A Database for Stylistic Human Gait Modeling and Synthesis

Joëlle Tilmanne, Raphaël Sebbe, and Thierry Dutoit

***Abstract*—This project aimed at recording a database which can be used as training data for the development of three-dimensional statistical models of various human gait styles. The three-dimensional motion data was captured using the IGS-190 motion capture suit from Animazoo, an integrated wireless motion tracking system equipped with inertial sensors. Fourty participants were recorded, walking at several speeds and performing several direction changes.**

***Index Terms*—motion capture, gait modeling, database.**

I. Introduction

Synthesis of realistic human motion is one of the greatest challenges in computer graphics, as classical animation methods are very time-consuming. A new trend in motion synthesis consists in using Machine Learning techniques to model and reproduce humanlike motion data. When using this kind of techniques, there is a need for representative data. This project is inscribed in this context, as its main goal is to record a database of three-dimensional human gait data, which could later be used for the training of statistical models of human gait and provide mechanisms to generate stylistic variations of these models.

In the next section, we give a brief introduction to the context of walk synthesis and character animation. In section 3, we introduce our database, and on what purpose it was built. In Section 4, we will give some information about the IGS-190 motion capture suit which is used for the recordings. Section 5 will explain the recording setup and instructions, and thus explain the content of the database. Section 6 will explain the format under which the data is represented. In the last section, we will make some remarks on the recorded data, and give some perspectives about the use of the database.

II. Animation of virtual characters

As of today, the only application where motion synthesis is truly used and is not only studied for research purposes is the animation field. Animation consists of the rapid display of a sequence of images, which creates an optical illusion of movement due to the human phenomenon of retinal persistence. The animation process is applied to various domains, ranging from the coarse motions seen in video games to the precise humanlike motions of 3D movies, and including fields like virtual reality or character animations for human-computer interactions. There are currently three main kinds of techniques which are used for motion synthesis in the animation field:

- keyframe animations (hand-driven),
- model-driven techniques in which a software based on a set of expert-based rules (i.e. the model) helps the animator for motion production,
- motion capture (data-driven technique).

Not that long ago, those methods were the only ones existing. Unfortunately, all of them have strong drawbacks. The solutions that exist for human character animation or synthesis of human motion in general are either time and labor consuming, or they are unrealistic and of poor quality.

For a few years now, a new trend has appeared. It consists of using statistical learning techniques to automatically extract the underlying rules of human motion, without any prior knowledge, directly from training on 3D motion capture data. Starting from the statistical models trained that way, new motion sequences can be automatically generated, using only some high-level commands from the user. Two movements generated by the same command (for example executing two consecutive steps in a walk sequence) will never be exactly identical. The result presents indeed a random aspect as can be found in the human execution of each movement, and becomes potentially more realistic than the repetition of the same capture sequence over and over. The movements produced that way are thus visually different, but are all stochastically similar to the training movements.

Bipedal locomotion is by nature a periodic time sequence of movements, which involve all body parts, with an underlying sequence of repeated phases. Following [1], human gait can be viewed as a dynamical process with a simple temporal structure (i.e., repetitions of basic stance and swing phases) but with several stylistic variations (i.e., walking, running, trudging, ...). Besides, the gait observations depict high variability that can be explained by the incapacity of a walker to produce perfectly twice the same movements and by the varying nature of its environment. All these elements create the naturalness that is so difficult to achieve in walking avatar animation. The modelling of such a process can be addressed by using probabilistic graphical models [2], and more especially Dynamic Bayesian Networks (DBN) [3] and Hidden Markov Models (HMM) [4], which provide mathematically tractable and computationally efficient solutions for their estimation.

III. Need for a database

In all Machine Learning techniques, the first major issue is to obtain enough representative training data. As only motion capture can give realistic 3D human motion data, it is the only way to obtain representative training data for statistical modeling of human motion. The first option is of course to use an existing database of captured motions. Several motion capture databases are available for free on the Internet, like the widely known Carnegie Mellon University (CMU) Graphics Lab Motion Capture Database [5], or motion provided for free by a society called Eyes, from Japan, referred to as "motion-data.com" [6], but in both cases the motions are only roughly described with a single definition word like "walk", "jump", etc. Some other sites sell the motions, like "Truebones" [7] but again, the description of those movements is very brief as they are aimed at a direct use in an animation sequence. There is thus almost no documentation on the recorded sequences of motion, on the people who performed them, or on the motion capture area. This makes it hard to compare walk sequences from different people as they are not all recorded in the same conditions, and the information about each recording is very sparse. The HDM05 database from Bonn University [8], [9] gives a little bit more information, as they have the same sequence played by five different subjects. The sequence is a series of different motions corresponding to instructions like walk 5 steps, turn right, walk 5 steps backwards, etc. But there are no speed variations nor extended study of direction changes, and the subdivision of the whole sequence into each separate class of motion is not made for style variations classes of walk.

For model training, we needed better documented data. Moreover, we are not interested in having as many different sequences as possible, but in acquiring enough examples of the same motion to train a model. If possible, we also want to take into account more precise parameters in the training, like the style of the motion, which is, along with other factors, linked to the subject performing it. This project gave the opportunity to massively collect human gait data with an accurate motion capture system. For each of the fourty subjects of the database, the recording conditions and instructions were the same, which enables us to truly compare the inter-people variations. One could then benefit from this valuable database to estimate complex statistical models. The data are collected for various gait conditions, with an inertial motion tracking system, the Inertial Gyroscopic System (IGS 190) from Animazoo [10].

IV. Motion capture equipment: the IGS 190

In most research, the three-dimensional nature of human gait is overlooked [11]. Observations are often reduced to measurements from the sagittal plane (e.g., video recordings from the side plane) and symmetry between left and right sides is assumed. From this limited perspective, much valuable information is lost. Observations can be improved by including measurements from other planes such as coronal (e.g., front plane) and transverse (e.g., top plane). However, human gait is ideally observed via three-dimensional measurements. To do so, various motion capture equipments can be proposed. All so called "motion capture" systems have the same goal, which is to record a real motion by transcribing it under mathematical form usable by a computer. This is achieved by tracking a number of key points through space across time, and combining them to obtain a tridimensional unified representation of the performance [12]. The subject's body is then most of the time represented under the form of a kinematic chain, whose root is situated at the hip level, and whose various segments represent a kind of simplified representation of the skeleton.

Several techniques can be used for motion capture. First, multiple camera views can be triangulated to estimate body marker positions. Such system requires fine calibration and robust computer vision algorithms to yield accurate data [13]. Marker-less optical solutions are more comfortable yet more complex [14]. Besides, mechanical solutions allow directly measuring joint angles but require intrusive sensors. In this project, we adopt the Inertial Gyroscopic System (IGS) developed by Animazoo [10]. The main advantage of this system is to deliver remotely accurate data in a unified manner with no extra solving and over large capture areas.

The IGS 190 motion capture suit contains 19 inertial sensors (InertiaCube3 from Intersense) consisting each of a 3 axis accelerometer, a 3 axis gyroscope and a 3 axis magnetometer. The data from those three sources are integrated and combined directly in the inertial sensor box, and angle data is thus provided straight from the sensor. Those angles which give the configuration of the inertial sensor are then postprocessed realtime trough the software provided with the suit. The output of the system is the set of angles between the body segments, at a rate of 30, 60 or 120 frames per second. In this database recording, the frame rate is chosen to be 60 fps (frames per second).

Although each of the sensors is wired to the main processing unit which is attached to the suit, the system is wireless as this main processing unit acts as emitter and communicates trough radio frequency with the receiver linked to the laptop trough USB connection. The wireless range of the suit is around 100 meter indoors and up to several hundred meters outdoors.

The AutoCal software enables the user to build a skeleton fitting any new actor wearing the suit with just two photographs taken in a given stance and within a 3D cube used for setting the scale of the picture. The actor will need to stand in the same stance, facing north, in the beginning of each motion capture session, in order to calibrate the system (which contains magnetometers).

There is no 3D position tracking system in the IGS suit, and the position of the actor is calculated by the software given a known initial position, using the length of the skeleton segments from the feet to the hip, the angles recorded between those segments for each frame, and always considering that the lower point of the skeleton is in contact with the ground.

Because of that, drifts can appear, but from our experiences with the equipment, if the calibration step is properly done, as well for the skeleton calibration as for the initial position of the actor at the beginning of each motion capture session, we obtain in most capture sequences very good quality of data.

V. Recording setup

This database contains 17 walk sequences for 40 subjects. Those 17 sequences correspond, for each subject, to rectilign walks at different speeds, and to segments with direction changes of various amplitudes in the trajectory. The subjects were participants of other projects of the workshop who agreed to volunteer for the motion capture recording.

The motion capture area was a 6m by 4m space, in which subjects were asked to walk given some instructions. The first step was to take the calibration pictures and to calibrate a skeleton for each participant. Once this was done, each subject came for approximately one hour of recording. They were first asked to remove their shoes and put on the motion capture suit. The walk in the database is thus always barefoot, on a perfectly even floor.

The first part of the motion capture session was the recording of the trajectory changes. Several crosses were drawn on the floor, as shown in figure 1. For each one of the five direction change sequences, the performer began on the red cross on the left upper side of the figure, facing right. The instruction was to walk straight until they reached the red line (which was also drawn on the floor), and then they were free to change direction the way they wanted to, they just had to reach one of the other crosses drawn on the floor, with different angles from the initial straight trajectory. For the first sequence, the cross to reach was just straight ahead (0 degree), and for the following ones, the angles of the direction change were respectively 45, 90, 135 and 180 degrees.

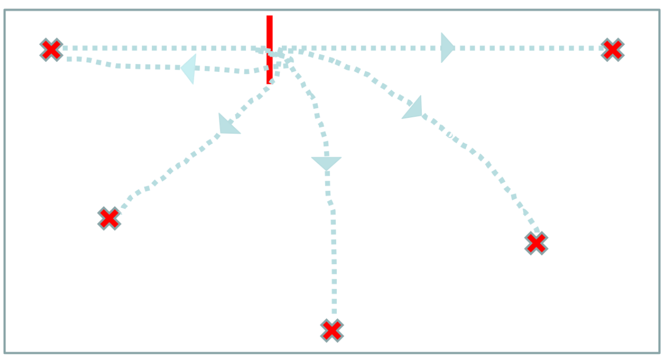

Fig. 1. Sketch of the motion capture area for the five direction change recordings. Only the red crosses and line were drawn on the floor.

For the second part of the motion capture, the aim was to study how walk changes when we change our velocity. This time, the performer was told to walk straight from one corner of the capture area to the opposite corner (approximately 7 meters). He was given four different instructions, and the sequence corresponding to each of these instructions had to be repeated three times. Those four instructions were the following:

- Walk (normally).
- Imagine you are in a park, the weather is nice, you take your time.
- You are going to work, but you are not late.
- You have to catch a train, and you are very late, but you are not allowed to run.

These instructions let to the performer the freedom to choose his speed for each scenario, even if the gradation from slower to faster was the same for everybody, with the first "walk normally" instruction as a reference for a standard self selected speed.

VI. Data representation

The output of our motion capture system are angles between body segments, and the calculated 3D position of the root (hips) which is evaluated from the angles and the segment lengths of the legs, as explained previously(section IV). That data is recorded under the form of a Biovision file format (.bvh), which is a "de facto" standard format for 3D motion data. A bvh file contains two parts. The first part is the hierarchical description of the skeleton, starting from the root (hip) to the extremities of each limb, and the geometric position of each joints, still standing. The second part of the file contains one line of data for each motion frame. The skeleton from the Animazoo software contains 20 body segments in the skeleton (see figure 2), and each data line contains thus 3 (3D position of the root) + 3×21 (3D angles for each segment and for the whole body) = 66 values. Three of the segments are in fact only added to make the skeleton look closer to the real skeleton but have no degree of freedom in the motion capture, and their value does thus not vary during motion capture. There are thus finally 57 values to analyze and model in our data.

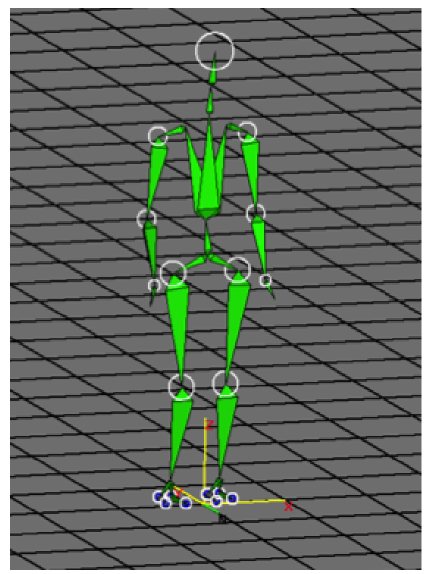

Fig. 2. Skeleton calibrated and displayed in Animazoo's software during motion capture.

As we have said in section IV, drifts can appear when using such an inertial motion capture system and it is the case for a few motions of the database, where the 3D position is not properly calculated. The appearance of the walk is then still good when looked at without 3D displacement, but the whole body looks like if instantly drifted backwards. Another small disturbance that can sometimes be observed is when one of the sensors moves because of the suit and not because of the subject's motion.

VII. Data visualization and segmentation

The project has a need for a tool to manually segment the motion recordings to keep only those parts needed by the training process. In this context, Motion Editor.app was developed, which consists of a 3D view of the skeleton and an interactive timeline. Moreover, this application allows the labeling of time segments that can be used for specific training purposes.

The software parses the Biovision file (.bvh) and builds a memory representation of its content. This includes both the skeletal representation, i.e. the relative offsets of the joints, and the time sampled 3-axis rotations of each joint [1]. Then, it displays the skeleton at any point in time as an Open Scene Graph [2] 3D scene view.

A timeline is also provided, which enables time navigation with immediate display of the skeleton. Additionally, the timeline permits to select temporal segments, and associated motion frames can be deleted. Alternatively, the time segment can also be given a label for identification (and further processing). The edited, in-memory frame representation can then be saved as a new .bvh file, and the labels, if any, can be saved to a separate labels file.

The graphical user interface of Motion Editor is shown of figure 3.

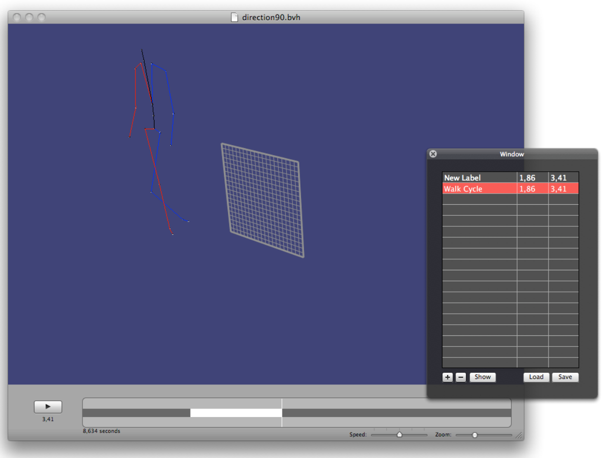

Fig. 3. Sketch of the motion capture area for the five direction change recordings. Only the red crosses and line were drawn on the floor.

VIII. Conclusion

The flexibility of the IGS 190 equipment enabled the recording of a large database of human walk and its variations. A preliminary visual analysis of the database already allowed us to observe the obvious style differences between walk from a woman and a man, the way that a person walks at different speeds, as well as the differences in walk parameters for two subjects with comparable morphology. From now on, this new database will allow us to study inter-subject variability in walk, walk variations linked to speed, and the way humans manage direction changes when they walk on an even ground without obstacles, and to train statistical models to model and to synthesize new motion.

[1]this 66 DOF record is called a frame

[2]Open Scene Graph is an open source library based on OpenGL use to display 3D content

Acknowledgment

The authors would like to thank all eNTERFACE'08 attendants who participated in the creation of the database.

This project was partly funded by the Ministry of Région Wallonne under the Numediart research program (grant $N^0$716631). J.Tilmanne receives a PhD grant of the Fonds de la Recherche pour l'Industrie et l'Agriculture (F.R.I.A.), Belgium.

References

[1] M. Brand and A. Hertzmann, "Style machines," in *Proc of SIGGRAPH 2000*, 2000, pp. 183–192. [Online]. Available: http://www.merl.com/reports/docs/TR2000-14.pdf
[2] S. L. Lauritzen, "Graphical models," 1996.
[3] Z. Ghahramani, "Learning dynamic bayesian networks," in *Adaptive Processing of Sequences and Data Structures*. Springer-Verlag, 1998, pp. 168–197.
[4] L. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition," in *Proc. of IEEE*, vol. 77, no. 2, February 1989, pp. 257–286.
[5] "The Carnegie Mellon University Graphics Lab Motion Capture Database," http://mocap.cs.cmu.edu.
[6] "The "mocapdata" database," http://www.mocapdata.com.
[7] "Truebones website," http://www.truebones.com.
[8] "The HDM05 motion capture database," http://www.mpi-inf.mpg.de/resources/HDM05.
[9] M. Müller, T. Röder, M. Clausen, B. Eberhardt, B. Krüger, and A. Weber, "Documentation mocap database HDM05," Universität Bonn, Tech. Rep. No. CG-2007-2, June 2007. [Online]. Available: http://www.mpi-inf.mpg.de/resources/HDM05/07_MuRoClEbKrWe_HDM05.pdf
[10] "Animazoo inertial motion capture system IGS-190," http://www.animazoo.com.
[11] C. L. Vaughan, B. L. Davis, and J. C. OConnor, *Dynamics of Human Gait, 2nd ed.* Kiboho Publishers, 1999.
[12] A. Menache, *Understanding motion Capture for Computer Animation and Video Games*. San Francisco, CA, USA: Morgan Kauffman Publishers Inc., 1999.
[13] B. Rosenhahn, T. Brox, and H.-P. Seidel, "Scaled motion dynamics for markerless motion capture," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2007.
[14] A. Colahi, M. Hoviatalab, T. Rezaeian, M. Alizadeh, and M. Bostan, "Implementation of optical tracker system for marker-based human motion tracking," in *Proceedings of the IASTED International Conference on Applied Simulation and Modelling (ASM)*, June 2006.

Joëlle Tilmanne holds an Electrical Engineering degree from the Faculté Polytechnique de Mons (Belgium) since June 2006. She did her master thesis in the field of sleep signals analysis, at Lehigh University (USA). She is pursuing a PhD thesis in the TCTS (Circuit Theory and Signal Porcessing) Lab of FPMs since september 2006, in the field of HMM based motion synthesis.

Raphaël Sebbe holds a PhD in the field of image processing as well as electrical engineering degrees from both FPMs and Supélec. Raphaël has worked on speech and image-related technologies and software development. He is now with TCTS Lab of FPMs, pursuing research on Numédiart project.

Thierry Dutoit graduated as an electrical engineer and Ph.D. in 1988 and 1993 from the Faculté Polytechnique de Mons, Belgium, where he is now a full professor. He spent 16 months as a consultant for AT&T Labs Research in Murray Hill and Florham Park, NJ, from July, 1996 to September, 1998. He is the author of two books on speech processing and text-to-speech synthesis, and the coordinator of the MBROLA project for free multilingual speech synthesis. T. Dutoit was an Associate Editor of the IEEE Transactions on Speech and Audio Processing (2004-2006) and a member of the INTERSPEECH07 organization committee. He was the initiator of eNTERFACE workshops and the organizer of eNTERFACE05.