Considerations for aggregation of motion-captured files in structured databases for DHM applications

A. KEYVANI*†§, D. LÄMKULL‡, G. BOLMSJÖ§ and R. ÖRTENGREN†

† Product and Production Development Department, Chalmers University of Technology, Gothenburg, Sweden
‡ Virtual Methods & IT, Volvo Car Corporation, Dept. 81620/PVÖS32, Gothenburg, Sweden
§ Department of Engineering Science, University West, Trollhättan, Sweden

Abstract

One way of enhancing motion simulation in digital human modeling (DHM) tools is to use data-driven methods which are based on real motion data. In spite of the availability of motion captured datasets which are offered for different purposes by commercial and research institutes, aggregation and integration of these motions in a unified and structured database system is not straight forward. Lack of this integration, limits the availability of existing data and causes DHM tools not to be able to use the data efficiently. Also for the researchers, comparison and analysis of data get very hard if not impossible. When searching for a specific motion pattern, it is optimal if the stored data in the database can be directly compared, analyzed and then retrieved if necessary. This study highlights several sources of incompatibility among motion capture files which shall be considered when implementing a comprehensive data management system for manipulating motion captured data. Subsequently, these incompatibilities are analyzed in more detail and necessary considerations and possible solutions are proposed in order to overcome the integration obstacles.

Keywords: Motion Database, Standardization.

1. Introduction

There are still major challenges for generating natural looking motions of human's daily tasks when using existing digital human modeling (DHM) tools (Artl and Bubb 2000; Raschke et al. 2005; Lämkull 2009). Data-driven methods which are utilizing motion data of real humans are one promising way to overcome the problem (Chaffin 2005). Such methods often need to use human motions which are generated by motion capture systems and stored in comprehensive databases (Reed et al. 2006). In spite of the availability of motion captured datasets which are offered for different purposes by commercial and research institutes, aggregation and integration of existing data in a unified and structured database system are not straight forward (Keyvani et al. 2011). Big sector of existing motion files belongs to game and animation industry, where the direct users of the data are still human operators (animators and game developers), and not software platforms (DHM tools). Therefore an automated system of motion selection and modification was irregularly developed since selection, adjustment, and fine tuning of motions were mostly done manually in these industries. In these cases, the motion files are mostly kept in big datasets (and not databases) which are folders of many motion files named by the motion content and are manually organized. On the other hand, direct use of motion data in DHM tools requires proper indexing of motions for automatic selection and retrieval purposes.

In the absence of a uniform solution to collect motion capture data from different sources in one database, DHM tools will not be able to gain the most and use these data efficiently. Also exchange of motion data, analysis results and simulations between researchers especially among multi-disciplinary areas is very hard at the moment (Paul and Wischniewski 2012). This study highlights several sources of incompatibility among motion capture files which shall be considered when implementing a comprehensive motion-data management system. Subsequently, each of these sources is analyzed in detail and necessary considerations and possible solutions are proposed in order to overcome the incompatibility issues.

*Corresponding author. Email: ali.keyvani@chalmers.se
2. Method

In order to identify, analyze and distinguish incompatibilities among motion files, different sources of information are looked through. These sources include reviewing the literature, checking different vendors and file formats, studying existing standards and inter-disciplinary conventions, interviewing specialists and end-users. Furthermore, to test and face the identified issues practically, an integrated database platform is implemented and filled with motion capture data from different available sources (Keyvani et al. 2012). New motion parsers are developed and current existing solutions are compared. The identified problems then discussed by looking into their current situation, available versions, existing standards/conventions, and by proposing converting methods to neutralize data and improve the integration level.

3. Results

3.1. Differences in file formats

Normally the motion capture vendor is the determinant to use a specific file format. While different vendors provide their own file formats, it is unlikely to convince them to stick to only one standard format. HTR by Motion_Analysis (2012), ASF/AMC by Acclaim, BVA/BVH by Biovision, and V/VSK by Vicon (2012) are examples of commercially available vendor-based Mocap file formats. For the Mocap end-users such as animation software tools, usually the functionality to import most of these formats is integrated in the animation software package. Therefore, at least in short term, the solution for integrating these formats in a motion database is to develop standard format converters and import all captured files into the database by using just one generic structure.

In terms of standards, X3D format presented in ISO/IEC 19775 in a larger scope deals with software system that integrates network-enabled 3D graphics and multimedia. Also for the 3D humanoid figures, H-Anim standard, ISO/IEC 19774 (2006) exists. However further standard establishment is needed in order to specifically manipulate Mocap data uniformly. A standard set of widely accepted CAD formats and simulation data formats integrated within each DHM tool and allowing transition of data from one tool to another in a seamless manner is a requirement urged by DHM tool users (Lockett et al. 2005; Wegner et al. 2007).

3.2. Naming Conventions

Motion capture systems are used widely in number of applications: biomechanical-based applications such as gait analysis, animation-based applications such as computer games and animation movies, and industry-based applications such as human simulation in virtual production lines. This diversity of applications and the fact that they are originated from different science groups lead to a chaos of conventions on naming of the bones, joints and body landmarks. Although there exist some guidelines about how to name different elements in a motion capture setup, still there is a long way toward a global agreement (Menache 2000). In other related areas such as computer graphics, numbers of established standards exist which can help handling of Mocap data in a more uniform way. H-Anim (2006) and ISO15536-1 (2008) standards are examples of these works toward standardization of humanoid figures. A temporary solution is to use mapping tables that connects corresponding names in each Mocap file to a reference table in the database. However, this mapping need to be often performed manually which is time consuming, not automated and vulnerable to errors.

3.3. Differences in skeleton configurations

In the context of digital human modeling, a common element in all articulated bodies is a body skeleton (often called a rig in animation terminology). A body skeleton is a set of hierarchical rigid segments connected by joints which actualize human body movements. In this regard, having fixed segment lengths, a motion can be defined by series of changes in the joint angles deciding the orientation of each segment relative to its proximal segment. The configuration of these hierarchical structures is not uniformly determined when it comes to virtual humans (Maciel et al. 2002; Guo and Li 2004; Song et al. 2011). In some areas of the body such as shoulder, the joint configuration is much more complex than to be modeled by a simple ball joint system. Therefore even if a simplification is not intended there is no single solution which can model the joint (Divr and Berme 1978; Engin 1980; Maurel and Thalmann 2000). In all these models, at least three distinguished joints, Acromioclavicular (AC), Sternoclavicular (SC), and Glenohumeral (GH) are considered for the shoulder complex. However if a simpler model is desirable then it is common to use a skeleton with only two joints (AC and GH) describing the shoulder movement.

Also some simplifications are usually considered in modeling spine. Based on the complexity of the model, this can be as simple as two joints or as complex as 24 joints with three degrees of freedom on each. While these joints are normally moving together, DHM tools normally use some dependency relationships in order to decrease the number of DOFs in the model. Another part where differences in modeling occur is leave nodes such as fingers, eyes and toes. However, this is rather an easy problem to solve as the motion of these nodes can be easily added on.
top of existing model without further recalculation

In order to store these motions with different
skeleton configurations in one unified database,
there is no single off the shelf solution available.

Often there is a need to retarget the skeleton which
is migrating from one skeleton configuration to
another (Gleicher 1998). Some of retargeting
methods are limited to migrating between human
models (Beurier and Wang 2004), while others can
retarget a motion to a new character which is not
necessarily human like (Hsieh et al. 2005). Among
these methods, the one proposed by Monzani et al.
(2000) which is using an intermediate skeleton to
retarget one rig into another can be used for the
purpose of this study. To do so, a standard skeleton
with a pervasive joint definition (e.g. H-Anim)
which includes all available joints in different
motion files is first predefined in the database. Then
each new motion is first retargeted to this standard
skeleton and then is stored in the database.

H-Anim standard (2006) also suggests considering
models with different level of articulation (LOA).
In this approach, based on the application a model
with a LOA differs from 0 to 4 is defined. Each
LOA includes a certain number of joints in the
model.

3.4 Inconsistency in marker placement and
marker-to-joint calculation

Many of motion capture systems are marker-based
(either optical or magnetic). This means that the
position of joint centers and orientation of body
segments has to be calculated based on the tracked
position of these markers which are attached on the
skin. Numbers of inconsistencies exist among end
users when it comes to practice:

First, the number and placement of markers are
different based on Mocap equipment vendor and
also research groups. Each vendor normally
suggests numbers of marker set templates which
can be different from one application to another.
For example in gait analysis two commonly used
marker sets are Helen Hyes (HHS) marker set
(Kadaba et al. 1990), and Cleveland Clinic marker
set (Sutherland 2002). The HHS does not use static
trials to define joint centers, but instead uses
anthropometric measures of the joints. Castagno et
al. (1995) have compared them in detail and
concluded minor differences in the result.

Second, there is no agreement how to calculate joint
centers from these marker sets. Many different
methods are suggested in the literature in order to
calculate joint centers (Holzreiter 1991; Gamage
and Lasenby 2002; Kirk et al. 2005; De Aguiar et
al. 2006; Xiao et al. 2009). Two common
techniques recognized as local and global
optimization are compared by Silaghi et al. (1998)
and a survey accomplished by Ehrig et al. (2006)
has also compared and classified many of these

methods. In addition, when using the magnetic
sensors, Mocap system receives both position and
orientation in comparison with optical markers
which are determining just the position. As a result
magnetic motion capture systems are using slightly
different methods than the ones used in optical
systems to calculate joint centers (O'Brien et al.
1999; Ringer and Lasenby 2004).

Third, sometimes the optical marker itself is used
directly on the output motion and this can lead to
confusion when a proper naming is not employed.
For example a marker named “hand” can be
mistaken for a marker in the palm of the hand, for a
marker on the back of the hand, or for a marker on
tip of index finger.

In spite of existing of all these methods, in most
cases these calculations are integrated as a solver
plugin inside each vendor’s Mocap software and
they are closed to the user access. As a result, using
the same skeleton configuration the output motion
can be different from one place to another.

3.5 Mismatches in joints’ degrees of freedom

In general, a segment in the space can be fully
defined by six parameters (three translational and
three rotational). For articulated bodies it is wise to
also consider a length factor which is determining
the changes in length of each segment through the
animation. However, leaving a set of hierarchical
segments with 7 degrees of freedom (DOF) for each
segment, one has to face huge amount of
complexity and computational redundancies. Most
of the Mocap systems have the ability to constraint
one or more of these DOFs during the marker to
joint conversion phase and through custom
definition of skeleton configurations. The first
common simplification is to constraint translational
DOFs for all the segments except the root segment
which needs to have translational DOFs in order to
move the whole body in the working space. Also
for a normal application which is not dealing with
bone deformation and fatigue issues, considering
constant bones lengths is a logical decision. In
practice, the main cause for variations in segment
length during a capture session is skin marker
displacements.

In addition to these general constraints, joint-
specific constraints can also be applied to the
skeleton definition. A good example is to apply
rotational constraints on the knee joint in order to
contain just one degree of freedom. Existence of
these types of joint specific constraints can greatly
help the solver engine to resolve redundancies
while generating a biomechanically acceptable
motion. However, disagreements can happen in
certain joints between DHMs on how to model the
body. For example, the elbow-wrist complex can be
modeled in at least two different ways: giving one
DOF to elbow and three DOFs for the wrist; e.g. in
Jack (Badler 1997) against assigning two DOFs for
the elbow and two for the wrist; e.g. in IMMA (Hanson et al. 2010). As a result, it is important how to handle these mismatches when accumulating motion files in a database. For normal usages, general constraints of constant bone length and no translation on child nodes are applicable for a wide range of motions. If a motion did not meet these requirements, the average bone length which is calculated from total motion frames can be substituted for each segment length and translational data can be just ignored if the values are within a range. Otherwise, reconfiguration of skeleton and recalculating of rotational data based on the new configuration are needed. For joint specific constraints, ISO 15536-2 (2007) standard has some guidelines on how to choose joints’ DOFs. In cases of incompatibilities, if transferring one joint DOF to another joint does not change the rotational sequence for those joints, mapping tables are valid solutions (eq.1). Otherwise, a recalculation of rotational data for both proximal and distal joints is needed (eq.2).

\[
R_{\text{proximal}}^{XZ} \ast R_{\text{distal}}^{XY} = R_{\text{proximal}}^{X} \ast R_{\text{distal}}^{ZX} \quad (1)
\]

\[
R_{\text{proximal}}^{XY} \ast R_{\text{distal}}^{X} \neq R_{\text{proximal}}^{X} \ast R_{\text{distal}}^{ZX} \quad (2)
\]

Where \(R_{\text{proximal}}^{XZ}\) represents rotational transform of the proximal segment with Euler sequence of \(X\) then \(Z\).

### 3.6. Double definition for using virtual joints

Fixed bones or virtual joints are the bones/joints that have no degree of freedom. The reason for defining such a joint is usually to create a constant offset relative to the parent joint without using any translational data in the transform matrix. Examples of this type of joints can be seen right/left hip joint. Two different methods can be used to move from parent joint (Root) to the child joint (Left Hip). First method is to create a transform matrix \(M\) which is translating ‘Root’ to the ‘Left Hip’ (\(T\)) and then rotating it 180 degree around X axis (\(R_x\)) (eq.3).

\[
\text{Root} \rightarrow \text{Left Hip} \\
M = T(0, -l \sin \theta, -l \cos \theta) \ast R_x(180) \quad (3)
\]

Second way is to define a ‘Dummy’ segment with the length of \(l\) which is connected to ‘Root’ and defined with transform \(M_1\). Subsequently, ‘Left Hip’ is redefined as a child of ‘Dummy’ segment and with transform \(M_2\) (eq.4).

\[
\text{Root} \rightarrow \text{Dummy} \rightarrow \text{Left Hip} \\
M_1 = R_x(90 + \theta) \text{ and } M_2 = R_x(90 - \theta) \quad (4)
\]

#### 3.7. Differences in rotation order conventions

It is very common in both motion capture files and DHM tools to use Euler convention to define rotation angles and orientation of segments in the space. For defining object orientation in the space other methods such as rotation matrices, vector-angles, and quaternions also exist (Diebel 2006). However, there are numbers of good reasons for choosing Euler angles compared to other methods:

- Euler angles are compact. Maximum three numbers are needed to describe any orientation in the space.
- Euler angles are easy to understand.
- Euler angles are widely used in many applications.
- And Euler angles are biomechanically self explanatory if suitable sequence order is chosen.

At the same time problems are also inherited when using the Euler angles:

- Euler angles suffer from gimbal lock.
- Same name can represent 12 different conventions (different orders).
- Euler angles are generally non-commutative.
- Euler angles can be problematic in motion interpolations.

In the subject of motion capture files, one issue is that some of file formats (e.g. HTR) only allow defining one global rotation order which is then propagated to all joints. This global order is normally stated in the file header and can be any of three combinations of non-repeating axes. Although this structure simplifies the calculations, in many cases it decreases the clarity of underlying motion. Choosing customized rotation order per joint in other supported formats gives the opportunity to the user to define more meaningful rotation sequences from biomechanical point of view. For example, international society of biomechanics’ (ISB) general recommendation is to choose the last rotation around an axis which is fixed to the distal segment (Wu et al. 2002; Wu et al. 2005) while choosing the ZXY convention when reporting global reference frame (Wu and Cavanagh 1995).
In general, researchers in clinical applications are recommending different rotation orders for different joints while this is not the case in animation and game applications because the end-users in those disciplines are rarely dealing with exact joint values in absolute numerical format. As a result, when using motion capture data for DHM applications in most of the times a conversion to the manikin’s joints specific conventions is needed. A simple solution is to transform the rotation from Euler angles to matrix format and convert it back to Euler angles but with new rotation order. This is rather an easy but computationally costly operation (Slabaugh 1999).

3.8. Mismatches in base pose definitions

Base pose, key pose or neutral pose are common terms but with no consistent meaning in the motion capture field. They can refer to a specific posture which is needed for the marker to joint solver engine in order to initialize optimization algorithm. This is usually a standing posture with arms opened to the sides almost 90 degrees (called T-Pose). The palm of the hands can be either facing down or forward. The terms can also refer to a specific posture in the animation usually named key pose which is a posture important for a specific scene when key framing the motion; e.g. see Yamane et al. (2010). In this case the key pose can be any possible posture from the motion. Finally, they can represent a standing posture with arms beside the body in a relaxed position when all the joint values are zero in their local joint coordinate system (JCS). No matter what definition is used to introduce the base pose, it is important that how it is treated in the motion capture files. Most of Mocap software packages have the possibility to consider a frame or any specific posture as a base pose and recalculate the rotational data with respect to this base pose. It means that each segment’s global orientation \( R_i^{G} \) is derived by applying its parent global orientation \( R_{i-1}^{G} \) followed by segment’s base orientation \( R_i^{B} \) and followed by current segment local orientation \( R_i \) (eq. 5).

\[
R_i^{G} = R_{i-1}^{G} \times R_i^{B} \times R_i
\]

The main benefit of presenting rotational data using a base pose is to create a meaningful reference posture which all the rotations are measured from it. However, when reporting motion data, there is no agreement between researchers in different disciplines on which posture shall be the zero reference for joint angles.

3.9. Gravity axis convention

Some applications prefer -Y direction to demonstrate gravity direction while others use -Z direction as gravity axis. While importing the motion to the database, if the file format stores which axis is used as gravity axis then it is easy to convert otherwise it is tricky to recognize it by just looking at the numbers. Keeping the X axis fixed, changing the gravity axis from Y to Z will replace \( Z \rightarrow Y \) and \( Y \rightarrow -Z \) and vice versa. There is no agreement between standards regarding the gravity direction relative to global coordinate system. For example ISO 15536-2 (2007) standard suggests Z axis while H-Anim (2006) standard suggests Y axis for up direction.

3.10. Differences in measurement and scaling units

Simple but important, motion capture files can come with different units both for length and angle. Before importing the data proper conversion of the scales shall be considered. In some formats such as HTR, a global scale factor exists in the header which has to be multiplied into translational data and segments lengths. Common length units are mm, cm, inch, and meter while common angle units are degrees and radian.

3.11. Differences in joint coordinate system

A hierarchical articulated body requires local coordinate systems (LCS) on each joint in order to solve the forward kinematic equations. Having these joint coordinate systems (JCS) defined, not only are the orientation and translation of each child joint then determined in relative to their parent but also the orientation of the distal segment (bone direction) is defined in corresponding to JCS. Although not necessary but it makes much more sense to define JCS in relation to a certain feature on the body such as plane of rotation, sagittal-coronal-transverse planes, bone direction, center of pressure, etc. (Grood and Suntay 1983; Wu et al. 2002; Wu et al. 2005). Furthermore, some restrictions can be imposed by the motion capture file format. HTR file format forces the solver engine to have one axis always at the same direction of next segment bone. This axis can be optionally chosen to be Y or Z in most of the applications. If the bone direction is fixed on one of JCS axes then only the bone length is required to define it while in a more generic file format such as AMC, both bone length and bone direction are needed.

4. Discussion

Agreeing on a single global standard is an ultimate solution but seems an unrealistic approach. Diversity in applications, vendors, equipment and formats are the hardware obstacles toward standardization. In addition, inter-disciplinary and cross-disciplinary disagreements in data presentation and result exchange are the soft obstacles against uniform aggregation of motion data.

Since the need for having a comprehensive motion database is increasing, the identified problems and proposed solutions in this study can help the
researchers to construct integrated database platforms to employ other available sources of motion data in their implementations. A DHM database architecture or framework would bring great benefit to industry. The architecture would provide a common human model data structure and interface that researchers and commercial DHM companies would utilize it to greatly reduce the amount of development required to migrate research to commercial DHM tools and industry developments from one DHM tool to another. The architecture would include a standard simulation data format for the simulation data transfer between different motion databases.

It is shown that in many cases a locally developed mathematical function can solve the incompatibility issues with the cost of additional computation. However, as these additional computations are only needed once when the motions are imported into the database, the solution is promising since it will greatly affect the database performance for future usages because all the motions which are imported to the database are generic and directly comparable afterwards.

5. Conclusion

DHM tools have improved significantly over the past 10 years. What has not improved is interoperability and consistency across tools. Recent developments have refocused the work on integration of real motion data into DHM tools. Standardization of motion databases and future improvements will benefit the DHM community (industry user, researcher and vendors) since it would result in decreased effort and increased consistency for simulation purposes. To successfully integrate motion databases, cross functional industry and research teams need to work cooperatively to scope work, perform research and transfer results for development into commercial tools.

References

Ali Keyvani, Considerations for aggregation of motion-captured files in structured databases