ARTICULATED MOTION ANALYSIS FROM MOTION CAPTURE DATA

J. Fayad[1], A. Del Bue[2] and P. M. Q. Aguiar[3]

1. ABSTRACT

We present computational methods to extract and model different joints of a generic subject, in an automatic way. The input for our methods is simply a set of trajectories of 3D points, obtained from a motion capture system (MOCAP). Due to the piecewise rigidity of the skeleton, these trajectories belong to different subspaces (of the space of all possible trajectories). We use this knowledge to derive computationally simple algorithms that are able to infer joint properties from the trajectories of the 3D points. This model-free approach enables analyzing directly specific subjects, rather than requiring user-defined a priori models of the skeleton. Such data-driven models are useful in the analysis of human walk and the evaluation of joint stress for pathology detection. Also, the customized skeleton models, together with physiological muscular information, allow to accurately analyze high level performances in sports.

2. INTRODUCTION

Computational models of human articulations are nowadays fundamental to perform an accurate analysis of the mechanical motion of a human body. Applications of these models span various fields, ranging from engineering to life sciences, where the analysis of human motion is crucial to accurate clinical analysis and credible animations of skeleton models[1]. Existing approaches to articulated motion analysis require a human operator to explicitly construct the model of the skeleton, including the definition and characterization of the joints, a costly and time-consuming task [2]. This can be done by measuring the dimensions of different body segments and articulation ranges of motion of a human subject, which can be troublesome. Alternatively one can use anthropometric tables to extract statistical measurements of those quantities and then use scale factors to resize the model to better fit a specific subject. However this last approach has the disadvantage of scaling the model in a statistical way and not modeling a specific individual, which can be required in some applications.

In opposition, we present computational methods to extract and model different joints from a generic subject, in an individualized and semi-automatic way. The only assumption made here is the availability of an initial segmentation of each near-rigid part composing the body. Such approach has relevant relations with recently developed techniques in the Computer Vision field [3,4,5,6] where, given the unpredictability of the data extracted from images, unconstrained analysis is often the only viable solution.

[1]Master Student, Institute for Systems and Robotics (ISR/IST), Av. Rovisco Pais 1, Lisbon, Portugal
[2] Researcher, Institute for Systems and Robotics (ISR/IST), Av. Rovisco Pais 1, Lisbon, Portugal
[3] Professor, Institute for Systems and Robotics (ISR/IST), Av. Rovisco Pais 1, Lisbon, Portugal

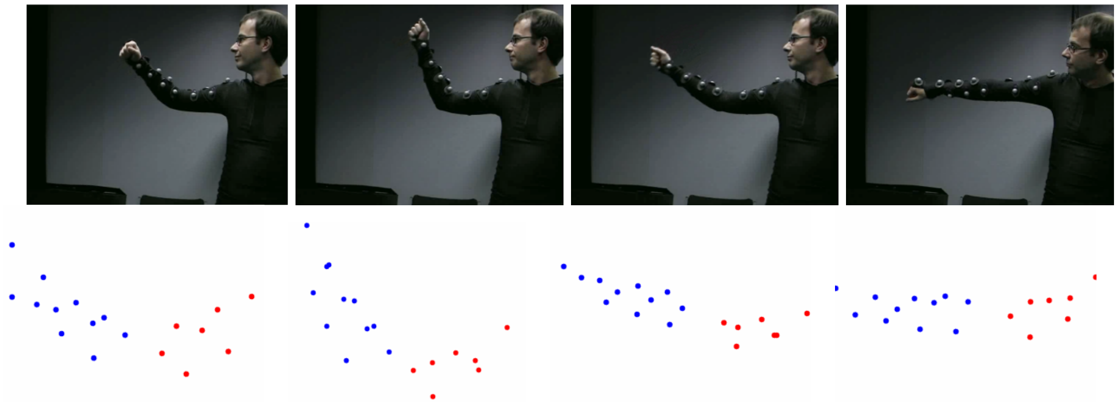

Fig 1. Four images and relative 3D information extracted from a MOCAP system each color corresponds to the set of points lying over the two bodies used to extract the arm joint.

3. A BILINEAR FORMULATION OF STRUCTURE AND MOTION

Our aim is to infer the musco-skeletal properties of a human body directly from the data acquired from MOCAP system. The main idea is to fulfill the estimation of these anatomical properties independently from the platform used to acquire the data. Such approach has the main feature of being independent to *a priori* models of the body. The approach is based on the key fact that a set of points, lying over a 3D object which is moving in time, shares common properties. For instance, if the body is moving rigidly, a simple rotation and translation can describe the motion of the set of 3D points. Similarly, points lying over an articulated structure can be described by the 3D position associated to both bodies, their relative translation and rotations and the common axis of rotation given the joint.

Interestingly, these properties become evident in a framework where the 3D object shape and the respective motion components are modelled as a bilinear form such that:

$$w_{ij} = M_i S_j \qquad (1)$$

where w_{ij} is a 3-vector representing the 3D shape point j captured at the time instant i. The 3xr matrix M_i represents the motion components of the 3D object (i.e. rotation and translation for a rigid shape) and the rxP matrix S_j collects a parameterisation of the 3D shape. If we collect every 3D point in a compact matrix form we obtain that:

$$W = \begin{bmatrix} W_1 \\ \vdots \\ W_F \end{bmatrix} = \begin{bmatrix} M_1 \\ \vdots \\ M_F \end{bmatrix} \begin{bmatrix} S_1 & \cdots & S_p \end{bmatrix} = MS \qquad (2)$$

where W is a 3FxP matrix, M a 3Fxr matrix and S a rxP matrix. The scalars F and P are, respectively, the number of frames captured and the number of points belonging to the shape. Each $3 \times P$ matrix W_i with $i=1\ldots F$ is such that:

$$W_i = \begin{bmatrix} w_{i1} & \cdots & w_{iP} \end{bmatrix}$$

The value of r, i.e. the dimensionality of the bilinear models, depends by the type of shape considered. In the following we show a set of example for these bilinear models for the human motion modelling scenario.

3.1 Rigid shapes

A body moving rigidly brings the dimensionality of the bilinear models to either $r = 3$

or $r = 4$, depending if we consider the rotating body with or without a translational component. In the case of no translation, the matrices M_i and S_j take the following form:

$$M_i = \begin{bmatrix} r_{i1} & r_{i2} & r_{i3} \\ r_{i4} & r_{i5} & r_{i6} \\ r_{i7} & r_{i8} & r_{i9} \end{bmatrix} \qquad S_j = \begin{bmatrix} x_j \\ y_j \\ z_j \end{bmatrix}$$

where M_i is a $3x3$ rotation matrix and x_j, y_j and z_j represent the three coordinates of the 3D point. Thus each point at each frame is given by the product as expressed in eq. (1). Given the grouping in eq. (2), we have that M and S are a $2Fx3$ and a $3xP$ full-rank matrices respectively. This forces a rank constraint on the measurements W (i.e. $rank(W) \leq 3$). Given this rank constraint, we can compute an initial factorization of W by performing a SVD giving:

$$W \xrightarrow{SVD} \sum_{i=1}^{P} u_i \sigma_i v_i^T = \sum_{i=1}^{r} u_i \sigma_i v_i^T = U_r \Sigma_r V_r^T$$

where U_r is a $2F \times r$ orthogonal matrix, Σ_r a $r \times r$ diagonal matrix and V_r a $P \times r$ orthogonal matrix. In the rigid case, we have that $\sigma_4 = 0$ if no noise and measurements errors are present. Thus, by enforcing the singular values after the third equal to zero, we obtain a numerical fit of the measurements in the sense of the Frobenius norm. This stage is likely to improve the measurements quality since even MOCAP systems have a certain uncertainty over the 3D location of points (Some systems provide also an indication of this uncertainty, thus a more accurate denoising could be performed). This initial decomposition via SVD can provide a first affine fit of the motion and shape components M and S such that:

$$\tilde{M} = U_r \text{ and } \tilde{S} = \Sigma_r V_r^T .$$

Since this transformation is valid up to an affine transformation i.e. $W = \tilde{M} Q Q^{-1} \tilde{S}$ we seek a specific transformation Q which enforces the metric properties of M (see Section 4 for a detailed description). The further advantage of this procedure is that we obtain a metric description of the shape S using the complete 3D information coming from the MOCAP system (i.e. all the frames and not only a singular frame). We avoid in such way to choose a 3D shape which relies only on a single instance of the captured motion.

Given a translation, the models can be similarly defined as:

$$M_i = \begin{bmatrix} r_{i1} & r_{i2} & r_{i3} & t_x \\ r_{i4} & r_{i5} & r_{i6} & t_y \\ r_{i7} & r_{i8} & r_{i9} & t_z \end{bmatrix} \qquad S_j = \begin{bmatrix} x_j \\ y_j \\ z_j \\ 1 \end{bmatrix}$$

where t_x, t_y and t_z are the translational component and S_j is now expressed in homogeneous coordinates.

3.2 Articulated Shapes – Universal Joint

In order to simplify the formulation, we consider articulated shapes composed by pairwise rigid bodies. In this case our measurements are given by the 3D points lying over both parts i.e. $W = [W^{(1)} \mid W^{(2)}]$. Two types of joints are here considered: the *universal joint* and the *hinge joint*. When two objects are linked by a universal joint the distance between the two centers of mass is constant (for instance, the head and the torso of a human body) but they have independent rotation components. At each frame the shapes connected by a joint satisfy the following relation:

$$t^{(1)} + R^{(1)} d^{(1)} = t^{(2)} + R^{(2)} d^{(2)} \qquad (3)$$

where $t^{(1)}$ and $t^{(2)}$ are the 3D shape centroids of the two objects, $R^{(1)}$ and $R^{(2)}$ the 3×3 rotation matrices and $d^{(1)}$ and $d^{(2)}$ the 3D displacement vectors of each shape from the central joint. The constraint expressed in eq. (3) reduces the dimensionality of the model. It is then possible to refer the articulated motion to a common reference frame registered on the centroid of the first object thus simplifying the shape matrix S as:

$$S = \begin{bmatrix} S^{(1)} & d^{(1)} \\ 0 & S^{(2)} - d^{(2)} \\ 1 & 1 \end{bmatrix} \qquad (4)$$

where S is a full rank-7 matrix. The motion for a frame i has to be arranged accordingly to satisfy equation (3) giving:

$$M_i = \begin{bmatrix} R_i^{(1)} & R_i^{(2)} & t_i \end{bmatrix}.$$

To obtain the block diagonal structure as in eq. (4) from a SVD of W, we consider the two shapes registered to the measured trajectories centroid:

$$\tilde{W} = \begin{bmatrix} \tilde{W}^{(1)} \mid \tilde{W}^{2} \end{bmatrix} = \begin{bmatrix} R^{(1)} & R^{(2)} \end{bmatrix} \begin{bmatrix} S^{(1)} & 0 \\ 0 & S^{(2)} \end{bmatrix}$$

which is a full-rank 6 matrix. However, the SVD of W computes the components of $V = [V^{(1)}\ V^{(2)}]$ that are dense. The remedy is to separate the two objects by multiplying the row space by a matrix N such that:

$$NV = \begin{bmatrix} null(V^{(2)}) \\ null(V^{(1)}) \end{bmatrix} \begin{bmatrix} V^{(1)} & V^{(2)} \end{bmatrix} = \begin{bmatrix} null(V^{(2)})V^{(1)} & 0 \\ 0 & null(V^{(1)})V^{(2)} \end{bmatrix}$$

where *null()* returns the left null-space of a matrix. The column space U is then multiplied by N^{-1} in order to retain consistency. Given the recovered block structure, separate metric constraints can be computed to recover the correct structure and motion.

3.3 Articulated Shapes – Hinje Joint

A hinge joint constrains the relative orientation of the 3D objects since both shapes have a common axis that is parallel to the axis of rotation. In this case, the overall rank of W drops to 6. To simplify visualization in a matrix form, we have aligned the axis of rotation to the x-axis. If $R^{(1)} = [p_1\ p_2\ p_3]$ and $R^{(2)} = [p_1\ p_4\ p_5]$ we can write that:

$$\tilde{W} = \begin{bmatrix} p_1 & p_2 & p_3 & p_4 & p_5 \end{bmatrix} \begin{bmatrix} x_1^{(1)} & \cdots & x_{P_1}^{(1)} & x_1^{(2)} & \cdots & x_{P_2}^{(2)} \\ y_1^{(1)} & \cdots & y_{P_1}^{(1)} & & & \\ z_1^{(1)} & \cdots & z_{P_1}^{(1)} & & & \\ & & & y_1^{(1)} & \cdots & y_{P_2}^{(1)} \\ & & & z_1^{(1)} & \cdots & z_{P_2}^{(1)} \end{bmatrix}$$

Similarly to the solution proposed for the universal joint, we have to compute a first fit using SVD and then to accordingly enforce the matrix structure of the row space V. In this case, we have to reduce some components of the row space of V to zero. However here we have a strong dependency on the x-axis which does not allow to consider the shapes separately. Thus we use a matrix N constructed such that:

$$N = \begin{bmatrix} null(V^{(2)}) \\ null(V^{(1)}) \end{bmatrix} \begin{bmatrix} b^T \\ null(V^{(2)}) \\ null(V^{(1)}) \end{bmatrix}$$

where $b^T = [1\ 0\ 0\ 0\ 0]$. The column space can be then transformed accordingly with N^{-1}. Now a hinge joint is defined as a line which define the axis of rotation and thus its center may be localized anywhere on the axis. However, the center lies in the null space of $[p_1\ p_2\ p_3\ p_4\ p_5\ (t_2 - t_1)]$, thus it can be extracted from the previous factorization.

4. SHAPE CALIBRATION

In the previous section we have found affine fits for two different types of joints. An upgrade to a metric space is necessary in order to compute a 3D shape that resembles the metric properties of the bodies. However, this stage is not directly affecting the previous computation of the joint position. The upgrade to metric is accomplished by imposing that the matrix M is a collection of rotation matrices at each frame. These constraints are obtained with the computation of a transformation matrix Q, which can correct the matrices at each frame to orthogonal. The motion matrix M can be given as:

$$\tilde{M} = \begin{bmatrix} \tilde{M}_1 \\ \vdots \\ \tilde{M}_F \end{bmatrix}$$

where each 3×3 block M_i can be written as:

$$\tilde{M}_1 = \begin{bmatrix} m_{1i}^T \\ m_{2i}^T \\ m_{3i}^T \end{bmatrix}. \qquad (5)$$

Thus, in order to obtain the solution for Q, we explicitly enforce orthonormality constraints in (5). These can be expressed as a set of equations in the form of $m^T_{di}Q^T Qm_{ei} = c_{de}$ where $d, e = 1, 2, 3$. The value of c_{de} depends explicitly on the given joint. A solution is obtained by considering the product $H = Q^T Q$, solving linearly for H using Least-Squares and then extracting Q with Choleski decomposition. The dimensionality of Q depends by the problem considered: $Q_{3\times3}$ for a single rigid body, $Q_{6\times6}$ for a universal joint, $Q_{5\times5}$ for a hinge joint.

5. EXPERIMENTS

In order to show the flexibility of our approach, we present results using two different databases of 3D points extracted using a VICON system. The first set shows point lying over an arm performing unknown motion. The task here is to obtain the location of the rotation axis associated to the elbow. The 3D points in Fig. 2 are related to images as shown in Fig. 1. The only information we require in order to run the approach is the association of each point to the related segment of the 3D body (i.e. a segmentation of the point trajectories).

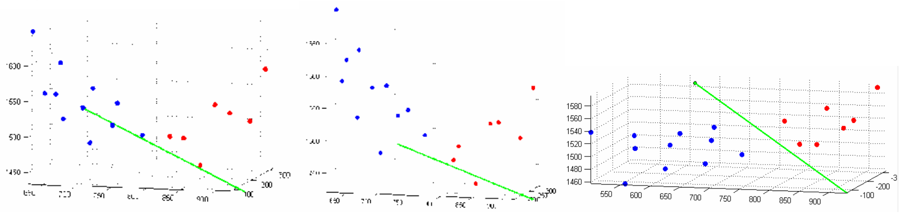

Fig 2. Three sample frames showing an arm and the computed hinge joint at the elbow

The second set of points was obtained from the freely available database of the Graphic Lab of Carnegie Mellon University (http://mocap.cs.cmu.edu/). In such experiment, an adult is running in an unconstrained environment. We were able to extract universal joints at each elbow and two hinge joints for each leg, located at the knee and ankle, as it can be seen in Fig. 3

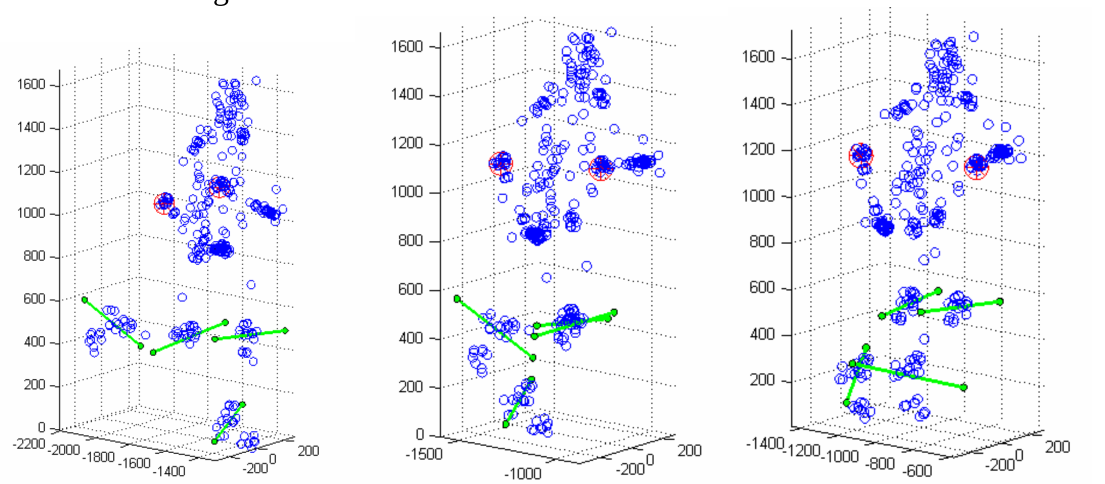

Fig 3. Three sample frames showing an adult human running. Two universal joints (red) and four hinge joints (green) are extracted from the full body trajectory.

6. CONCLUSIONS

We have presented a framework for computing the joint positions from a set of observations obtained from MOCAP systems. Our analysis is not constrained by the specific setup used for the acquisition but it accounts for high degrees of flexibility. As future work, we aim to extend these methods to deal with missing entries in the measurements and to add consistent modelling of the muscular properties of the body.

7. ACKNOWLEDGMENTS

This work was partially supported by Fundação para a Ciência e a Tecnologia (ISR/IST plurianual funding) through the POS_Conhecimento Program that includes FEDER funds and grant PTDC/EEA-ACR/72201/2006, “MODI - 3D Models from 2D Images”.

8. REFERENCES

[1] Mündermann, L. and Corazza, S. and Andriacchi, T.P., The evolution of methods for the capture of human movement leading to markerless motion capture for biomechanical applications. J. of NeuroEngineering and Rehabilitation. Vol. 3(1)

[2] Kroemer, K. and Snook, S. and Meadows, S. and Seutsh, S., Ergonomic models of anthropometry, human biomechanics, and operator-equipment interface, 1988. National Academy Press.

[3] Tomasi, C. and Kanade, T., Shape and motion from image streams under orthography: a factorization method. International Journal of Computer Vision, 1992.

[4] Bregler, C. and Hertzmann, A. and Biermann, H., Recovering non-rigid 3D shape from image streams. IEEE Conference on Computer Vision and Pattern Recognition, 2000.

[5] Tresadern, P. and Reid, I., Articulated Structure From Motion by Factorization. IEEE Conference on Computer Vision and Pattern Recognition, 2005.

[6] Yan, J. and Pollefeys, M., A factorization-based approach to articulated motion recovery. IEEE Conference on Computer Vision and Pattern Recognition, 2005.