Portuguese Journal of Sport Sciences
11 (Suppl. 3), 2011

THE EVOLUTION OF POSE ESTIMATION ALGORITHMS FOR 3D MOTION CAPTURE DATA: COPING WITH UNCERTAINTY

W. Scott Selbie
HAS-Motion Inc.

At the heart of many biomechanical analyses is the estimation of the pose (position and orientation) of a multi-segment model based on recording of 3D motion data. The principle assumption of most pose estimation algorithms is that sensors move rigidly with the body segments to which they are attached. It is accepted, however, that sensors attached to the skin move relative to the underlying skeleton and that this idiosyncratic Soft Tissue Artifact (STA) is challenging to model. Usually pose is estimated with discriminative algorithms that are ill-suited to the uncertainty of STA. Emerging algorithms based on probabilistic inference may mitigate STA by encoding the pose and any prior knowledge about the pose probabilistically, and capture the “artifacts” using a generative model.

KEY WORDS: optimization, soft tissue artefacts, probabilistic inference

Biomechanical models are often defined as set of rigid segments, with subject-specific scaling defined by palpable anatomical landmarks and anthropometric measurements. Segmental interactions are described by joint constraints permitting 0 to 6 degrees of freedom, with environmental interactions either modeled mathematically or recorded. The model is actuated by muscles represented by individual muscle elements or lumped parameters. Regardless of model complexity, a principal goal of modeling human movement is to represent a recorded performance in a hierarchical form suitable for analysis.
Mathematically, the assumption of rigid segments means that the kinematics of a segment are defined completely by a local segment coordinate system (SCS) fixed in the segment. Generally, the pose of an unconstrained rigid segment requires six independent variables, commonly known as degrees of freedom. Here we will restrict ourselves to recordings of 3D marker data from optical motion sensors. Pose estimation is the mapping of the markers $(\vec{m_i})$ to the pose of the hierarchical model, where the pose is represented by a vector from the Origin $(\vec{O})$ of the SCS to a root segment, and by three rotation matrices $(\mathrm{R}_1, \mathrm{R}_2, \mathrm{R}_3)$.

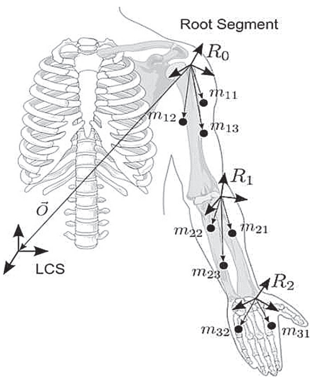

Figure 1: **An example of a three segment hierarchical model of the upper arm. Each rigid segment is defined by a local coordinate system aligned typically with anatomical axes. The pose of a root segment has 6 degrees of freedom (the location of the origin and the orientation) relative to the laboratory (LCS), and all other segments are mobilized with respect to a parent segment by 0 to 6 degrees of freedom.**

This presentation is an overview of approaches to pose estimation. The principle assumption of deterministic pose estimation algorithms is that markers move rigidly with their associated body segments (i.e., marker coordinates within the SCS are fixed). Markers attached to the skin, however, move idiosyncratically relative to the underlying skeleton (Cappozzo, Catani, Leardini, Benedetti & Della Croce, 1996). Such soft tissue artifact (STA) cannot be compensated adequately by deterministic methods, but probabilistic algorithms offer a new approach to this problem.

Portuguese Journal of Sport Sciences
11 (Suppl. 3), 2011

Discriminative Pose Estimation

Our nomenclature for the three discriminative algorithms is derived from (Lu & O'Connor, 1999). **Direct Pose Estimation** computes the SCS of each segment in a motion trial at each frame without assuming rigid segments, allowing the expected distribution of the markers and the segment length to change during movement. Thus, it is not possible to declare the number of degrees of freedom of a model, except instantaneously. There is no redundancy in the markers and there is no leniency in marker placement. If there is an error in the location of a marker, it will result in a direct error in the estimation of the SCS. Thus direct pose estimation is the least effective pose estimation algorithm.

For conciseness, we will present the **Segment Optimization** and **Global Optimization** pose estimations in a statistical framework. A point m_i attached rigidly to a segment, has a location represented by a static vector $\vec{a}_i$ in a Segment Coordinate System (SCS) and by a dynamic vector $\vec{b}_i$ in a Laboratory Coordinate System (LCS). For a single segment with markers $i = 1,2,\cdots,N$ the relationship between $\vec{a}_i$ and $\vec{b}_i$ is:

$$\vec{b}_i = \mathrm{R}\vec{a}_i + \vec{O} \tag{1}$$

where

R = rotation matrix from SCS to LCS

$\vec{O}$ = translation vector from SCS to LCS.

This can be generalized to a relationship between recorded data $\vec{d}$ and pose $\vec{q}$ of an entire model:

$$\vec{d} = f(\vec{q},\vec{w}) \tag{2}$$

where for *N* markers:

$\vec{d} = [\vec{b}_1, \vec{b}_2, \cdots, \vec{b}_N]$ the recorded marker data in the LCS,

$\vec{w} = [\vec{a}_1, a_2, \cdots, \vec{a}_N]$ the locations of the markers in the corresponding SCSs

and, for M generalized coordinates:

$\vec{q} = [\vec{q}_1, \vec{q}_2, \cdots, \vec{q}_M]$ represents the degrees of freedom in the model.

If we assume that all error $\vec{\in}$ in the data $\vec{w}$ is from noise, and is normal and independent, the generative model may take the form of the conditional probability distribution:

$$P(\vec{d}|\vec{q},\vec{w}) \sim N_d(f(\vec{q},\vec{w});\Sigma) \tag{3}$$

where

$N_d(f(\vec{q},\vec{w});\Sigma)$ is the multivariate normal distribution with a mean of $f(\vec{q},\vec{w})$ and a covariance matrix Σ of the sensor noise $\vec{\in}$

Equation 3 is read as the probability of seeing data $\vec{d}$ given the state of the model $\vec{q}$ and $\vec{w}$.

Expressing the distribution function explicitly:

$$P(\vec{d}|\vec{q},\vec{w}) = \frac{1}{\left((2\pi)^{\frac{3N}{2}} det(\Sigma)^{\frac{1}{2}}\right)} e^{-\left(\frac{1}{2}\right)\left(\vec{d}-f(\vec{q},\vec{w})\right)^T \Sigma^{-1}\left(\vec{d}-f(\vec{q},\vec{w})\right)} \tag{4}$$

and taking the negative log of this distribution function we get,

$$-\log\left(P(\vec{d}|\vec{q},\vec{w})\right) = C + \left(\frac{1}{2}\right)\left(\vec{d} - f(\vec{q},\vec{w})\right)^T \Sigma^{-1}\left(\vec{d} - f(\vec{q},\vec{w})\right) \tag{5}$$

In the deterministic algorithms the noise in one measurement is taken to be independent of all other measurements (thus $\Sigma = I$). For a deterministic solution we can reduce equation 5 to an Error function $E(\vec{q},\vec{w})$which is minimized with respect to $\vec{q}$ at each frame of data.

$$E(\vec{q},\vec{w}) = \left(\vec{d} - f(\vec{q},\vec{w})\right)^T \left(\vec{d} - f(\vec{q},\vec{w})\right) \tag{6}$$

In the simplest case of **Segment Optimization**, some of the elements of the vector $\vec{q}$ form a rotation matrix, so the problem is a constrained minimization problem, which can be solved using Lagrangian multipliers (Spoor & Veldpaus, 1980). This least squares solution can be considered a pattern recognition algorithm; the configured pattern of the tracking markers in each LCS is specified in a standing trial, and this pattern is fit to the homologous marker

Portuguese Journal of Sport Sciences
11 (Suppl. 3), 2011

configuration in each frame of motion capture data. The Segment Optimization approach to pose estimation is useful because it is straightforward and the solution has no local minima. Segment Optimization methods treat segments as independent (6 DOF), but links them implicitly by the motion capture data i.e. segments do not come apart because the subject does not come apart). Movement at a joint may be real (e.g. the knee joint axis is not) or may be caused by noise. Segment Optimization places no restrictions on marker placement, which allows exploration of marker placements that reduce STA (Cappozzo et al., 1997) or the number of markers on a segment (in an over specified system N>3, if noise and/or STA is uncorrelated, the computed pose will act to minimize the effects of the noise).

Lu & O'Connor (1999) introduced **Global Optimization** where physically realistic joint constraints are added to a model to minimize the effect of STA and measurement error. Global Optimization is dependent explicitly on the specification of a hierarchical model because the task is to identify an articulated figure consisting of a set of rigid segments connected with joints. Global Optimization is the search for an optimal pose of a multi-link model for each data frame such that the overall differences between the measured and model-determined marker coordinates are minimized in a least squares sense across all the body segments. It considers measurement error distributions in the system and provides an error compensation mechanism between body segments which can be regarded as optimal at the system level. For Global Optimization $f(\vec{q},\vec{w})$ is considerably more complex than for Segment Optimization, and more importantly, it is not possible to determine generally if the marker data are sufficient to compute a unique pose for the model, but it is beyond our page limitations to elaborate. Global Optimization is an extension to segment optimization because if all joints have six degrees of freedom, Global Optimization and Segment Optimization are equivalent.

As reported by Cereatti, Della Croce & Cappozzo (2006) there have been several attempts to modify optimization methods to minimize STA, but none of the approaches have been satisfactory because discriminative models have no mechanism to compute a compensation for systematic but idiosyncratic errors even when the presence of the STA can be modeled.

Probabilistic Pose Estimation

Todorov (2007) proposed that pose estimation from noisy motion capture data is better tackled by assuming uncertainty in the data and using well-established probabilistic algorithms based on Bayesian inference. Bayesian statistics is particularly well-suited for dealing with uncertain data because it provides a framework for making optimal inferences from uncertain information (Figure 2). For those of us who have always used discriminative models the probabilistic approach requires a conceptual leap because it seems to turn the problem on its head. The solution to the pose of the model given a set of data is oddly enough, not to solve for the pose directly (as in the discriminative model), but to solve for the possible data sets that are consistent with the pose in the context of a predicted pose; i.e. we must specify how we assume the data were produced.

The Bayesian formulation for the estimation of pose $\vec{q}$ is expressed as:

$$P(\vec{q},\vec{w}|\vec{d}) = \frac{P(\vec{d}|\vec{q},\vec{w})P(\vec{q},\vec{w})}{P(\vec{d})} \tag{7}$$

where the **Posterior** $P(\vec{q},\vec{w}|\vec{d})$ is the estimation of the model pose $\vec{q}$ based on the recorded data $\vec{d}$ and the marker locations $\vec{w}$, which is our goal. The **Normalization** term $P(\vec{d})$ is constant and *luckily* does not affect the estimation of pose (Todorov, 2007). The **Likelihood** $P(\vec{d}|\vec{q},\vec{w})$ is an estimate of the distribution of the data given the state of the system and the fixed marker locations $(\vec{q},\vec{w})$.The Likelihood is the Global Optimization estimate. The **Prior** $P(\vec{q},\vec{w})$ is described by a generative model (e.g. a probability distribution over possible poses), centered at a **predicted pose** $\vec{q}$, and its variance encodes how uncertain we are about the prediction. The probability distribution is based on extrapolations from previous states, and/or our understanding of the expected kinematics/kinetics of the movement.

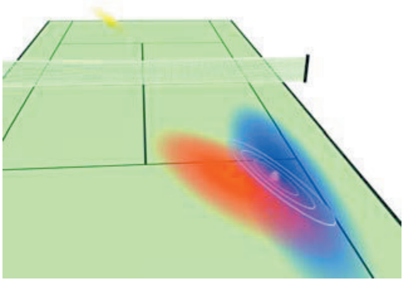

Figure 2: An example of Bayesian inference used to estimate optimally the bounce location of an incoming tennis ball. From vision we can estimate the *Likelihood* of different bounce locations (left hand ellipse). *Prior* experience as a tennis player may suggest that our opponent is an experienced player and thus the ball will tend to land close to the line (right hand ellipse). Integrating these two sources of information gives the control ellipse that denotes the *Posterior* (ball in center of inner ellipse) which indicates the most probably bounce location. Thus a player can get an optimal estimate of where the ball will land by using information in addition to following the flight of the ball (the Likelihood). (Reproduced with permission; Wolpert & Ghahramani [in press])

It is possible to account soft tissue deformations by modifying the Likelihood, based on the relationship between the residual vectors and the inferred joint angles, and changing the generative model to incorporate correlations between them. If there was no STA and all residuals were due to sensor noise, there would be no correlations. STA cause such correlations, because the tissue deforms in the same way every time you are in the same pose. Thus the correlations capture the effects of STA (to first order).

The Prior could be interpreted as a correction factor on the Global Optimization solution. The simplest **prior** would be to assume that the state at time t is very similar to the state at time t-1. The probability distribution would then take the form:

$$P_t(\vec{q}_t, \vec{w}) \sim N_d(\vec{q}_{t-1}; R_t) \qquad (8)$$

where R_t is an approximation to the covariance matrix. R_t declares how much you believe in the Prior compare to how much you trust the data. Models of STA and/or dynamics can be incorporated into the prior to minimize errors due to STA.

Probabilistic pose estimation algorithms based on Bayesian inference represent the next generation of algorithms. These algorithms provide a mechanism for enforcing dynamical consistency on the pose, for mitigating soft tissue artifact, for fusing data from redundant sensors, and given creative Prior rules promise for solving sparse data sets.

REFERENCES:

Cappozzo, A., Catani, F., Leardini, A., Benedetti, M. & Della Croce, U. (1996) Position and orientation in space of bones during movement. Experimental artifacts. *Clinical Biomechanics*, 11, 90–100.

Cappozzo, A., Cappello, A., Della Croce, U. & Pensalfini, F. (1997) Surface-Marker Cluster Design Criteria for 3D Bone Movement Reconstruction. *IEEE Transactions on Biomedical Engineering*, 44, 1165-1174.

Cereatti, A., Della Croce, U. & Cappozzo, A. (2006) Reconstruction of skeletal movement using skin markers: comparative assessment of bone pose estimators. *Journal of NeuroEngineering and Rehabilitation*, 3:7, 3-7.

Lu, T.W. & O'Connor, J.J. (1999). Bone position estimation from skin marker coordinates using global optimization with joint constraints. *Journal of Biomechanics*, 32, 129-134.

Spoor, C.W. & Veldpaus, F.E. (1980) Rigid body motion calculated from spatial co-ordinates of markers. *Journal of Biomechanics*, 13, 391- 393.

Todorov, E. (2007) Probabilistic inference of multijoint movements, skeletal parameters and marker attachments from diverse motion capture data. *IEEE Transactions on Biomedical Engineering*, 54, 11.

van den Bogert, A.J. & Su, A. (2007) A weighted least squares method for inverse dynamic analysis.

Computer Methods in Biomechanics and Biomedical Engineering, 11:1, 3-9.

Wolpert, D.M. & Ghahramani, Z. (in press). Bayes rule in perception, action and cognition. R. L. Gregory (Edt.) *The Oxford Companion to the Mind.*