A REAL-TIME HUMAN BODY SKELETONIZATION ALGORITHM FOR MAX/MSP/JITTER

André Baltazar

INESC Porto and
Portuguese Catholic University -
School of the Arts
andrebaltaz@gmail.com

Carlos Guedes

INESC Porto
University of Porto, School of
Engineering
carlosguedes@mac.com

Bruce Pennycook

The University of Texas at Austin
bpennycook@mail.utexas.edu

Fabien Gouyon

INESC Porto
fgouyon@inescporto.pt

ABSTRACT

In this paper we present an algorithm for real–time full-body skeletonization and visualization implemented as two external objects for Max/MSP/Jitter. These objects are intended to provide an accurate description of bodily motion as captured by a video camera, to be used as musical rhythm controller in interactive music systems.

1. INTRODUCTION

The use of video cameras as gestural controllers for games and other applications is expanding thanks to the increase in computational speed and the developments in computer vision. Sony's Eye Toy camera has been used in Playstation 2 consoles since 2003 [1]; Microsoft announces a "revolutionary new way to play: no controller required" thanks to the use of video cameras as gestural controllers for soon-to-be-released Project Natal for Xbox consoles [2].

Utilizing video cameras as gestural controllers for real-time music generation is one of the aims of project "Kinetic controller driven adaptive and dynamic music composition systems." This joint research project involves INESC Porto, the New University of Lisbon in Portugal, and the University of Texas at Austin, with YDreams [3] and Casa da Música [4] as partners. The project includes the development of new techniques and strategies for computer-assisted composition in the context of real-time user control with non-standard human interface devices. The research team is designing and implementing real-time software that will provide tools and resources for music, dance, theatre, installation artists, interactive kiosks, computer games, and internet/web information systems.

Systems employing video analysis of human body motion for the purposes of creating and controlling electronic music have been developed since the mid 1990s. Earlier work of composers Todd Winkler [5] and Richard Povall [6], the seminal work of Antonio Camurri and colleagues with the EyesWeb platform in expressive gesture analysis [7], choreographer Robert Weschler's work with Palindrome [8], Mark Coniglio's continued development of his Isadora programming environment [9] plus the groundbreaking work Troika Ranch [10] has done in interactive dance stand out as important references on how video analysis technologies have provided interesting ways of movement-music interaction.

1.1. Extracting rhythms from video analysis of the human body

Previous work by one of the authors [11] provided ways of enabling dancers to generate musical rhythmic sequences from movement, by analyzing periodicities in the frame-differencing (quantity of motion) analysis of a video stream. Although this work provided satisfactory results in terms of rhythmic control and generation, this type of analysis does not allow looking at the rhythms produced by different body parts simultaneously. More recent work by Luiz Naveda [12], has utilized video segmentation techniques to analyze the rhythm of separate body parts (trunk, arms, legs and head) of samba dancers as a means of figuring out the contribution of each body part in recreating that rhythm through movement. Naveda's analysis is not done in real time but nevertheless clearly shows how the movement of each body part is related to different rhythmic strata in samba.

The accurate segmentation of the human body is thus an important issue for increased gestural control using video cameras. In this paper, we present an algorithm for real-time human body skeletonization for Max/MSP. This algorithm was inspired by existing approaches and adds some important improvements, such as means to acquire a better representation of the human skeleton in real time. By doing accurate segmentation analysis, we expect increased rhythmic control from bodily action captured by a video camera.

1.2 Current algorithms for real-time skeletonization

Several human action skeletonization methods were proposed in the past few years (see for example [13] for three major techniques used in computer vision). However, there are still very few skeletonization methods that work in real time for human motion analysis. One of the most relevant in this field is the method proposed by Fujyoshi et al [14]. They propose a method for analyzing the motion of a human target in a video stream. Moving targets are detected and their boundaries extracted. A “star” skeleton is produced (Figure 1) to form these boundaries. Two motion cues are determined from this skeletonization: body posture, and cyclic motion activities such as walking or running. From these, the target’s gait can be computed. Skeletonization does not require an *a priori* human model or a large number of pixels in the target. Furthermore, it is computationally inexpensive, and thus ideal for real-time video analysis applications such as outdoor video surveillance.

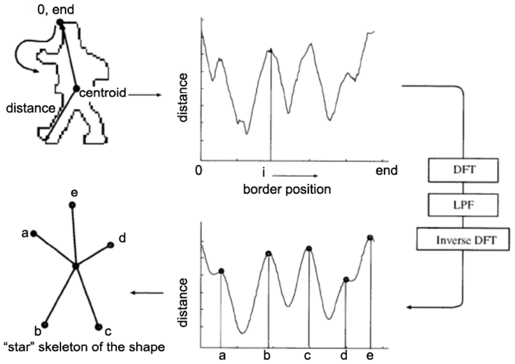

Figure 1 – From [14]. The boundary is “unwrapped” as a distance function from the centroid. This function is then smoothed and external points are extracted (used with permission from the authors).

The work of Chen et al [15] follows a similar approach. They present a Hidden Markov Model (HMM)-based methodology for action-recognition using a star skeleton as a representative descriptor of human posture. They implement a system to automatically recognize ten different types of actions, which has been tested on real human action videos with a great deal of success. They also designed a posture codebook, which contains representative star skeletons of each action type and defined a star distance to measure the similarity between actions.

2. OUR APPROACH

Our approach uses the star skeleton method as a point of departure with some modifications. We propose a more accurate skeleton representation for providing improved rhythmic control from bodily action in real time. The goal is to have a skeleton representation as close as possible to a real human skeleton so that one can better measure the limbs articulation and trajectory in relation to their point of origin (e.g. shoulders and neck) instead of the center of the star skeleton.

We make use of Jean-Marc Pelletier’s cv.jit.lib for image analysis in Max/MSP/Jitter [16] for the necessary background subtraction and silhouette extraction together with new externals developed by us for the real-time full-body skeletonization and skeleton drawing.

2.1. Objects and Algorithm Description

We developed two objects for Max/MSP/Jitter for real-time full-body skeletonization – kin.skel and skel.draw. Object kin.skel is responsible for the output of several important elements of the skeleton: mass center coordinates, hands and feet coordinates, arms and legs angle. Object skel.draw is responsible for the drawing of the skeleton in an output video matrix. They are inherently connected since skel.draw provides the visual output to kin.skel. A demo patch of the algorithm as well as objects kin.skel and skel.draw can be downloaded at http://sites.google.com/site/kineticproject09/home

2.2. kin.skel

This object was developed based in the method described by [14] with some modifications. The most relevant is the fact that Fujiyoshi et al. determine the extremities of the skeleton by finding the maximum differences from the edges to the mass center. In our case, we calculate them by analyzing the whole edge frame shot and finding the rightmost and leftmost active pixels on the edge above and below the mass center, as well as the uppermost pixel in the edge.

The real-time extraction of the skeleton is articulated in four stages: edge detection, mass center and extremities calculation, skeleton drawing, and outputs. Below we give a brief description of these stages.

2.2.1. Edge detection

For the edge detection we use the algorithm developed by Canny [17]. It takes a binary image as input and outputs the silhouette of the human body. It is possible to change the edge detection threshold at run time.

2.2.2. Mass center and extremities calculation

Once the edge of the subject is obtained, we use it for calculating the mass center coordinates by simply adding all the pixels’ coordinates and dividing them by the number of pixels. The extremities calculations and the skeleton drawing are based on the correlations of ideal human proportions with geometry described by the ancient Roman architect Vitruvius in Book III of his treatise De Architectura [18].The Vitruvian human proportions laws were made famous by Leonardo Da Vinci's drawing "Vitruvian Man" - named in honor of the architect. From Vitruvius's description of the human body's proportions, we extrapolate the shoulders' positions and use them to connect the arms, thus obtaining a more accurate representation of the human body than the one given by [14] (Figure 2).

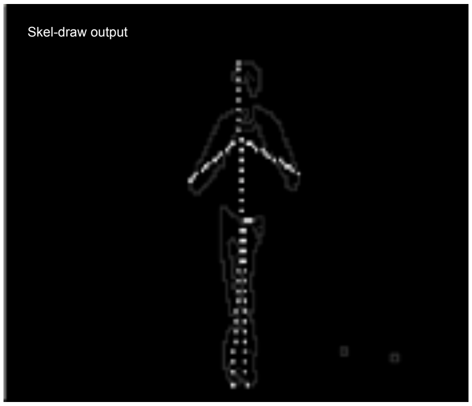

Figure 2 – The output of our algorithm based on Vitruvius's description of the human body´s proportions.

2.2.3. Outputs from the object

There are two main outputs from the kin.skel object. One output is the set of points coordinates that will feed the skel.draw object for the skeleton visualization. The other output sends all the characteristics obtained by kin.skel (mass center, right arm angle, left arm angle, right leg angle, left leg angle) packed as a list for further analysis in the Max environment.

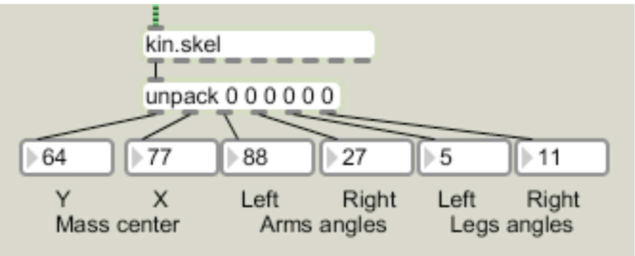

Figure 3 – Output of kin.skel in list format.

2.3. skel.draw

Once the major points coordinates are obtained, they are passed to a routine responsible for the skeleton drawing. The several point coordinates that will constitute the line that unites the major points are calculated by successively finding the middle point between extremities. To retain real-time output, each line is drawn with only 30 pixels, which produces an adequate representation of the skeleton. Figure 4 depicts all image processing stages from video input to skeleton drawing.

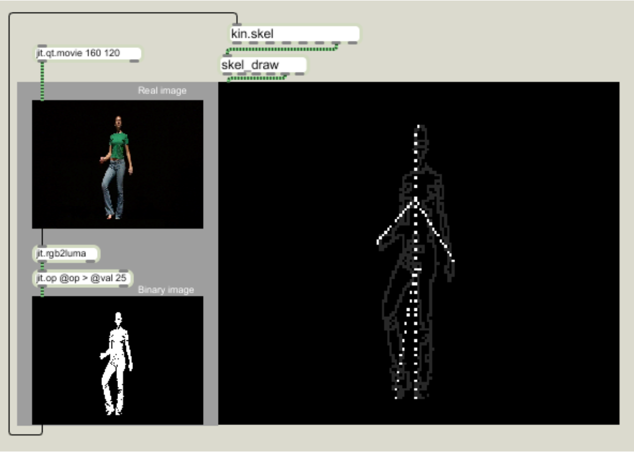

Figure 4 – The video input on top left, the binary image extraction below skeleton processing and draw by objects kin.skel and skel.draw respectively

3. APPLICATIONS

The approach we present here has potential for multi-layered rhythm generation in real time from the human body and for applications involving automatic music generation and rhythmic analysis of the human body. As we show in Figure 5, the output of the algorithm can be used to analyze in real time the variation of the angles of the arms and legs of the skeleton, as well as the variation of the mass center position. This can be utilized to analyze how each body part relates to a given stratum of a rhythm for example. More interestingly, this information can be used to enable humans to generate rhythms using different body parts for applications involving interactive music systems and automatic music generation.

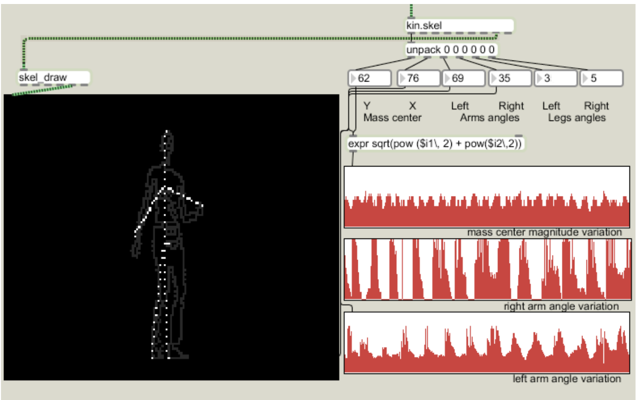

Figure 5- Visualization of kin.skel output over time: top: mass center magnitude; middle: right arm angle; bottom: left arm angle. Note how different periodicity rates can be observed from each of the body parts.

4. CONCLUSIONS AND FUTURE WORK

In this paper we presented an algorithm for real-time human-body skeletonization and drawing consisting of two external objects for Max/MSP/Jitter. This algorithm is a modified version of the star method for skeletonization proposed by Fujyoshi et al [14], a computationally inexpensive algorithm suitable for real-time applications. Our algorithm provides a more accurate description the human skeleton, suitable to be used as a musical rhythm controller in real-time for interactive music systems. Of particular interest, is the ability of the algorithm to output the variation in time of different body parts, namely arms, legs and center of mass, which increases rhythmic control in real time.

Future developments of the algorithm aim for a better representation of the skeleton including shoulders and joints by following the Vitruvian description of human proportions. Other developments will include methods for pattern recognition similar to those by Chen et al [15] for real-time posture recognition in order to improve real-time full-body analysis in interactive environments.

5. AKNOWLEDGMENTS

This joint research project is made possible thanks to the support from the Portuguese Foundation for Science and Technology (grant UTAustin/CD/0052/2008) and the UT Austin | Portugal Program in Digital Media [19]. We would like to thank Jaime Cardoso from INESC Porto's Multimedia and Telecommunications Unit for his input on the algorithm's development and Jean-Marc Pelletier for his help with unraveling the Jitter Software Development Kit. We also thank Luiz Naveda from the University of Ghent for lending us some videos of dancers for testing, and to dancer Juliana Morais for letting us use captions of her dancing in this paper.

6. REFERENCES

[1] http://en.wikipedia.org/wiki/EyeToy

[2] http://www.xbox.com/en-US/live/projectnatal/

[3] http://www.ydreams.com/

[4] http://www.casadamusica.com/

[5] Winkler, T. (1995). Making motion musical: Gesture mapping strategies for interactive computer music. *Proceedings of the International Computer Music Conference*, 261-264.

[6] Povall, R. (1998). Technology is with us. *Dance Research Journal,* 30(1), 1-4

[7] Camurri, A., Mazzarino, B. and Volpe, G. (2004) Analysis of Expressive Gesture: The EyesWeb Expressive Gesture Processing Library, in A. Camurri, G. Volpe (Eds.), *Gesture-based Communication in Human-Computer Interaction*, LNAI 2915, Springer Verlag.

[8] http://www.palindrome.de

[9] http://www.troikatronix.com/isadora.html

[10] http://www.troikaranch.org/

[11] Guedes, C. (2006). Extracting Musically-Relevant Rhythmic Information from Dance Movement by Applying Pitch-Tracking Techniques to a Video Signal. *Proceedings of the Sound and Music Computing Conference SMC06*, Marseille, France.

[12] Naveda, L. and Leman, M. (2009). A cross-modal heuristic for periodic pattern analysis of samba music and dance. *Journal of New Music Research* 38 (3), 255-283

[13] http://www.inf.u-szeged.hu/~palagyi/skel/skel.html #Skeletonization

[14] Fujiyoshi, H., Lipton, A. J., and Kanade, T. (2004). Real-Time human motion analysis by image skeletonization. *IEICE Transactions on Information and Systems E Series D*, 87(1), 113-120.

[15] Chen, H. S., Chen, H. T., Chen, Y. W. and Lee, S. Y. (2006). Human Action Recognition Using Star Skeleton. *VSSN '06***:** *Proceedings of the 4th ACM international workshop on Video surveillance and sensor networks***,** October 2006

[16] http://www.iamas.ac.jp/~jovan02/cv/

[17] Canny, J. F. (1986). A computational approach to edge detection. *IEEE Trans Pattern Analysis and Machine Intelligence*, 8(6): 679-698, Nov 1986.

[18] Vitruvius, *De Architectura*: THE PLANNING OF TEMPLES, Book 3, Chapter I

[19] http://www.utaustinportugal.org/