varona, J.M. Buades, F.J. Perales, "Hands and face tracking for VR applications", Computers & Graphics, 29(2):179-187, 2005.

[Vic05a] Vicon Systems, 2005. http://www.vicon.com.

[Wan03a] L. Wang, W. Hu, T. Tan, "Recent developments in human motion analysis", Pattern Recognition 2003, 36(3):585-601.

[Zha01a] L. Zhao, "Dressed human modeling, detection and parts localization", PhD thesis, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, July 2001.