Functional Body Mesh Representation - a Simplified Kinematic Model

Przemysław Skurowski* and Magdalena Pawlyta†

*Institute of Informatics, Silesian University of Technology, Gliwice, Poland
†Polish-Japanese Institute of Information Technology, Poland

Abstract. This paper describes *Functional Body Mesh* a simplified representation of a body kinematic structure intended to form a framework for marker-wise signal processing of recorded motion trajectories of motion capture of virtually any vertebrates. The parent-child and sibling relationships are inferred on a coherence of movement and constancy of distances. There is proposed a GMM based method for inferring FBM segments from the raw motion capture data based on a relationship between the points. For creating groups representing specific body parts we propose an incremental multicriterial clustering algorithm employing Gaussian mixture models. To infer body parts hierarchy we propose utilizing a consensus method.

Keywords: Motion capture, Body analysis, Clustering, Gaussian Mixture Model
PACS: 07.05.Rm, 06.30.Bp, 07.05.Kf, 07.05.Tp

INTRODUCTION

Optical motion capture (MoCap) systems [1] register trajectories of tracking points - markers which are organized for further processing needs into some object model [2] during the initialization stage. It is common for motion processing algorithms [3] to use arbitrary human model - manually pre-edited skeletal structure (with limb lengths and joint locations) tightly [4] connected to the predefined marker locations (mesh).

Prior to the obtaining skeletal animation it is usually necessary to process the markers recorded positions. In this stage, there are used such filters as denoising and occluded marker reconstruction. The markers are processed either independently (e.g. [5]) or there is required reference to the estimated skeleton [6]. Skeleton free processing, utilizing rigid-body mesh of markers is also possible, but in a pipeline of a typical commercial software it is limited by the arbitrary (human) model assigned or tuned to the markers [4]. In a mocap software, to create of a new subject model or atypical topology of markers it is usually heuristic requiring some manual work and user experience.

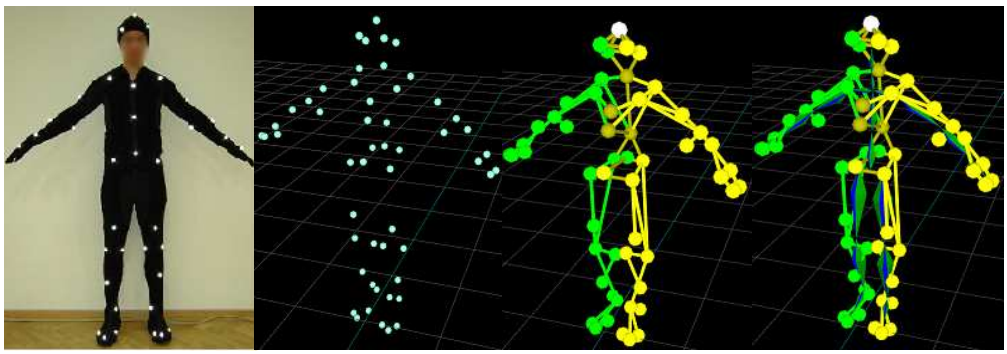

FIGURE 1. Motion capture pipeline: an actor in A-pose, raw markers, arbitrary mesh, skeleton assigned to the mesh

The principal goal of the work was to propose a representation of the subject's kinematic structure almost free of prior assumptions on the body structure and thus allowing to adopt easily to virtually any vertebrate. The secondary goal was that the representation should be easy to obtain with a little user interaction. The concept of a body representation in *functional body mesh* (FBM) can be considered as an intermediate between skeletal and marker form. It comprises of a tree of partial submeshes reflecting the body structure, where the submeshes represent specific body parts and the tree representing hierarchy. Such an approach would allow to process the mocap data using all the information in raw xyz marker trajectories (without aggregating into skeletal animation) whereas it would also

allow to use the knowledge of the body hierarchy. The representation form a kind of framework for marker-wise signal processing.

The FBM representation can be employed in two basic ways: straightforward in current software as a simple visualization method for a new kinds of subjects and as a preliminary step for further marker-wise processing like motion prediction, classification, filtering and skeleton inferring where the resulting mesh model should be consistent analytically and morphologically with the body structure. To the authors knowledge the most of current body structure analysis methods focus on parameterizing of the skeletal structure inferred on the body motion (e.g. [7]), of which some, employing a local scheme, share the stage of marker segmentation with our approach. Another approach, similar to ours, a concept of groups of correlated markers was employed to recover gaps in mocap recordings [8].

THE INFERENCE METHOD

The inference method for an articulated body structure we propose is intended for the basic and most common area - optical motion capture with sparse tracked markers (features), attached to the body. It comprises of three basic steps:

1. sorting markers in special order (top-down/center-boundary - TDCB) to achieve meaningful body hierarchy,
2. clustering of siblings (using GMM based thresholds) into groups representing specific body parts,
3. hierarchy recovery by selection of a parent for each of marker groups using consensus method.

It is based on the corollaries of a rigid body assumption - coherence of movement and constancy of distances:

- Siblings markers are located on common body parts: 1) they move together so their movements (gradients) are similar; 2) they are located on rigid parts so their relative distance is constant.
- Each sibling group has a single parent group that: 1) is located in another group; 2) cannot be located in a child node (no loops); 3) has a closest and constant distant single point to the group, 4) the sibling and parent body parts are connected so distances cannot vary very much.

For the measurement of a distance we used simple Euclidean distance. Constant distance over the whole sequence was verified for each pair of markers statistically as a range of values. To filter out noise existing in records we used the inter-quantile distance between lower (L=0.5$^{\text{th}}$) and upper (U=99.5$^{\text{th}}$) percentile (P) of the whole N frame sequence.

$$D^E_{A,B}(n) = \sqrt{(x_A - x_B)^2 + (y_A - y_B)^2 + (z_A - z_B)^2}, \quad R^E_{AB} = P_U\left\{D^E_{A,B}(1..N)\right\} - P_L\left\{D^E_{A,B}(1..N)\right\}. \tag{1}$$

The movement of the k^{th} marker is well described with a gradient: $\Delta_k(n) = [\Delta_x, \Delta_y, \Delta_z] = [x_n - x_{n-1}, y_n - y_{n-1}, z_n - z_{n-1}]$. Similarity of gradients was measured using weighted cosine distance (D^c)and coherence in the sequence was calculated as a range of values. In order to filter out the noise we used the interquantile distance between the lower (L=1st) and upper (U=99th) percentile of the whole sequence :

$$D^c_{\Delta_A,\Delta_B}(n) = w\,(1 - \cos(\Delta_A \Delta_B)) = w\left(1 - \frac{(\Delta_{Ax}\Delta_{Bx} + \Delta_{Ay}\Delta_{By} + \Delta_{Az}\Delta_{Bz})}{|\Delta_A| + |\Delta_B|}\right), \; R^c_{AB} = P_U\left\{D^c_{\Delta_A,\Delta_B}(1..N)\right\} - P_L\left\{D^c_{\Delta_A,\Delta_B}(1..N)\right\} \tag{2}$$

The use of weighting of a simple cosine distance with a length of movement is intended to solve the problem when markers placed on the same body part have small non-consistent (opponent) movements due to deformation caused by the elasticity of the human body - like chest during breathing or foot flattening under weight of a body.

$$w = w(\Delta_A, \Delta_B, n) = 0.5 \cdot (|\Delta_A(n)| + |\Delta_B(n)|) \tag{3}$$

where: $n, n-1$ number of two successive frames, x_A, y_A, z_A, x_B, y_B, z_B - coordinates of points A, B for n^{th} frame.

Inter-marker relationship is determined by assigning of their pairs into classes according to the aforementioned rationales. One can intuitively assign each pair into one of four different classes: peer, close, independent and opponent ($c_1..c_4$). The distance functions described above should reflect these classes so one should consider a multimodal probability distribution function (PDF) for both distance functions. Gaussian Mixture Models [9] (GMM), combining two or more Gaussian probability distributions (g_i): $G = \sum_{i=1}^{C} g_i = \sum_{i=1}^{C} w_i\, N(\mu_i, \sigma_i)$, where: C number of modes (c_i), w_i - weights, $N(\mu_i, \sigma_i)$ normal probability distribution of a μ_i, mean value and σ_i std deviation; estimated with expectation maximization, appeared to fit the data well.

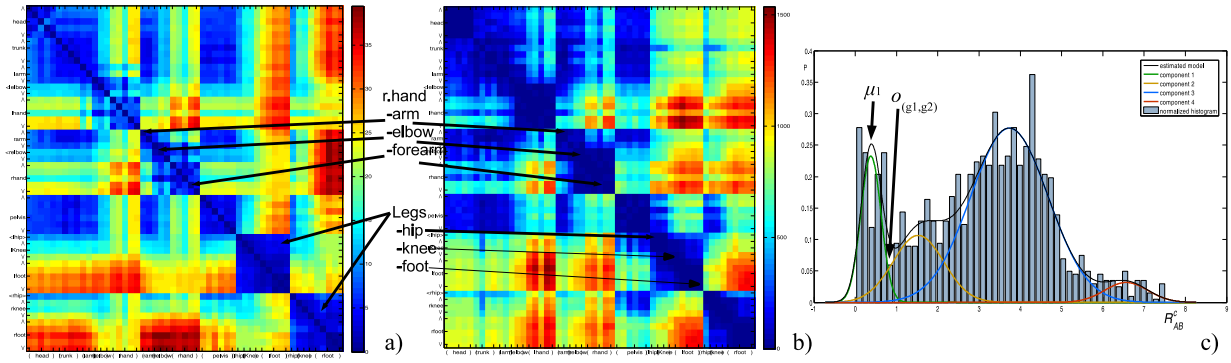

FIGURE 2. Inter-marker distances for IM - exemplary subject: a) matrix of R^c, b) matrix of R^E, c) GMM for R^c annotated for threshold evaluation (see further in the text)

TDCB is a special order of processing which is crucial for further proper clustering process that would preserve the body hierarchy - i.e. forearm is subordinate to arm or thigh to shank not otherwise. It is achieved by bicriterial top-down/center-boundary (TDCB) sorting according to the main body axis (A) in the frame with the smallest sum of gradients, which allowed us to identify the T-pose or just relaxed standing. Main body axis A was identified with peak marker P_0 and first principal component (PC_1) obtained with PCA (principal component analysis) performed on the markers coordinates. We want markers to be sorted from the highest one (top of the head) to the bottom (feet) and from the main axis of the body (which roughly conforms spine in vertebrates) to the most distant parts (hands). The first criterion (top-down) is the i-th marker position (P_i) along the main axis (A) - we measure it as the distance(D^{TD}) between the given marker projection onto A and the peak marker. The second criterion was the Euclidean distance (D^{CB}) of the P_i from A. They are simply calculated as:

$$D_i^{TD} = (PC_1 \cdot \vec{P_0P_i})/|PC_1|, \quad D_i^{CB} = |PC_1 \times \vec{P_0P_i}|/|PC_1| \tag{4}$$

For the first criterion we employed bucket sorting with 10 equally spaced buckets, next within each bucket all markers were sorted with *qsort* according to the second one. It would result in meaningful hierarchies even if the subject is lying, although, there might be exceptions - for example coiled snake would not exhibit main body axis.

Clustering of siblings requires checking for the two conditions: coherence of movements and constancy of distances. The clustering algorithm is a simple incremental, iterative process of scanning markers in TDCB order and attaching to the cluster currently processed marker if it fulfills both criteria with respect to the initial marker (no linkage updating) of the cluster. The thresholds are based on GMM models. When all the possible to attach markers are assigned to cluster we take next non-clustered (free) marker as initial one for new cluster and again join free markers to a new cluster.

Threshold scaling appeared in experiments to be necessary. As we heuristically identified pure GMM classification to be sufficient for only well recorded ROM sequences. We decided to leave the threshold level for both measures as a tuning parameter for the end user with the value corresponding to the pure GMM classification set as default. The threshold value T is scaled along the GMM results according to the tuning parameter s. For negative s the threshold is scaled as a fraction of range between 0 and μ_1; for positive s between μ_1 and the value where the first and second mode intersects $o_{(g_1,g_2)}$ - see the annotations in Fig. 2c. The threshold value is computed as:

$$T = \{if\ s < 0,\ \mu_1 - |s| \cdot \mu_1;\ \ if\ s = 0,\ \mu_1;\ \ if\ s > 0,\ \mu_1 + s \cdot (o_{(g_1,g_2)} - \mu_1)\} \tag{5}$$

where: s is the tuning parameter with default value 1 for pure GMM.

Body hierarchy is obtained with identification of a parent for each of the group of siblings finally forming a tree. For the identification of a parent we propose a consensus approach. According to the assumptions a single group of siblings can have one common parent. In the first step each marker in the group, using constancy of distance, identifies a parent candidate marker, next, the parent group is chosen as the one containing the most of the candidates. The candidate marker is the one having the smallest R^E to the proposing one (presumably located at the joint) with the limitation that it is neither the member of the current group nor any other located in the subordinate tree below the current group. All the voting process is performed in bottom-up order which ensures the head group to be the root of a tree.

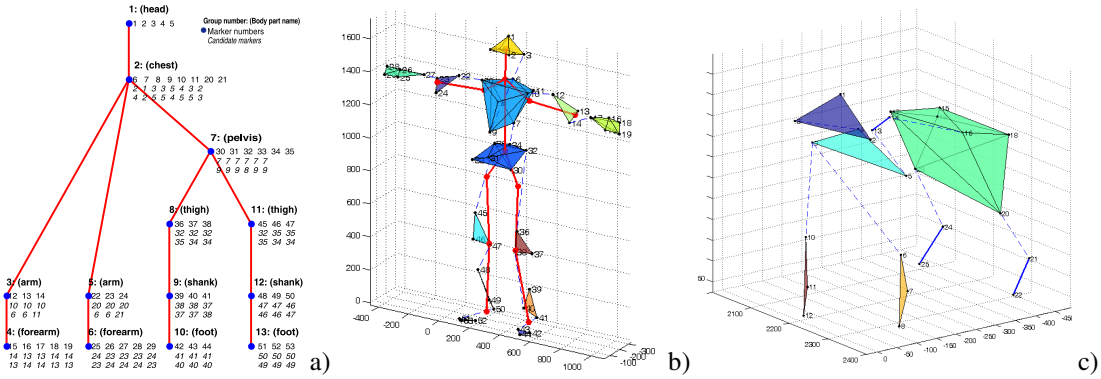

FIGURE 3. Hierarchy: a) tree with voting for IM, b) resulting FBM for IM subject c) FBM for a dog

SUMMARY OF RESULTS

In the empirical verification, we used regular mocap recorded sequences, none of which were recorded intentionally for our research. Due to lack of 'ground truth' for the resulting model we had to examine the results visually, whether they conform subjects body. The most of data were recorded in PJIIT HM lab using a Vicon MX system with Vicon Blade software for general purpose and Vicon Nexus for medical sequences. They are commonly used calibrating sequences - a ROM (*range of movement*) sequence [4] which is described as a human subject moves all limbs and does exercising all rotation extremes for every joint, which were collected for both medical and animation needs and are of different quality from perfect ones to the ROMs of heavily disabled persons. Additionally, we used a sequence of dog mocap available in a public repository and as inadequate to the method - a face spelling alphabet (non-rigid) which was recorded in our lab. Furthermore, we verified the results against ROMs from the CMU repository.

The results (see Figs. 3b,c) were obtained by an experienced operator who manually tuned s parameter with respect to the data. The s values were typically in between 0.9 and 1.5, although, it was necessary to set it up to 2.5 value for the dog or the poorly performed sequences. The result in the Figs. 3b, is demonstrating the subjects pose in the most stable frame (T-pose) so body hierarchy is visible.

Both the inference method and the FBM reprefor as the representation for signal processing. To verify its usefulness we applied FBM to control denosing filter and skeleton estimation (see Fig. 3b). Additionally, we also used it for marker prediction and artifact detection. The other methods of structure inference, utilizing additionally the proximity of markers, are also under consideration - we are currently working on artificial neural networks for fully automatic inference process. The usage of FBM for non-rigid subjects - human faces - is at this moment at the conceptual stage.

ACKNOWLEDGMENTS

This project has been supported by the National Centre for Research and Development, Poland. (Project INNOTECH In-Tech ID 182645 "Nowe technologie wysokorozdzielczej akwizycji i animacji mimiki twarzy.")

REFERENCES

1. M. Kitagawa, and B. Windsor, *MoCap for Artists: Workflow and Techniques for Motion Capture*, Focal Press, 2008.
2. T. B. Moeslund, and E. Granum, *Comput. Vis. Image Underst.* **81**, 231–268 (2001), ISSN 1077-3142.
3. T. B. Moeslund, A. Hilton, and V. Krüger, *Comput. Vis. Image Underst.* **104**, 90–126 (2006), ISSN 1077-3142.
4. *Vicon Blade, How to Setup Characters. Blade, A Step-by-Step Ref. Guide Rev. 1.0*, Vicon Motion Systems (2008).
5. J. H. Challis, *Journal of Applied Biomechanics* **15**, 303–317 (1999).
6. A. Aristidou, and J. Lasenby, *The Visual Computer* **29**, 7–26 (2013), ISSN 0178-2789, 1432-2315, 00011.
7. M.-C. Silaghi, R. Plünkers, R. Boulic, P. Fua, and D. Thalmann, "Local and Global Skeleton Fitting Techniques for Optical Motion Capture," in *CAPTECH '98 Proc. of the Int. Workshop on Modelling and Motion Capture Techniques for Virtual Environments*, LNAI, Springer-Verlag, London, UK, 1998, pp. 26–40, ISBN 3-540-65353-8.
8. G. Liu, and L. McMillan, *The Visual Computer* **22**, 721–728 (2006), ISSN 0178-2789, 1432-2315, 00042.
9. M. A. T. Figueiredo, and A. K. Jain, *IEEE Trans. Pattern Anal. Mach. Intell.* **24**, 381–396 (2002), ISSN 0162-8828.