Extraction and tracking of a body skeleton from multiple views: A survey

Alexander Pinzon Fernandez

***Abstract*—This paper presents a survey of the motion capture process by extracting the skeleton. The skeleton is the easy way to capture the essential dynamic relationships and movement of a body. The skeletons do not have a formal definition, then not exist accurate way to evaluate them, then looking some features as homotopic, invariant, centeredness, to meet the requirements. For this reason exist are many methods for skeleton extraction, that will be studied in this article.**

***Index Terms*—Skeleton extraction, motion capture.**

CONTENTS

I Introduction 1

II Motivation 1

III Previous Work 1
- III-A Data Acquisition 1
- III-B Image segmentation 1
- III-C Three-dimensional reconstruction 2
  - III-C1 Image-based Reconstruction 2
- III-D Skeleton Extraction 2

IV Problem Definition 2

References 2

I. INTRODUCTION

Body dynamics is an important area of research devoted to understanding their dynamic physical behavior in a complex environment that can be used for biometric applications. We are particularly interested in understanding and interpreting human behavior in complex environments. In a number of applications it is important to identify the actions of certain parts of the body, e.g. hand-gestures, gait analysis, and facial expression analysis[10]. Finally, the modeling of human behavior can be used for a number of applications such as generating natural animation or graphics, understanding normal and pathological behaviors, and analysis of data for medical applications [18].

The skeletons are 1D structures that represent a simplified version of the geometry and topology of a 3D object. While the 3D representation is invaluable, many applications require alternative representations as a skeleton of these models[3]. Skeletons reduce dimensionality of digital objects but at the same time a way to capture the essential dynamic relationships[16].

Alexander Pinzon Fernandez, Bioingenium Research Group

II. MOTIVATION

Technological advances in the fields of telecommunication, graphic hardware and geometry processing during the last decade, have contributed to an evolution of the digital data being manipulated and transmitted over the Internet. Nowadays, static and dynamic three-dimensional meshes constitute the emerging multimedia content[12]. For this reason it is vitally important to develop methods for manipulating and modeling of three-dimensional meshes, which can be useful in numerous applications.

III. PREVIOUS WORK

The extraction and skeleton tracking is achieved using a set of organized activities inside a process. These activities are: data acquisition from a real object, the 3D reconstruction from data capture, skeleton extraction from object reconstructed three-dimensional, and tracking the skeleton taken as an iterative capture process with a skeleton topologically invariant in time.

A. Data Acquisition

There are different approaches in the acquisition of spatial information of real objects, some based on image processing, sensor-based other (mechanical, electromagnetic and inertial among others), and others like laser scanner, tomography, ultrasound and so on [8].

The acquisition methods based on signal processing using sequences of images produced by one or more cameras. When using multiple cameras is exploited spatial coherence given by stereoscopic vision through the use of markers (bright, reflective, or detected by filters), and the silhouettes of the objects obtained by segmenting images [18].

When using a single image using a priori models are also added restrictions on the search for solutions, these models can be based on probabilistic approaches and machine learning.

B. Image segmentation

The image segmentation is a set of methods for dividing an image into regions, given certain characteristics. Segmentation can be automatic ie without user interaction, or semi-manual requiring supervision[9]. Here are some of the most widely used segmentation methods:

- Methods for active contour segmentation [7].
- Snake method using filters with gradient direction.
- Methods that use space-time consistency constraints forms especially boxes and ellipses [11].
- Probabilistic methods with Bayesian networks to model the interaction of fields [13].

C. Three-dimensional reconstruction

Three-dimensional reconstruction is done with techniques in which real objects are represented by data structures or functions with emphasis on features like volume and shape. There is currently a great deal of reconstruction techniques and methods of 3D Mesh (13), given a set of unconnected 3D points[5].

1) Image-based Reconstruction: The image-based reconstruction has two major approaches used with each brand, and one without them, we must also take into account the intrinsic parameters (focal length, distortion, etc..) And extrinsic (position, orientation, etc.). camera [6].

In the reconstruction with the use of markers is filtering the image and locate those markers which are generally little points of any color at the scene as white, red or deep green, are also used markers that emit light and small lights, or the use of black and white images taken with a recognizable pattern.With these points located on the segmented images are necessary to make the solution of the equations of the corresponding linear transformations based on a stereo system with a pinhole camera model [17].

D. Skeleton Extraction

The extraction of the skeleton is a problem currently is treated from different angles, some estimates made of the skeleton in 2D and then used to estimate a priori information was transformed as the actual three-dimensional skeleton [14]. Other approaches perform an iterative reduction of the number of polygons with some restrictions on contraction and energy minimization, until found a skeleton irreducible [2].

Other methods use different topological representation of a three dimensional object by means of graphs, then perform a simplification of the graph, and finally used a skeleton model which transformed to minimize the error between the skeleton and the graph given by some heuristics based on metrics [4].

Other methods use of probabilistic graphical models to resolve both issues: incorporate the spatial dependencies between the vertices into the clustering process, while providing a globally optimal solution. In particular, base on Markov Gibbs Random Fields (MRF) [12].

To select a method of extracting is needed is an analysis of the desired properties of the curve-skeleton, as required by the various applications, and how the various existing curve-skeletonization methods satisfy these properties[15].

Other methods to extract such a skeleton on-the-fly, both from point clouds and polygonal meshes. The algorithm is based on a deformable model evolution that captures the object's volumetric shape. The deformable model involves multiple competing fronts which evolve inside the object in a coarse-to-fine manner. Frst track these fronts' centers, and then merge and filter the resulting arcs to obtain a curve-skeleton of the object[1].

What is needed is an analysis of the desired properties of the curve-skeleton, as required by the various applications, and how the various existing curve-skeletonization methods satisfy these properties[3], because the requirements of a skeleton differ with applications[13].

IV. Problem Definition

The motion tracking is performed with expensive conventional capture systems ranging from 70.000 to approximately 200.000 dollars. This track of the movement has these disadvantages.

- Use markers in the form of clothing and devices attached to the body and limbs, which alter the naturalness of the movements. In addition, experts need to position the markers because the markers must be located in specific anthropometric points.
- Stereo systems that perform a three-dimensional reconstruction from multiple viewpoints handle large volumes of data for the geometry of the body, which requires high performance machines.

References

[1] Ariel Shamir Leif Kobbelt Andrei Sharf, Thomas Lewiner. On-the-fly curve-skeleton computation for 3d shapes. In *Computer Graphics Forum*, volume 26 of *323-328*, School of Computer Science, Tel Aviv University; Departament of Mathematics, PUCRio de Janiero; Efi Arazi School of Computer Science, The Interdisciplinary Center, Herzliya; Computer Graphics Group, RWTH Aachen, 2007. Skeleton Extraction.

[2] Oscar Kin-Chung Au, Chiew-Lan Tai, Hung-Kuo Chu, Daniel Cohen-Or, and Tong-Yee Lee. Skeleton extraction by mesh contraction. *ACM Transactions on Graphics*, 27(3):10, 2008. Skeleton Extraction.

[3] Nicu D. Cornea and Patrick Min. Curve-skeleton properties, applications, and algorithms. *IEEE Transactions on Visualization and Computer Graphics*, 13(3):530–548, 2007. Skeleton Extraction Survey Member-Silver,, Deborah.

[4] Edilson de Aguiar, Carsten Stoll, Christian Theobalt, Naveed Ahmed, Hans-Peter Seidel, and Sebastian Thrun. Performance capture from sparse multi-view video. In *SIGGRAPH '08: ACM SIGGRAPH 2008 papers*, pages 1–10, New York, NY, USA, 2008. ACM. 3D Reconstruction.

[5] Edilson de Aguiar, Christian Theobalt, Marcus Magnor, and Hans-Peter Seidel. Reconstructing human shape and motion from multi-view video. In *2nd European Conference on Visual Media Production (CVMP)*, pages 42–49, London, UK, December 2005. The IEE. 3D Reconstruction.

[6] Teofilo E. de Campos and David W. Murray. Regression-based hand pose estimation from multiple cameras. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 1:782–789, 2006. Motion Capture.

[7] Matthew Flagg, Atsushi Nakazawa, Qiushuang Zhang, Sing Bing Kang, Young Kee Ryu, Irfan Essa, and James M. Rehg. Human video textures. In *I3D '09: Proceedings of the 2009 symposium on Interactive 3D graphics and games*, pages 199–206, New York, NY, USA, 2009. ACM. Other.

[8] D. M. Gavrila. The visual analysis of human movement: a survey. *Comput. Vis. Image Underst.*, 73(1):82–98, 1999. Motion Capture Survey.

[9] Ning Jin and F. Mokhtarian. Image-based shape model for view-invariant human motion recognition. In *Advanced Video and Signal Based Surveillance*, pages 336–341, Sept. 2007. 3D Reconstruction.

[10] Paul G. Kry. *Interaction Capture and Synthesis of Human Hands*. PhD thesis, University of British Columbia, 2005. Motion Capture.

[11] J.-L. Landabaso, L. Lizcano, and M. Pardas. Shape from inconsistent silhouette for free viewpoint video. In *Image Processing*, pages 213–216, Oct. 2008. 3D Reconstruction.

[12] Guillaume Lavoué and Christian Wolf. Markov random fields for improving 3d mesh analysis and segmentation. In *Eurographics 2008 Workshop on 3D Object Retrieval*, April 2008. Skeleton Extraction.

[13] Wan-Chun Ma, Fu-Che Wu, and Ming Ouhyoung. Skeleton extraction of 3d objects with radial basis functions. In *SMI '03: Proceedings of the Shape Modeling International 2003*, page 207, Washington, DC, USA, 2003. IEEE Computer Society. Skeleton Extraction.

[14] Fabio Remondino and Andreas Roditakis. 3d reconstruction of human skeleton from single images or monocular video sequences, 2003. 3D Reconstruction.

[15] J. Starck, G. Miller, and A. Hilton. Video-based character animation. In *SCA '05: Proceedings of the 2005 ACM SIGGRAPH/Eurographics symposium on Computer animation*, pages 49–58, New York, NY, USA, 2005. ACM. Motion Capture.

[16] S. Svensson, I. Nyström, and G. Sanniti di Baja. Curve skeletonization of surface-like objects in 3d images guided by voxel classification. *Pattern Recognition Letters*, 23(12):1419 – 1426, 2002. Skeleton Extraction.

[17] Thorsten Thormählen and Hans-Peter Seidel. 3d-modeling by ortho-image generation from image sequences. In *SIGGRAPH '08: ACM SIGGRAPH 2008 papers*, pages 1–5, New York, NY, USA, 2008. ACM. Modeling 3D.

[18] Jessica Junlin Wang and Sameer Singh. Video analysis of human dynamics - a survey. *Real Time Imaging*, 9:321–346, 2003. Motion Capture survey.