Applied Mathematics & Information Sciences
An International Journal

http://dx.doi.org/10.18576/amis/100107

Functional Body Mesh Representation,, A Simplified Kinematic Model, Its Inference and Applications

Przemysław Skurowski[1,*] *and Magdalena Pawlyta*[2]

[1] Institute of Informatics, Silesian University of Technology, ul. Akademicka 16, Gliwice, Poland
[2] Branch Faculty of Information Technology of Polish-Japanese Academy of Information Technology, Al. Legionów 2, Bytom, Poland

Received: 10 Jun. 2015, Revised: 8 Aug. 2015, Accepted: 9 Aug. 2015
Published online: 1 Jan. 2016

Abstract: The paper describes a simplified representation of a body structure and a GMM based method for inferring from the motion capture data based on a functional relationship between the points. The proposed representation can be efficiently used for marker-wise processing of the data. The parent-child and sibling relationships are inferred on a coherence of movement and constancy of distances. 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

1 Introduction

Optical motion capture (MoCap) [13] registered data consists of recorded spatial coordinates (trajectories) of tracking points (usually markers) over time. In the initialization stage markers are organized into some object model [15] for further processing needs. It is common for motion processing algorithms [16] to use arbitrary human model - manually pre-edited skeletal structure (with limb lengths and joint locations) tightly [19] connected to the predefined marker locations (mesh).

Prior to the obtaining smooth skeletal animation it is usually necessary to process the markers recorded positions. In this stage, incorporating such filters as denoising and reconstruction, there are preocessed markers either independently (e.g. [5]) or there is required reference to the estimated skeleton [1]. Skeleton free processing, utilizing rigid-body mesh is also possible, but in a pipeline of a typical commercial software the arbitrary (human) model is simply assigned to the markers or tuned to the markers [4], [19]. To create of a new subject model in the software it is required to do manual work, although there is some assistance. Such templates are stored in external files.

The proposal of a body representation - a *functional body mesh* (FBM) stems from the fact that marker-wise processing of movement data still remains an area open for improvements, alas most of current body meshes which can be found in software are defined arbitrarily so even simple change of marker locations or adding an item to an actor can result in skeleton template modification (not to mention of modelling different species), meanwhile the proposed solution is intended to be a model-free representation of the subject's kinematic structure.

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. Such a representation should form a framework for further marker-wise signal processing. The secondary goal was that the representation should be easy to obtain with as little a user interaction as possible in the automated or semi-automated processes.

The proposed body structure representation can be considered as an intermediate between skeletal and marker form. It comprises a tree of partial submeshes reflecting the body structure, where the submeshes represent specific body parts and the tree represents their interconnections and hierarchy. Such an approach would allow to process the MoCap data using all the information stored in raw xyz marker trajectories (without aggregating

* Corresponding author e-mail: przemyslaw.skurowski@polsl.pl

it into skeletal animation) whereas it would also allow to benefit from the knowledge of the body hierarchy.

The FBM representation can be employed in two basic ways: straightforward in current software as a simple visualization method (see p. 4.2) 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.

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 is based on the movement and distance relationships of marker points which stem from the corollaries of a rigid body assumption. We perform an analysis of inter-marker relationships (parent-child and sibling) and provide a machine learning based method utilizing clustering and statistical modelling for automated generation of a subjects body structure.

Our algorithm proposal is able to adapt to the most vertebrate subjects, such as animals which movements are recorded. The proposed approach can be also adapted to marker-less MoCap systems based on tracking numerous features (such as SIFT or SURF [3,9]) which could be affected by the fact that these features are randomly located and noise prone. Therefore, such a method like our proposal seems appropriate due to its ability to adopt to random feature configuration and due to its statistical nature to overcome the noise and gaps.

In p. 5, as a proof-of-concept, we provide an exemplary application of the functional mesh in the marker smoothing filter; next we demonstrate its correspondence to the skeletal structure and we discuss possible extension.

2 Related work

The most of current body movement analysis methods focus on parameterizing of the skeletal structure inferred on the body motion (e.g. [17,22,18,12,6],), some of which, employing a *local* scheme, share the stage of marker segmentation on the basis of rigid body principle with our approach [18,6,12].

A similar approach to our proposal, although less formalized, a concept of hierarchized groups of correlated markers to represent subjects topology was employed to detect and fill missing markers (gaps)in MoCap recordings [14]. The other approach was presented in [23] where authors propose automatic marker labeling (assignment to known segments) using segmentation with a rigid body assumption, they also demonstrated usability of their model to recover gaps in MoCap recordings.

Other similar topics which allow for the variability of subjects models are - adaptation to the variability of marker locations to drive a known skeleton was proposed in [11], another interesting idea of semantic rigid segment matching to represent body structure was proposed in [20]. A hybrid model of a skeleton-mesh intended to the animation of semi-rigid models was proposed in [7].

3 The inference method

Our solution results in forming tree structure of marker points on the rigid body assumption. It is based on two criteria - coherence of movement and constancy of distance. The method is based on following rationales:

–Siblings 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 that:
1. is located in another group,
2. cannot be located in a child node (no loops),
3. is 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.

A planar example of two body parts, parent (AB) and child (CD) anchored to B, is shown in Fig. 1. There are depicted body rotations (two) which plus translation of base are sufficient to describe motion of the human body. As one can see gradient angles vary very little within a single body part meanwhile they differ much more between markers inside and outside body part. Such a difference is even stronger for unconnected body parts which can move independently or in opposite directions like legs during walking. Similar observation can be made for the distance between connected markers for siblings and partially for parent-child - distances should only differ due to noise, meanwhile they can vary for body parts not directly connected - see the distance between base and remote markers in the example.

The method comprises three basic steps:

1. sorting markers in special order *top-down/center-boundary* (TDCB) to achieve meaningful body hierarchy in next step

2. clustering of siblings into groups representing body part

3. hierarchy recovery by selection of a parent for each of marker groups.

In further paragraphs of the article we will explain how we described above points using mathematical terms. For the siblings grouping task, since classical clustering algorithms [2] require quite complicated modifications to use more than one criterion, we propose using of incremental clustering which can simply check all the criteria in each iteration.

The key problem in the proposed algorithm was to determine the inter marker relationship prior to the clustering - we need to decide whether they fulfill the criteria of coherent movement and constant distance or not. To solve the problem we employed the Gaussian Mixture Models (GMM) [10] which appeared to fit the experimental data very well and have good rationale based motivation. GMM classified each pair of markers into one of the predefined classes describing the degree of motion coherence or constancy of distances.

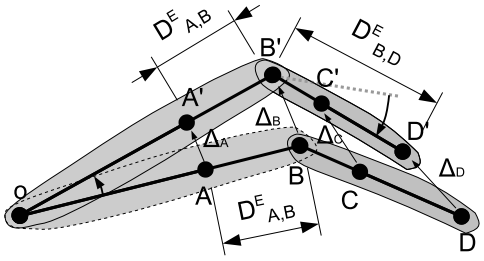


Fig. 1: Example of moving two groups (AB, CD) - with gradients, distances and angles

The sketch in Fig. 1 illustrates also a problem which results from the fact that group membership of markers located on joints is ambiguous. Please note the marker B which is located on joint might be sometimes considered as a part of group of a BCD markers meanwhile the desired group partitioning is an AB-CD. Joints are considered as parts of body segments (bones) located higher in the hierarchy - e.g. the elbow is considered as a part of the arm rather than the forearm. Thus prior to the clustering stage, it was necessary to use special top-down/center-boundary (TDCB) marker processing order in clustering algorithm to achieve this feature. Moreover, proper sorting of markers would provide also a meaningful structure of connections, with the head as a root, and with limbs connected to the torso (not otherwise).

3.1 Distance functions

For the measurement of a distance we used 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:

$$D^E_{A,B}(n) = \sqrt{(x_A - x_B)^2 + (y_A - y_B)^2 + (z_A - z_B)^2} \tag{1a}$$

$$R^E_{AB} = P_U\left\{D^E_{A,B}(1..N)\right\} - P_L\left\{D^E_{A,B}(1..N)\right\} \tag{1b}$$

where: x_A, y_A, z_A, x_B, y_B, z_B - coordinates of respective points A, B for n^{th} frame.

The movement of the k^{th} marker is well described with gradient:

$$\begin{aligned}\Delta_k(n) &= [\Delta_x, \Delta_y, \Delta_z] = \\ &= [x_n - x_{n-1}, y_n - y_{n-1}, z_n - z_{n-1}]\end{aligned} \tag{2}$$

where: $n, n-1$ - number of successive frames; x_n,x_{n-1},... - marker coordinates in two successive frames; Δ_x, Δ_y, Δ_z - differences of respective coordinates for two successive frames. Angular coherence of gradients was measured using weighted cosine distance:

$$\begin{aligned}D^c_{\Delta_A,\Delta_B}(n) &= w(1 - \cos(\Delta_A \Delta_B) = \\ &= w(1 - \frac{(\Delta_{Ax}\Delta_{Bx} + \Delta_{Ay}\Delta_{By} + \Delta_{Ax}\Delta_{Bz})}{\sqrt{\Delta^2_{Ax} + \Delta^2_{Ay} + \Delta^2_{Az}}\sqrt{\Delta^2_{Bx} + \Delta^2_{By} + \Delta^2_{Bz}}}).\end{aligned} \tag{3}$$

As the value of weighting function w we used averaged length of gradients:

$$w = w(\Delta_A, \Delta_B, n) = \frac{1}{2}\left(L_A(n) + L_B(n)\right) \tag{4}$$

where L_k is distance of a movement of k marker is:

$$\begin{aligned}L_k(n) &= |\Delta_k| = \\ &= \sqrt{(x_n - x_{n-1})^2 + (y_n - y_{n-1})^2 + (z_n - z_{n-1})^2}\end{aligned} \tag{5}$$

where: $n, n-1$ number of two successive frames.
Cosine distance provides useful information about vectors conformance - it results in zero value when they are the same, 1 when orthogonal and 2 when opponent. In case of an ideal rigid body, pure cosine distance would be satisfactory for our needs, but during tests, it appeared necessary to incorporate weighting by length of gradient. The cosine distance does not consider the length of the movement, meanwhile markers placed on the same body part could have small non-consistent (opponent) movements due to deformation (see Fig. 2) caused by elasticity of the human body. To overcome this problem and suppress the influence of small deformations we decided to use weighting as it is described above.

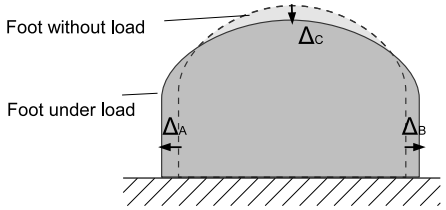


Fig. 2: Deformation scheme of the foot under load of a body. Frontal view with three gradients of virtual markers A,B,C

Finally, the coherence of gradients of a markers in the whole sequence was calculated for each pair of markers as a range of values. Again, to filter out noise existing in

records we used the interquantile distance between lower (L=1st) and upper (U=99th) percentile of the whole sequence:

$$R^c_{AB} = P_U\left\{D^c_{\Delta_A,\Delta_B}(1..N)\right\} - P_L\left\{D^c_{\Delta_A,\Delta_B}(1..N)\right\}. \quad (6)$$

Please note that the proposed values for percentiles in both range measures appeared to be crucial for proper results of the rigid body - they were tuned using the experimental data. The proposed distance functions for each pair of markers can be collected in the distance matrix. When markers are arranged in the order reflecting human anatomy (TDCB) then one can observe (Fig. 3) desired existence of small distances ('dark squares' along the diagonal) for groups of markers placed on common body parts. Generally, both measures conform each other but they also provide some supplementary information. In Fig. 3a internal structure of a hands is easily visible while Fig. 3b exhibits parts of legs better.

3.2 Determining inter-marker relationship

Inter-marker relationship is determined by assigning of their pairs into classes according to the rationales mentioned in p. 3. One can intuitively assign each pair into one of C - four different classes: peer (c1), close (c2), independent (c3) and opponent (c4). The distance functions described in previous subsection should reflect these classes so one should consider a multimodal probability distribution function (PDF) for both distance functions. The Gaussian mixture models [10] appeared to fit the data very well as it is demonstrated in Fig. 4a. The GMM is a mixture model summing two or more Gaussian probability distributions (g_i):

$$G = \sum_{i=1}^{C} g_i = \sum_{i=1}^{C} w_i\, N(\mu_i, \sigma_i), \quad (7)$$

where: C number of components (in our case $c_1..c_4$), w_i - weight, $N(\mu_i, \sigma_i)$ normal probability distribution of a μ_i, mean value and σ_i std deviation. These parameters are estimated in the process of *expectation maximization* [8]. Classification using GMM is a selection of a class with the highest mode value for the given object (x):

$$c(x) = \underset{i=1..C}{\arg\max}\ \{w_i N(\mu_i, \sigma_i)(x)\}. \quad (8)$$

The Fig. 4b demonstrates classification of marker pairs according to the model visible in Fig. 4a. for the R^c_{AB} data shown in Fig. 3a. The same satisfactory behaviour was observed for both the GMM for the R^E_{AB} measure and for classification - not included in a figure. Alas, during the experimentation, it appeared that pure GMM classifiers were too loose as they would connect also a remote markers. It appeared to be necessary to recompute the threshold values on a basis of the GMM results - the

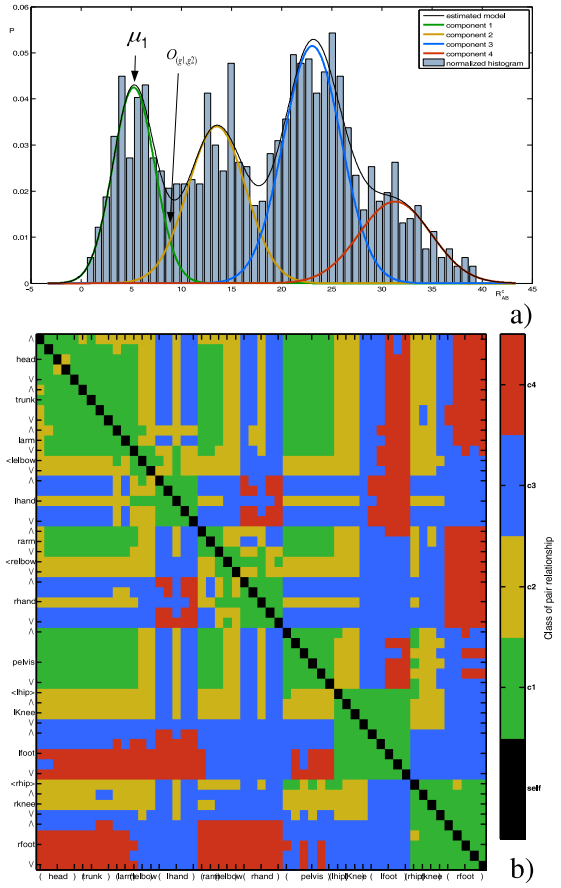


Fig. 4: Determining inter marker relationships for subject IM: a) GMM model for R^c_{AB}, b) resulting classification

proposed rescaling scheme is piecewise linear given with Eq. (9).

In further experiments we heuristically identified simple GMM classification to be sufficient for only well recorded *range-of-motion* (ROM) sequences (see p. 4.1). 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. 4a. The threshold value is computed as:

$$T = \left\{ \begin{array}{l} \text{if } s < 0,\ \mu_1 - s\cdot\mu_1 \\ \text{if } s = 0,\ \mu_1 \\ \text{if } s > 0,\ \mu_1 + s\cdot(o_{(g_1,g_2)} - \mu_1) \end{array} \right\}, \quad (9)$$

where: s is the tuning parameter with default value 1 for pure GMM. The typical values used during experimentation appeared to be 1-1.5.

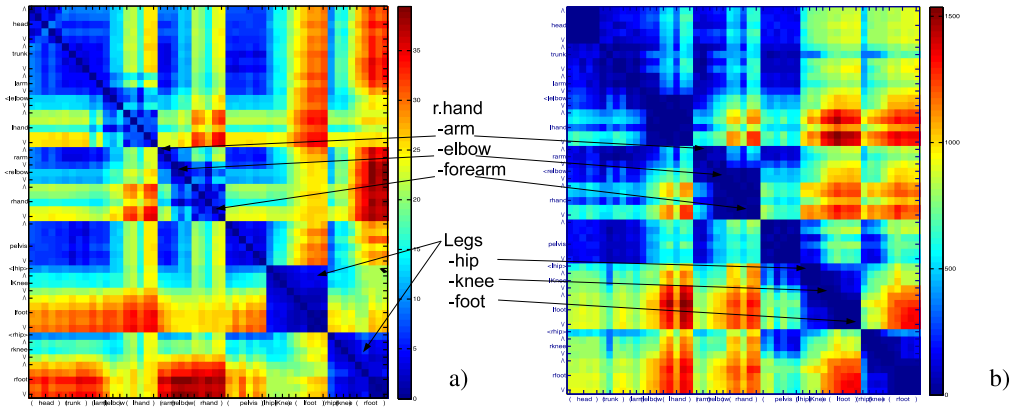


Fig. 3: Exemplary distance matrices for subject IM: a) Range of weighted cosine distance R^c_{AB}, b) Range of Euclidean distance percentiles R^E_{AB}

3.3 Top-down/center-boundary sorting

The special order of processing is crucial for proper clustering process that would preserve the body hierarchy. It is achieved by bicriterial top-down/center-boundary (TDCB) sorting. We sorted the data using the most steady frame in sequence (smallest sum of gradients) which allowed us to identify the T-pose or just relaxed standing in the dog. Markers are processed from the highest one (top of the head) to the bottom (feet) and from the main axis of the body (spine) to the most distant parts (hands). Main body axis is identified with PCA (principal component analysis) which is performed on markers coordinates in the most steady frame. The PCA does its task by diagonalizing the covariance matrix (C) into a matrix (Λ) of eigenvalues (λ_i) and produces the new orthogonal base vectors (PC - eigenvectors) which points out the direction of data variability stored as columns in change of basis matrix (E).

$$C = E\Lambda E^T . \tag{10}$$

Eigenvalues (and corresponding eigenvectors - PCs) are sorted in descending order, thus the first principal component ($PC_1 = [px_1, py_1, pz_1]$) identifies the direction of the largest elongation of the body. The first PC with a peak marker (P_0) located on the top of a head identify the main body axis (A) which roughly conforms the spine in vertebrates.

Bicriterial sorting is performed with respect to the main axis. The first criterion (top-down) is the i-th marker position (P_i) along the main axis (A) - we measure it as the distance between the given marker projection onto A and the peak marker P_0 , calculated as:

$$D_i^{TD} = \frac{PC_1 \cdot \mathbf{P_0P_i}}{|PC_1|} . \tag{11}$$

The second one was the Euclidean distance of a marker P_i from A:

$$D_i^{CB} = \frac{|PC_1 \times \mathbf{P_0P_i}|}{|PC_1|} . \tag{12}$$

We applied a combined approach for sorting. For the first criterion we employed bucket sorting with 10 equally spaced buckets. The number was chosen empirically using various subjects. Within each bucket all the markers were sorted with *qsort* according to the second criterion. The TDCB order can be noted as:

$$P_i \prec P_j \Leftrightarrow \left\{ \begin{array}{l} \text{if } D_i^{TD} \lesssim D_j^{TD}, \text{ or} \\ \text{if } D_i^{TD} \approx D_j^{TD} \wedge D_i^{CB} \leqslant D_j^{CB} \end{array} \right\} . \tag{13}$$

This way of sorting results in meaningful hierarchies even if the subject would be lying on the floor in the most steady frame, although, there might be exceptions. For example coiled snake would not exhibit main body axis, also for some humans like yoga performers who are able to stand still in extraordinary positions the identification of steady frame can be improper thus body main axis can be far from spine location.

3.4 Clustering of siblings

The clustering of sibling markers requires checking for the two conditions: coherence of movements and constancy of distances. These requirements are well described with the membership to the proper classes of a respective GMM. 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. When all the possible to attach markers are assigned to cluster we take next nonclustered (free) marker as initial one for a new cluster and again join free markers to new cluster. The process can be described using the following pseudocode:

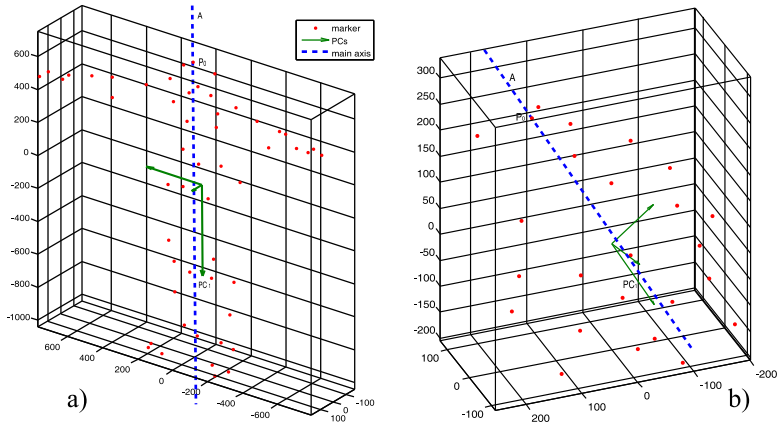


Fig. 5: Location of main body axis in: a) human, b) sitting dog

```
const sG, sR; // tuning parameters
int n;  //number of markers
double R(n,n); //dist matrix range measure
double Dc(n,n);//dist matrix weight cosine
type marker :={double x,y,z};
marker M(n);  //nonassigned marker list
C(.,.) := []; //empty 2D cluster-marker list
TDc:=evalthreshold(Dc, sG);
TR:=evalthreshold(DR, sR); // see eq.(9)
M:=sortTDCB(M);
for all i in M(1..n)
  ind_i++;
  ind_j:=1;
  C(ind_i,ind_j):=i; //init new cluster
  M:=M-M(i); // remove ith marker from list
  for all j in M(i+1..n)
    if R(i,j)<=TR & Dc(i,j)<=TDc
      ind_j++;
      C(ind_i,ind_j):=j; //add to cluster
      M:=M-M(j); // remove j-th marker
    end if
  end for
end for
```

The results for the above algorithm are demonstrated in Figs. 7, 8 where each group is wrapped or with convex hull either marked with bold line.

3.5 Identification of a parent

According to the rationales a single group of siblings can have one common parent. For the identification of a parent we propose a consensus approach. In the first step each marker in the group identifies, using constancy of distance, a candidate marker being a member of the proposed parent; next, the parent is chosen as the group having the most of the candidates. The candidate marker is the one having the smallest R^E_{AB} distance 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 subtree below the current group.

Base element results from the organization of markers. The overall tree building process runs upwards from the bottom elements - last in the list. Therefore the root of a tree will be identified as a group consisting of markers from the beginning of the list. It is a common organization of the data where the initial markers are located on the head of a subject and the last ones are located on the legs. A simple improvement for the basic voting step is to allow each marker to propose more than one candidate for a parent to solve when voting for a single parent is nonconclusive (e.g. equal result).

```
int l;  //number of clusters
int m;  //length of cluster
int n;  //number of markers
int f;  //number of candidates per marker
int C(l,m);  // matrix of clusters
double R(n,n);  // R distance matrix
int CParent(l)=0; //
for i = l..1
  M=length(i);//num. or mrkrs in ith clust
  mrkrs = C(i,1..M); //get current markers
  tmpR = R(1..n, mrkrs); //dists for i clust
  children = getchildren(Cparent, C);
    //recursively get markers of children
  tmpR({mrkrs, childern},1..M) = infinity;
    //exclude self and children markers
  if isallinf(tmpR)
    break; // the root found
  end
  inds = sort(tmpR); // get sorted dists
  candidates = inds(1..f,1..M);
  //get proposals for each of M markers
  votes = getclustnumber(candidatetes);
  winner = max( count(votes));
  CParent(i) = winner;
end for
```

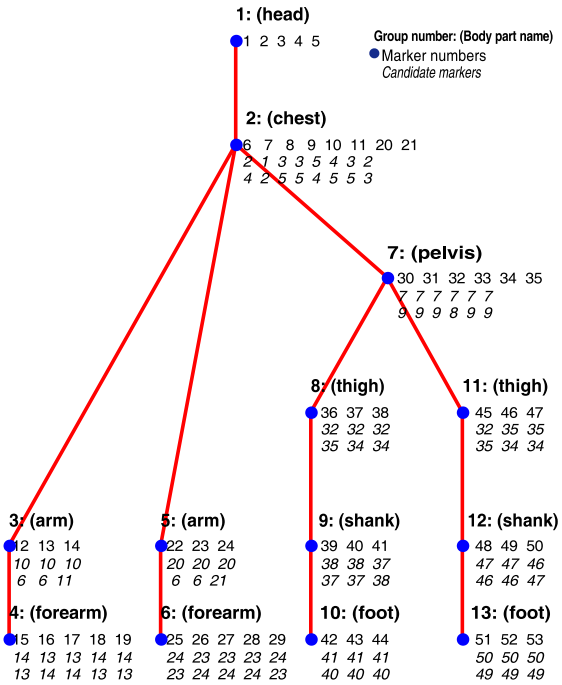


Fig. 6: Proper parent-child relationship for subject IM annotated with marker numbers (for parameter $s = 1$) and two candidate markers for each member of a cluster

4 Testing

4.1 The data

The most of test data were recorded in PJIIT HML lab using Vicon MX a MoCap system equipped with 10 T40 Vicon NIR cameras and Vicon Blade software for the general purpose recordings and Vicon Nexus for medical sequences. The data were collected for both medical and animation needs, none of them were prepared especially for the needs of this algorithm. They are ordinary, commonly used calibrating sequences - a ROM (*range of movement*) sequence [19] which is described as a human subject moves all limbs and does exercising all rotation extremes for every joint. Additionally, we used a sequence of a face spelling alphabet which was recorded in our lab and a sequence of a dog MoCap available in a public repository [24].

4.2 Results

Due to lack of a 'ground truth' for the resulting model the best we could do was to examine the results visually. We tested a set of various MoCap sequences, among others these include:

1. proper calibration ROM sequence according to Vicon manual (Fig. 7a,b)
2. other species (dog) non-calibrating, but representative sequence (Fig. 7c)
3. improper calibration sequence (Fig. 8b)
4. ROM of a person with movement disabilities (hip endoprosthesis and after a stroke incident) - captured for medical purposes (Fig. 8a)
5. non rigid body (face) - an inadequate case for the algorithm (Fig. 8c)

The results demonstrated in Figs. 7 and 8 were obtained by an experienced operator who tuned s parameter with respect to the data. The results in the figure are demonstrating the subjects pose in the most stable frame (T-pose usually) so the body hierarchy is visible. The visual representation of a movement is depicted in Fig. 9, where a few frames form a walk sequence starting with T-pose just after the ROM recording are shown.

Table 1: Thresholds set up for the exemplary sequences

Recording	s_R	s_G
HJ	0.9	1.1
IM	0.5	1
dog	2.5	2.5
JB	1	1
AR	0.9	0.9
face	2	2

4.3 Disussion of results

When a subject performed carefully whole the ROM sequence, then it appeared that the default value (pure GMM classifier) worked properly (Fig. 7b). In case when a subject did not perform the whole ROM sequence well then it was necessary to increase the thresholds. The dog sequence, although it was not a calibration sequence, did quite well as it contained a representative set of moves for most of the joints.

In cases when the sequence was incomplete as shown in Fig. 8b it was impossible to specify proper thresholds. The other, interesting, observation (Fig. 8a) we made for medical recording of an aged person with limited movement abilities who performed ROM sequence as it was possible for him - in such a situation a model appeared to exhibit the stiffness of joints observed in the person. The face experiment showed the method to be irrelevant to the non-rigid body - we were unable to use the method effectively.

For the testing purposes, we also verified operating of the algorithm against ROMs from other data sets. One of them was the well known CMU repository [25] - using 41 markers (instead of 53 set used in our lab). We noted proper functioning of proposed method but mainly for the later ROM records - early ones were not representative enough.

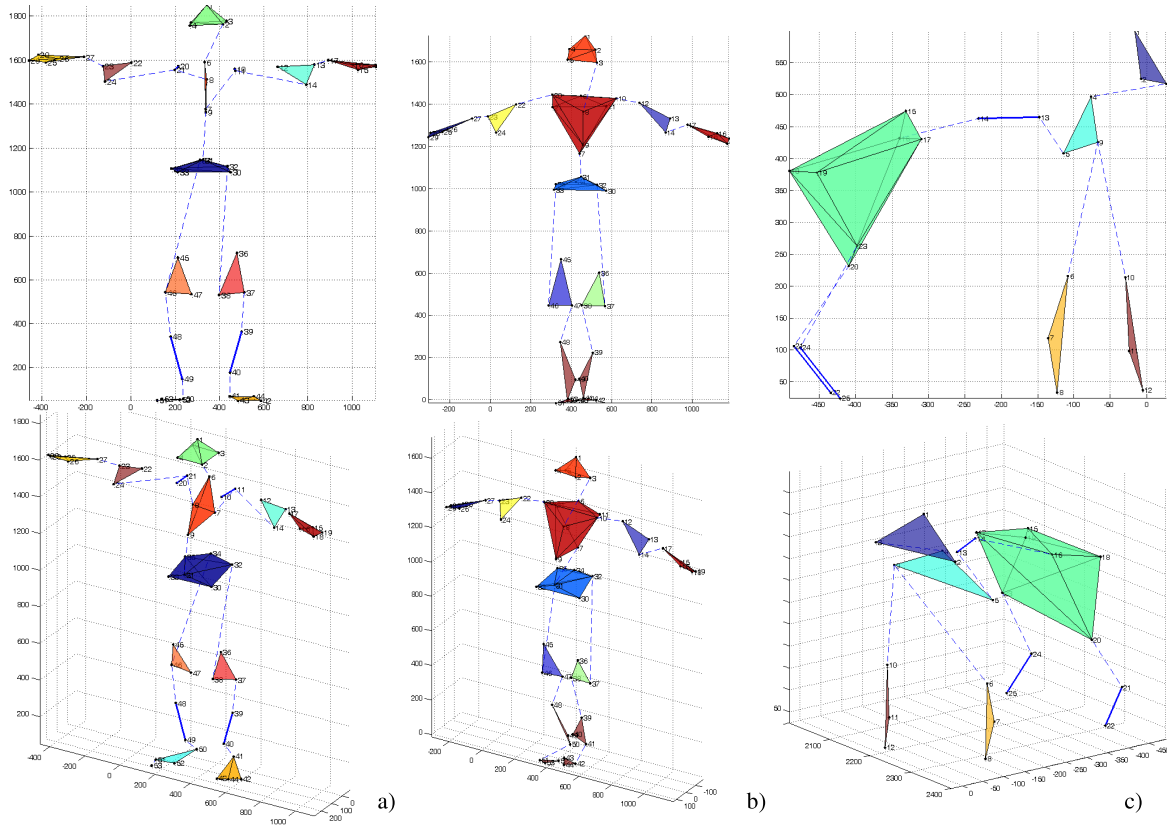


Fig. 7: Exemplary satisfactory results for: a) human subject HJ (male), b) human subject IM (female), c) an animal (dog)

5 Applications and extensions

The obvious application of the proposed representation is to organize and visualize the structure of a body of a new subject type. Visual interpretation of submeshes can be various depending on the data provided. For sparse points as in p. 4.2 the convex hull provides visually efficient representation. Alternatively, for relatively dense point clouds (like in [9]), the solid representing body part could approximate the real body surface. Therefore, a surface mesh spanned (triangulated) on the points cloud could be an effective representation.

Aside of the simple visualization of a kinematic structure, the resulting meshes can be used for other tasks, these are: structure aware signal filtering, recontruction, artifact detection and prediction. We have already prototyped all these, but since they are out of scope of this paper, as a proof-of-concept we demonstrate: that FBM can be easily used to estimate skeleton and that using it as a framework for structure aware marker smoothing. Finally, we discuss optional extension of the basic scheme (with multiple parents per segment).

5.1 Skeleton inferring

Skeleton inferring is an obvious application of the mesh resulting from our proposal (see Fig. 10). We employed an approach based on the classical local method [18]. In this case one can consider inferring of a functional mesh as a rigid body partitioning stage of classical algorithm with additional information about body part connections. The joint locations were estimated for the parent-child connected groups of markers. We calculated them as rotation centers for trajectories of markers in child group using local (relative) coordinate system which was established according to the position and rotation of the parent body part.

5.2 Marker-wise filtering

Another proposed exemplary application is to control processing of the raw marker data at the cleaning stage. One of such techniques is a low pass Butterworth IIR filter [21] which is used to eliminate a motion jitter. We employed the results of movement analysis within the body mesh to control cutoff frequency. Standard deviation of euclidean distance between all markers within group was used - one can consider it as a marker's deviation

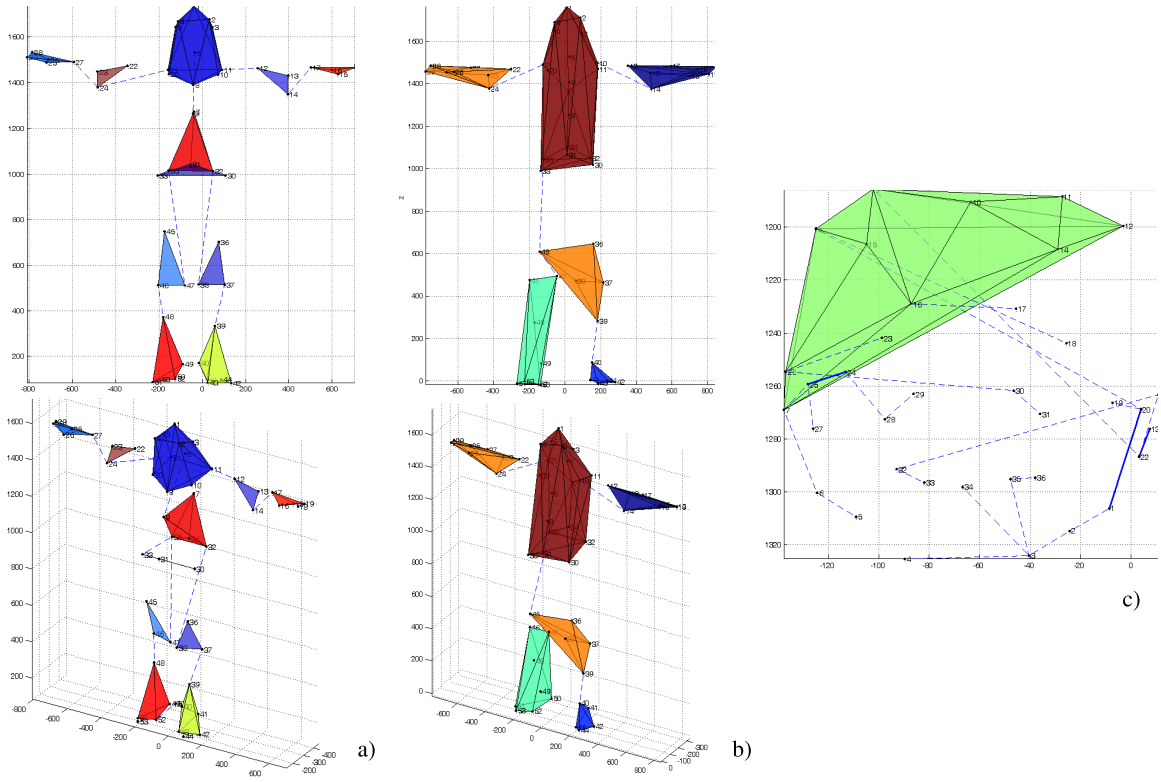


Fig. 8: Exemplary non-satisfactory results for: a) aged and impaired subject JB (male), b) incomplete ROM sequence for subject AR (male), c) non-rigid body (face - markers are placed on the muscles)

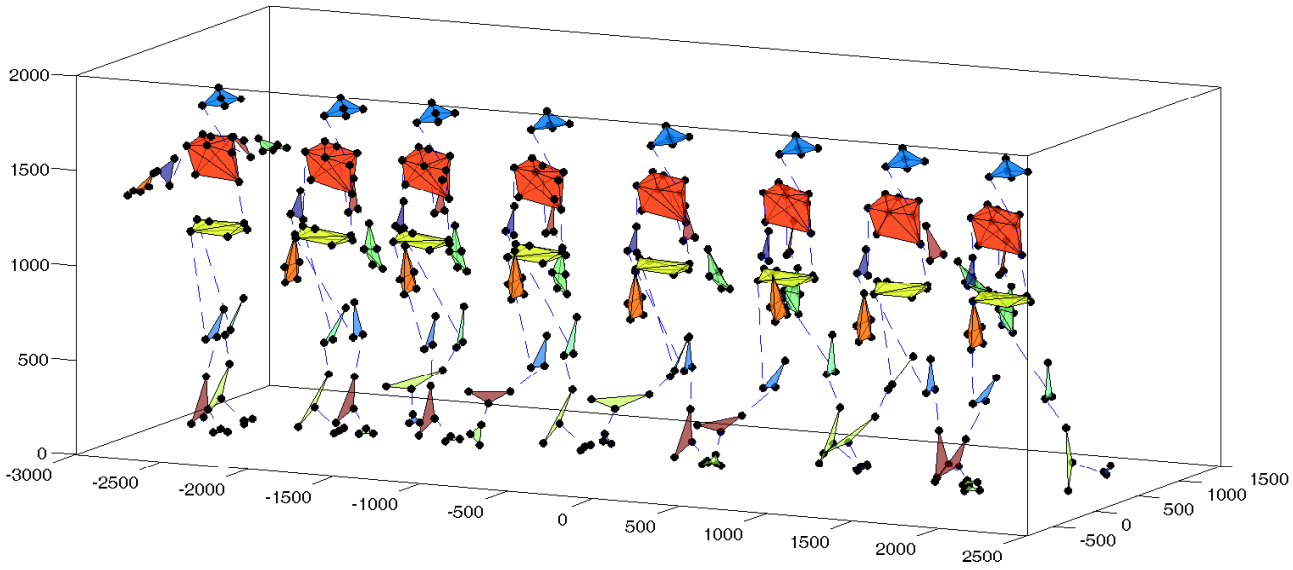


Fig. 9: A walk sequence for HJ subject

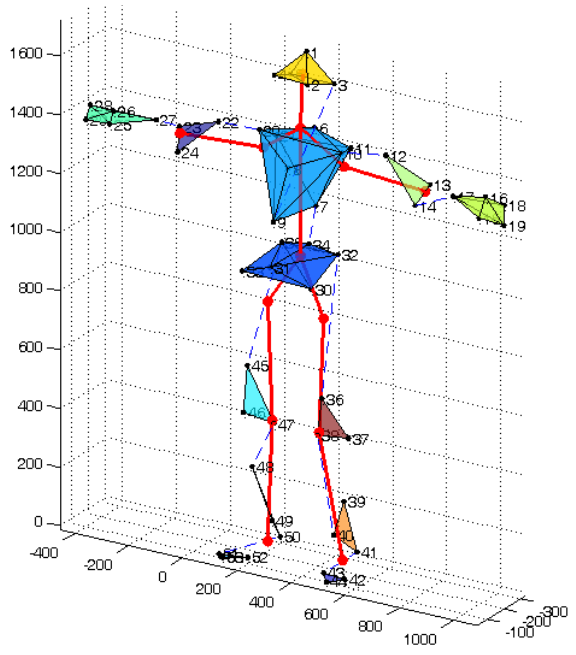


Fig. 10: A skeleton on the corresponding source functional mesh for IM subject

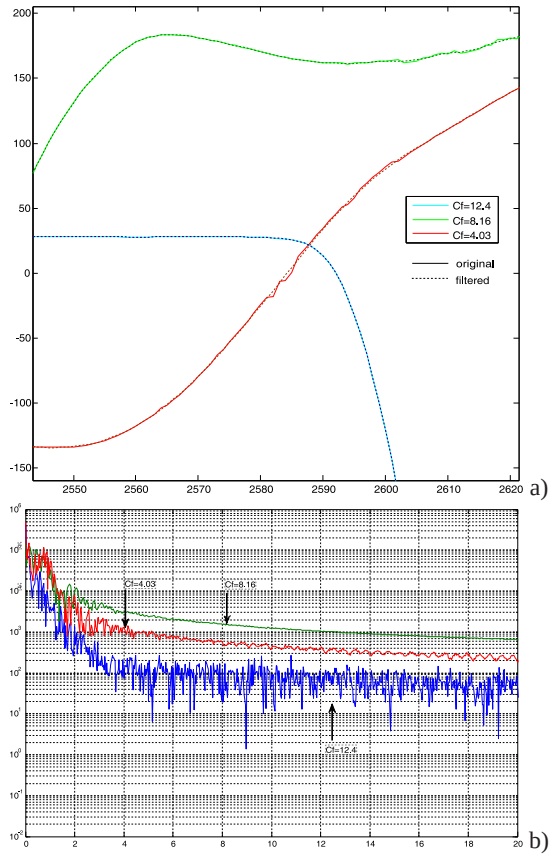


Fig. 11: Rigid body controlled Butterworth low pass filtering of trajectories: a) original and filtered for 3 exemplary markers (min, avg and max deviation from the rigid body model), b) their Fourier amplitude with cut-off frequencies annotated

from rigid body model. For each marker in a group we calculated deviation to all other sibling markers - the larger resulting value the proportionally lower cutoff frequency (stronger filtration) is assumed for the second order Butterworth filter. criterion:

$$V_i = \frac{1}{|M|-1} \sum_{m \in M, m \neq i} \sigma_{i,m}, \tag{14}$$

where: $|M|$ is a number of markers in a group M, $\sigma_{i,m}$ is a variance of the distance between i and m markers, the additional condition of $m \neq i$ is due to $\sigma_{i,i} \triangleq 0$. Next, the averaged variance is remapped using $g$function which adopts V_i to the cut-off frequency Cf. In our case is simple linear scaling between arbitrary minimal and maximal limits.

$$Cf_i = g(V_i) \tag{15}$$

Demonstrative results are demonstrated in the Fig. 11a, where one can see examples of the filtration for minimal, average and maximal value of the V_i within a body part. The corresponding amplitude spectra, depicted in Fig. 11b, show that cut-offs (Cf) are *not* simply related to the amplitude of high frequencies.

5.3 Mesh extensions

The rigid body model assumed in previous paragraph suggests that an extended version of a basic functional mesh can be useful - each separate rigid segment (cluster) could also point more parent markers from the parent segment for which the rigid body requirements (thresholds) are fulfilled - usually these markers would be located by the joints. Resulting mesh hierarchy would be slightly redundant - hierarchical graph, instead of tree - but including these points into convex hulls representing rigid body model parts could make visual results even more realistic. This step can be performed very easily by taking the results of the voting stage (p. 3.5) and including the candidate markers from the parent cluster into the current one if they fulfill the thresholds from the clustering stage (p. 3.4).

6 Summary and future work

The inferring method provides results as it was expected - it is a realistic mesh of connections between markers. It allows for semi-automatic inferring of a body structure without prior knowledge about the subject model and requires minimal operator interaction.

The other methods of structure inference, which are based on other clusterization techniques and using

additionally proximity of markers, are also under consideration - we are currently working on artificial neural networks to ensure fully automatic inference process - the results are very promising requiring no user interaction and small number of errors. The inference method for non-rigid subjects - human faces specifically - is at this moment at conceptual stage.

Use of FBM structure made implementing of an exemplary signal processing aware of kinematic structure very easy - we can easily process multivariate signals simply referring between values of the explicitly given (as parents and siblings) variables.

Both the inference method and the proposed representation can be easily adopted into the already existing motion editing software. We successfully applied the proposed representation for the artifact detection, signal reconstruction and denoising of the MoCap sequences such as in the example mentioned in p. 5.2. These results will be brought to the wider audience in prospective papers.

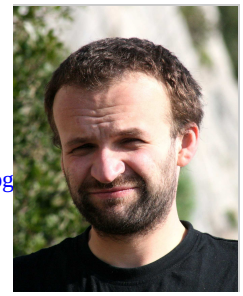
Acknowledgement

The authors are thankful to the anonymous referee for a careful and helpful review. 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.”).
This work was also partly performed using the infrastructure supported by POIG.02.03.01-24-099/13 grant: GCONiI - Upper-Silesian Center for Scientific Computation.

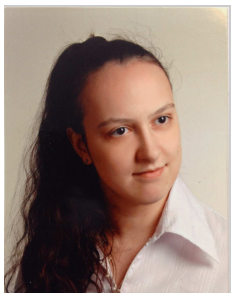
References

[1] Andreas Aristidou and Joan Lasenby, *Real-time marker prediction and CoR estimation in optical motion capture*, The Visual Computer **29** (2013), no. 1, 7–26 (en), 00011.
[2] Max Bramer, *Principles of data mining*, Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2007.
[3] T. Brox, B. Rosenhahn, J. Gall, and D. Cremers, *Combined region- and motion-based 3d tracking of rigid and articulated objects*, IEEE Trans. on Pattern Analysis and Machine Intelligence (2009).
[4] C-Motion, *V3Dwiki tutorial: Visualizing data associating movement trials and model*.
[5] John H. Challis, *A procedure for the automatic determination of filter cutoff frequency for the processing of biomechanical data*, Journal of Applied Biomechanics **15** (1999), no. 3, 303–317.
[6] Edilson de Aguiar, Christian Theobalt, and Hans-Peter Seidel, *Automatic learning of articulated skeletons from 3D marker trajectories*, Advances in Visual Computing (George Bebis, Richard Boyle, Bahram Parvin, Darko Koracin, Paolo Remagnino, Ara Nefian, Gopi Meenakshisundaram, Valerio Pascucci, Jiri Zara, Jose Molineros, Holger Theisel, and Tom Malzbender, eds.), Lecture Notes in Computer Science, no. 4291, Springer Berlin Heidelberg, January 2006, pp. 485–494.
[7] Edilson De Aguiar, Christian Theobalt, Sebastian Thrun, and Hans-Peter Seidel, *Automatic conversion of mesh animations into skeleton-based animations*, Computer Graphics Forum **27** (2008), no. 2, 389–397.
[8] A. P. Dempster, N. M. Laird, and D. B. Rubin, *Maximum likelihood from incomplete data via the em algorithm*, J. Roy. Statist. Soc. Ser. B **39** (1977), no. 1, 1–38.
[9] Ashish Doshi, Jonathan Starck, and Adrian Hilton, *An empirical study of non-rigid surface feature matching of human from 3D video*, Journal of Virtual Reality and Broadcasting **7** (2010), no. 3.
[10] Mario A. T. Figueiredo and Anil K. Jain, *Unsupervised learning of finite mixture models*, IEEE Trans. Pattern Anal. Mach. Intell. **24** (2002), no. 3, 381–396.
[11] R. Haratian, R. Twycross-Lewis, T. Timotijevic, and C. Phillips, *Toward flexibility in sensor placement for motion capture systems: A signal processing approach*, IEEE Sensors Journal **14** (2014), no. 3, 701–709.
[12] Adam G. Kirk, James F. O’Brien, and David A. Forsyth, *Skeletal parameter estimation from optical motion capture data*, Proc.of the 2005 IEEE Comp. Soc. Conf. on Computer Vision and Pattern Recognition (CVPR’05) (Washington, DC, USA), vol. 2, 2005, pp. 782–788.
[13] Midori Kitagawa and Brian Windsor, *Mocap for artists: Workflow and techniques for motion capture*, Focal Press, 2008.
[14] Guodong Liu and Leonard McMillan, *Estimation of missing markers in human motion capture*, The Visual Computer **22** (2006), no. 9-11, 721–728 (en), 00042.
[15] Thomas B. Moeslund and Erik Granum, *A survey of computer vision-based human motion capture*, Comput. Vis. Image Underst. **81** (2001), no. 3, 231–268.
[16] Thomas B. Moeslund, Adrian Hilton, and Volker Krüger, *A survey of advances in vision-based human motion capture and analysis*, Comput. Vis. Image Underst. **104** (2006), no. 2, 90–126.
[17] David A. Ross, Daniel Tarlow, and Richard S. Zemel, *Learning articulated structure and motion*, International Journal of Computer Vision **88** (2010), no. 2, 214–237 (en).
[18] Marius-Calin Silaghi, Ralf Plünkers, Ronan Boulic, Pascal Fua, and Daniel Thalmann, *Local and global skeleton fitting techniques for optical motion capture*, CAPTECH ’98 Proc. of the Int. Workshop on Modelling and Motion Capture Techniques for Virtual Environments (London, UK), LNAI, Springer-Verlag, 1998, pp. 26–40.
[19] Vicon Motion Systems, *Vicon blade, how to setup characters. blade, a step-by-step ref. guide rev. 1.0*, June 2008.
[20] Xiaopeng Wei, Boxiang Xiao, Qiang Zhang, and Rui Liu, *A rigid structure matching-based noise data processing approach for human motion capture*, Proc. of the 2011 Workshop on Digital Media and Digital Content Management, DMDCM ’11, IEEE Comp. Soc., 2011, pp. 91–96.
[21] D.A. Winter, *Biomechanics and motor control of human movement*, John Wiley & Sons, 2004.
[22] Zhidong Xiao, Hammadi Nait-Charif, and Jian Zhang, *Automatic estimation of skeletal motion from optical motion*

capture data, Motion in Games, LNCS, vol. 5277, Springer, 2008, pp. 144–153.
[23] Qian Yu, Qing Li, and Zhigang Deng, *Online motion capture marker labeling for multiple interacting articulated targets*, Computer Graphics Forum **26** (2007), no. 3, 477–483 (en).
[24] *Free motion capture data*, http://motioncapturedata.com/2009/05/animal-motion-capture-dog
[25] *CMU graphics lab motion capture database*, http://mocap.cs.cmu.edu/resources.php.



Przemysław Skurowski received both M.Sc. Eng. degree and PhD degree in computer science from Silesian University of Technology in 2000 and 2007 respectively. He is currently an associate professor in the Institute of Informatics in Silesian University of Technology. His work is focused on the multimedia, image processing and human centered audiovisual information processing.



Magdalena Pawlyta received an engineer degree from Polish-Japanese Institute of Information Technology in 2013. She is a professional motion capture editor in Human Motion Laboratory at PJIIT with practical experience in recording of human motion sequences for medical, animation and game development purposes. Currently she is working on her M.Sc. thesis.