Tracking Stick Figures with Hierarchical Articulated ICP

D. Moschini, A. Fusiello
Dipartimento di Informatica, Università di Verona
Strada Le Grazie 15, 37134 Verona, Italy
andrea.fusiello@sci.univr.it

Abstract

This paper presents an ICP-based algorithm for tracking an articulated skeletal model of the human body (stick figure) in 3D. The data are 3D points distributed roughly around the limbs' medial axes. The algorithm fits each stick to a limb in a hierarchical fashion, traversing the body's kinematic chain, while preserving the connection of the sticks at the joints. Experimental results illustrate the algorithm.

1 Introduction

Tracking or capturing the motion of a human body is a problem that has a long history in Computer Vision (see [11] for a survey) and several real-world application, such as human-computer interfaces, motion transfer, animation of virtual characters, activity/gesture/gait recognition, biomechanical studies. Marker-based commercial systems are available that works at very high frame rates and very high precision. While it is out of doubt that such speed/accuracy combination is necessary in biomechanics, it is questionable whether it is needed when animating a virtual character in a videogame or building a user-interface. There is therefore a niche for less expensive systems that work markerless at a reduced speed (up to 100 Hz). In this paper we present part of an ongoing project aimed at building a system with those characteristics.

The literature on markerless body tracking in three dimensions can be broadly split into two groups: those using a stick model for the human body [5, 9], roughly corresponding to its skeleton, and those using a full 3D model of the body's shape, in the form of a polygonal mesh or a volumetric model [2, 8, 12]. Since we aim at a real-time system, we are forced to work with a stick model. Indeed, a stick (or skeletal) model has fewer dependencies on anthropometric parameters than a shape model and can be tracked much faster because of its simplicity.

We use an approach based on the well-known Iterative Closest Point (ICP) algorithm [4]: the model is an articulated stick figure representing the body and it's kinematics, the data are 3D points distributed roughly around the medial axes of the limbs. The data are registered to the model using a hierarchical approach that proceeds by traversing the kinematic chain.

Differently from [12] we do not enforce joints constraints *a-posteriori* (thereby interfering with the result of ICP) but *during* the registration process. To the best of our knowledge this is the only ICP-based approach with this feature. Other approaches based

on the EM algorithm enforce the joint constraints, but they are much more computation intensive than ICP.

2 Background

2.1 Human Body Model

In this section we describe the articulated model representing the human body pose we used in the paper. It consists of a kinematic chain of ten sticks and nine joints, as depicted in Figure 1. The torso is at the root of tree, children represents limbs, each limb being described by a fixed-length stick and the corresponding rotation from its parent. Hence, the motion of one body segment can be described as the motion of the previous segment in the kinematic chain and an angular motion around a body joint. Only the torso contains a translation that accounts for the translation of the whole body. Rotations are represented with 3×3 matrices. For the sake of simplicity, each joint has three degrees of freedom and there are no limits on the angles.

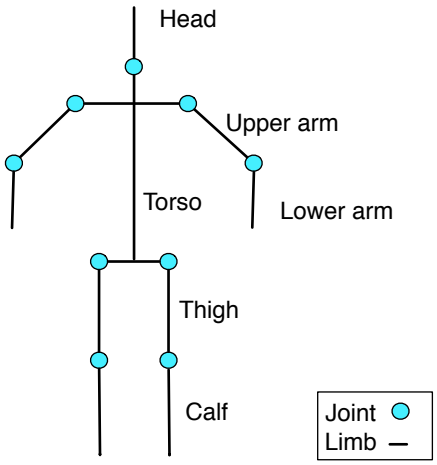

Figure 1: The stick figure body model.

2.2 Iterative Closest Point

The Iterative Closest Point (ICP) [6, 4] is customary used for registration of 3D sets of points. The standard algorithm estimates the rigid motion between a given set of $3D$ data points and a set of $3D$ model points. In summary:

Algorithm 1 ICP

Input: two sets of p 3D points, the data $\{\mathbf{a}_i\}_{i=1\ldots p}$ and the model $\{\mathbf{b}_i\}_{i=1\ldots p}$
Output: the rigid motion T that brings the data onto the model

1. For each point $\mathbf{a}_i$ of the data set find the closest point $\mathbf{b}_i$ in the model set.
2. Given these putative correspondences, estimate the global motion transformation T between all the points by solving an Extended Orthogonal Procrustes Problem (see below).
3. Apply the transformation T to the data points.
4. If the distance between the two sets is less than a given threshold then quit, otherwise repeat from step 1.

The *Extended Orthogonal Procrustes Problem* (EOPP) [13] can be stated as follows: transform a given matrix A into a given matrix B by a similarity transformation (rotation, translation and scale) in such a way to minimize the sum of squares of the residual matrix. For reason that will be clear in the following, we consider instead the *Weighted Extended Orthogonal Procrustes Problem* (WEOPP) problem, with weights on the points. In formulae:

$$\arg\min_R \left\| (cRA + \mathbf{t}\mathbf{u}^\top - B)W \right\|_F^2 \quad \text{subject to } R^T R = I \tag{1}$$

where matrices A and B are $(3 \times p)$ matrices containing p corresponding point in 3-D space, R is (3×3) orthogonal rotation matrix, $\mathbf{t}$ is a (3×1) translation vector, c is scale factor, $\mathbf{u}$ is a $p \times 1$ vector of ones, W is a $(p \times p)$ diagonal matrix weighting the p points, and $\|\cdot\|_F$ denotes the Frobenius norm.

The solution to the the problem (derived in [1]) is based on the Singular Value Decomposition (SVD). Let

$$UDV^\top = A_w \left(I_p - \frac{\mathbf{u}_w \mathbf{u}_w^\top}{\mathbf{u}_w^\top \mathbf{u}_w} \right) B_w^\top \tag{2}$$

be the SVD decomposition of the matrix on the right-hand side[1], where $A_w = AW$, $B_w = BW$, and $\mathbf{u}_w = W\mathbf{u}$. The sought transformation is given by (we omit the scale c that is not needed in our case):

$$R = V \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \det(VU^T) \end{bmatrix} U^\top \tag{3}$$

$$\mathbf{t} = (B_w - RA_w) \frac{\mathbf{u}_w}{\mathbf{u}_w^\top \mathbf{u}_w} \tag{4}$$

The diagonal matrix in (3) is needed to ensure that the resulting matrix is a rotation [7]

The *Weighted Orthogonal Procrustes Problem* (WOPP) problem is a special case of WEOPP and the solution can be derived straightforwardly by setting $\mathbf{u} = \mathbf{0}$.

[1] Please note that $A\frac{\mathbf{u}_w \mathbf{u}_w^\top}{\mathbf{u}_w^\top \mathbf{u}_w}$ is a matrix of the same size as A with identical columns, each of them equal to the centroid of the points contained in A.

3 Hierarchical Articulated ICP

This section describes our contribution, namely the Hierarchical Articulated ICP algorithm for registering an articulate stick model to a cloud of points. We assume that the data are 3D points distributed roughly around the medial axes of the body's segments. They can reasonably come from the skeletonisation of a volumetric reconstruction of the body, as in [9, 5, 10]

The data are registered to the model using a hierarchical approach that starts from the torso and traverse the kinematic chain down to the extremities. At each step ICP computes the best rigid transformation of the limb that i) fits the data and ii) satisfy the kinematic constraints (namely, that the limb is connected to its ancestor through the joint) .

The closest point search works from the data to the model, by computing for each data point its closest point on the body segments. Only the matches with the current segment are considered, all the other should be – in principle – discarded.

However, the rotation in 3D space of a line segment cannot be computed unambiguously, for the rotation around the axis is undetermined. In order to cope with this problem we formulate a *Weighted* Extended Orthogonal Procrustes Problem and give a small but non-zero weight also to points that match the descendants of the current segment. In this way they contribute to constrain the rotation around the segment axis. Think, for example, of the torso: by weighting the points that match the limbs as well, even if they are still to be aligned, the coronal (aka frontal) plane can be recovered.

This is the complete algorithm described step by step:

Algorithm 2 HIERARCHICAL ARTICULATE ICP

Input: The model $\mathscr{S}$ composed by segments and the data set $\mathscr{A}$ of 3D points
Output: a set of rigid motions (referred to the kinematic chain) that brings the model onto the data

1. Traverse the body model tree structure using a level-order or a preorder traversal method.
2. Let $s_j \in \mathscr{S}$ be the current body segment.
3. Compute the closest points:
   (a) For each data point $\mathbf{a}_i \in \mathscr{A}$ and for each segment $s_\ell \in \mathscr{S}$ compute its projection $\mathbf{p}_{i\ell}$ onto the line containing s_ℓ ;
   (b) if $\mathbf{p}_{i\ell} \in s_\ell$ then add $\mathbf{p}_{i\ell}$ to $\mathscr{M}$ (the set of the closest-point candidates), otherwise add the endpoint of s_ℓ to $\mathscr{M}$.
   (c) Find $\mathbf{b}_i$, the closet point to $\mathbf{a}_i$ in $\mathscr{M}$.
4. Weight the points: If $\mathbf{b}_i$ belongs to s_j than its weight is 1, otherwise it is ε (chosen heuristically) for all the descendant and 0 for all the others.
5. If the distance of $\mathbf{b}_i$ to $\mathbf{a}_i$ is above a given threshold then the weight is set to 0.
6. Solve for the transformation of s_j with ICP using WEOPP for the torso and WOPP (only rotation) for the limbs.

The output of the algorithm represents the pose of the body. In a tracking framework, the pose obtained at the previous time-step is used as the initial pose for the current frame.

4 Experimental Results

The body tracker presented in section 4 has been implemented and tested on sequences taken from the HumanEva-I dataset [14]. All the sequences in HumanEva-I have been calibrated using the Vicon's proprietary software and the motion data saved in the common `c3d` file format. The dataset contains multiple subjects performing a variety of actions like walking, running, boxing, etc. In particular we used the sequences called *"S1 Box 3"*, *"S2 Throwcatch 3"* and *"S3 Jog 1"*.

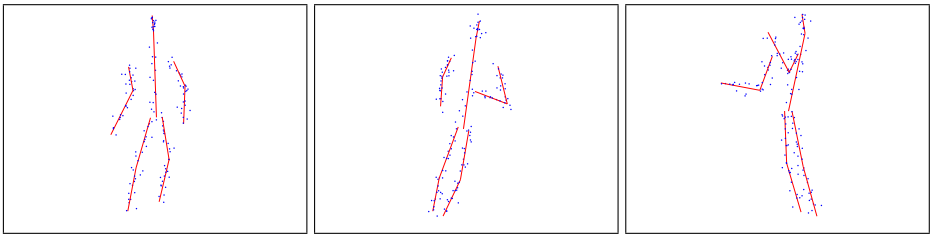

Figure 2: Sample frames of the synthetic data used in the experiments.The ground-truth stick figure and the data points corrupted by noise are shown.

The synthetic input data for our experiments has been created as follows (see Fig. 2). First the sequences have been sub-sampled at $40 fps$ instead of the original $120 fps$. Then, from each frame of the `c3d` file, reference points have been extracted and used to fit the skeletal model. The segments of the model are then sampled uniformly to obtain about 300 3D points. These points are finally perturbed using a Gaussian noise with an amplitude of half the hips distance.

Validation of the tracking is done by comparing the angles of the ground-truth with the angles of the computed model. Figure 3 reports the ground truth and estimated joint angles of the torso, right shoulder and right elbow in the three sequences. It can be seen that the estimated angles follows closely the ground truth, even without temporal smoothing. No significant mistracking occours.

For a quantitative comparison we computed the following angular error for each joint, in each frame of the sequence:

$$e(R_1, R_2) = \angle(R_1 R_2^\top) \tag{5}$$

where $\angle(\cdot)$ denotes the angle of the axix-angle representation of the rotation, and can be computed with $\angle(R) = \arccos((\mathrm{tr}(R) - 1)/2)$.

Mean and standard deviation of the error are shown in Table 1. The magnitude of the error is comparable with results reported in the state-of-the art literature [3, 15]. For the "leaves" limb, as head, lower arm and calf the angles error is usually bigger than the other, as expected.

The algorithm has been implemented in MATLAB, hence the performances are far from the desired real-time: it takes about 3.5 seconds to process a frame on a laptop with an Intel Core Duo Processor T2250.

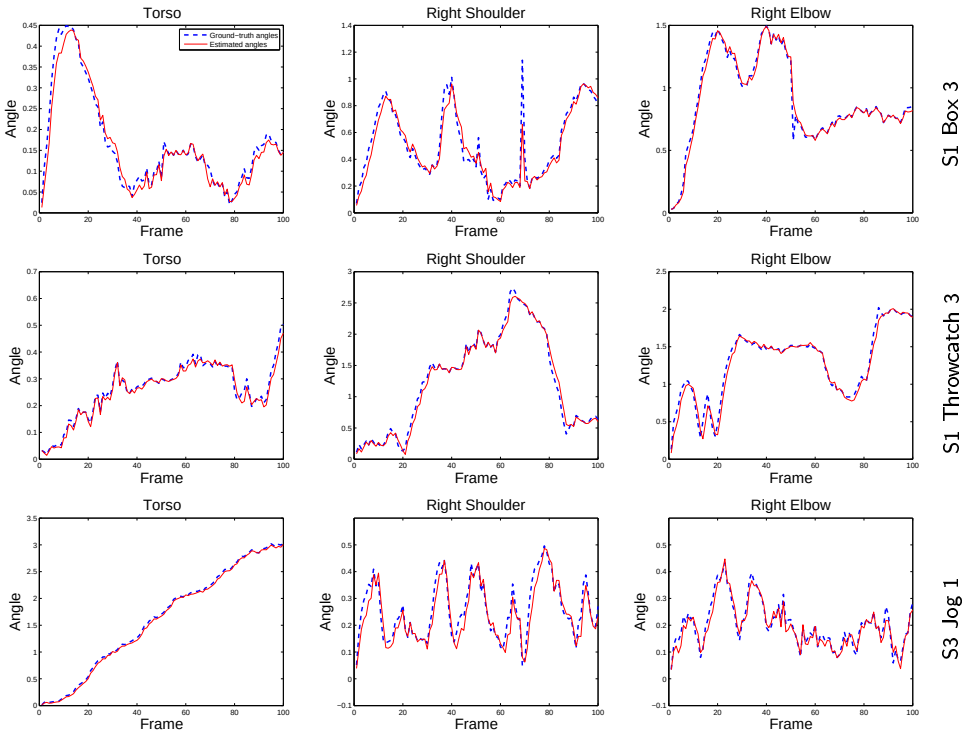

Figure 3: Plots comparing ground truth and estimated joint angles of the torso, right shoulder and right elbow in the three sequences used for the experiments (the sequence name is on the right).

5 Conclusions and Future Work

This paper has proposed a new ICP-based algorithm for tracking articulated skeletal model of a human body. The proposed algorithm takes as input $3D$ data points distributed around the torso and limbs medial axes. It fits the skeletal body model in each frame using a hierarchic tree traversal version of the ICP algorithm that preserves the connection of the segments at the joints. The proposed approach uses only the kinematic constraints and no other assumptions are made on the position of the body. This implies that we can recognized potentially all the body configuration.

The synthetic results presented here demonstrate the feasibility of the approach, which is is intended to be used in complete system for vision-based markerless human body tracking. Therefore, future work will consider improving performances by including a Kalman filter to smooth the estimates and provide a better prediction of the body's pose (which should allow ICP to converge in less iterations) and implementing the front-end of the pipeline, i.e., the shape from silhouette and skeletonisation modules that feeds the algorithm presented here.

Acknowledgments. This paper was partially supported by PRIN 2006 project 3-SHIRT.

		Box	*Throwcatch*	*Jog*
Torso	*Mean*	0.032	0.024	0.029
	Std. dev.	0.017	0.012	0.014
Neck	*Mean*	0.152	0.207	0.232
	Std. dev.	0.078	0.110	0.268
Left shoulder	*Mean*	0.074	0.071	0.063
	Std. dev.	0.039	0.041	0.030
Right shoulder	*Mean*	0.102	0.073	0.061
	Std. dev.	0.051	0.067	0.029
Left hip	*Mean*	0.181	0.120	0.111
	Std. dev.	0.126	0.089	0.075
Right hip	*Mean*	0.376	0.088	0.105
	Std. dev.	0.308	0.060	0.089
Left elbow	*Mean*	0.070	0.056	0.049
	Std. dev.	0.045	0.038	0.028
Right elbow	*Mean*	0.064	0.060	0.049
	Std. dev.	0.040	0.040	0.027
Left knee	*Mean*	0.061	0.049	0.052
	Std. dev.	0.035	0.028	0.030
Right knee	*Mean*	0.066	0.043	0.052
	Std. dev.	0.055	0.025	0.032

Table 1: Mean and standard deviation of the errors (in radians) for each joint of the body nfor the three sequences used in the experiments.

References

[1] D. Akca. Generalized procrustes analysis and its applications in photogrammetry. Technical Report, ETH, Swiss Federal Institute of Technology Zurich, Institute of Geodesy and Photogrammetry, 2003.

[2] D. Anguelov, D. Koller, H.-C. Pang, P. Srinivasan, and S. Thrun. Recovering articulated object models from 3D range data. In *Proc. of the 20th conference on Uncertainty in Artificial Intelligence*, pages 18–26, Arlington, Virginia, United States, 2004.

[3] A. Ashbrook, R. B. Fisher, N. Werghi, and C. Robertson. Construction of articulated models from range data. In *Proc. British Machine Vision Conference*, pages 183–192, 1999.

[4] P. J. Besl and N. D. McKay. A method for registration of 3-d shapes. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 14(2):239–256, 1992.

[5] G. J. Brostow, I. Essa, D. Steedly, and V. Kwatra. Novel skeletal representation for articulated creatures. In *ECCV04*, pages Vol III: 66–78, 2004.

[6] Y. Chen and G. Medioni. Object modelling by registration of multiple range images. *Image and Vision Computing*, 10(3):145–155, 1992.

[7] K. Kanatani. *Geometric Computation for Machine Vision*. Oxford University Press, 1993.

[8] S. Knoop, S. Vacek, and R. Dillmann. Modeling joint constraints for an articulated 3D human body model with artificial correspondences in ICP. In *Proc. 5th IEEE-RAS International Conference on Humanoid Robots*, pages 74–79, 2005.

[9] C. Ménier, E. Boyer, and B. Raffin. 3d skeleton-based body pose recovery. In *Proceedings of the 3rd International Symposium on 3D Data Processing, Visualization and Transmission*, Chapel Hill (USA), June 2006.

[10] B. Michoud, E. Guillou, and S. Bouakaz. Human model and pose Reconstruction from Multi-views. In *International Conference on Machine Intelligence*, November 2005.

[11] T.B. Moeslund, A. Hilton, and V. Kruger. A survey of advances in vision-based human motion capture and analysis. *Computer Vision and Image Understanding*, 103(2-3):90–126, November 2006.

[12] L. Mundermann, S. Corazza, and T.P. Andriacchi. Accurately measuring human movement using articulated ICP with soft-joint constraints and a repository of articulated models. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–6, June 2007.

[13] P. Schönemann and R. Carroll. Fitting one matrix to another under choice of a central dilation and a rigid motion. *Psychometrika*, 35(2):245–255, June 1970.

[14] L. Sigal and M.J. Black. Humaneva: Synchronized video and motion capture dataset for evaluation of articulated human motion. Technical Report CS-06-08, Brown University, Department of Computer Science, 2006.

[15] Y. Sun, M. Bray, A. Thayananthan, B. Yuan, and P.H.S. Torr. Regression-based human motion capture from voxel data. In *Proc. British Machine Vision Conference*, pages I:277–286, 2006.