

Data-based Motion Prediction

Julian J. Faraway
University of Michigan

Copyright © 2003 Society of Automotive Engineers, Inc.

Abstract

A complete scheme for motion prediction based on motion capture data is presented. The scheme rests on three main components: a special posture representation, a diverse motion capture database and prediction method.

Most prior motion prediction schemes have been based on posture representations based on well-known local or global angles. Difficulties have arisen when trying to satisfy constraints, such as placing a hand on a target or scaling the posture for a subject of different stature. Inverse kinematic methods based on such angles require optimization that become increasingly complex and computationally intensive for longer linkages. A different representation called stretch pivot coordinates is presented that avoids these difficulties. The representation allows for easy rescaling for stature and other linkage length variations and satisfaction of endpoint constraints, all without optimization allowing for rapid real time use.

The validity of this scheme also rests on the availability of motion capture data. There are two situations - in one case the user has access to a larger database of motions relevant to the particular problem while in the second case, the user collects a small amount of motion capture data concerning the task of interest. At the Human Motion Simulation Laboratory (HuMoSim) at the University of Michigan, we have collected several large databases on various types of automobile and materials handling motions. A prediction models based on these databases is presented.

Two contrasting prediction methods are demonstrated. One is parametric and uses functional regression analysis to predict the stretch pivot coordinates used in the postural representation as they vary over the time of the motion. This regression-type model allows the use of subject-

based variables such as stature and age and task based variables such as target location and object weight to influence the predicted motion. This also allows the scientific study of the effect of such factors using statistical significance testing. The second type of prediction method is based on the idea of nearest neighbor nonparametric regression. A small number (perhaps even just one) of motions is selected that have characteristics similar to the motion we wish to predict. These motions are then averaged in a special way using the stretch pivot coordinates to produce the predicted motion with the required features such as stature, target location etc. This method lends itself to the prediction of motion based on small special purpose motion capture databases collected by the user for some specific problem.

Keywords: Ergonomics, Motion Warping, Trajectory Modeling, Inverse Kinematics.

INTRODUCTION

Accurate predictions of human motion are valuable in many industrial applications. See [3] for a discussion of the utility of models for motion prediction. Many vehicle interiors and workplaces are first designed using a CAD system. Physical prototypes are useful for evaluating the ergonomic characteristics of the design, but these are expensive and time consuming especially when a design must proceed through several iterations. The ability to place an authentically moving virtual human within software such as *Jack* ([2]) helps the designer detect problems with the layout that can be rapidly corrected, thus speeding the design process. In some other applications, it is sufficient to produce motions that appear to be correct, but this application demands that the predicted motions be close to how people really move. Such demands impose greater costs in terms of the quantity of motion capture data required. Even so, the methodology pre-

sented below is generalizable and would be useful across a wide range of applications.

There is a huge amount of research on human motion. Some predictive models for human motion rely heavily on theories of human physiology and psychology. For example, see recent presentations at Digital Human Modeling Conferences, [10] and [1]. However, in recent years, motion capture technology has allowed us to collect large databases of human motion that make the construction of empirical models of motion feasible. The type of models we present here use relatively little knowledge of biomechanics, but rely on observations of real people from which they draw their validity.

In animation applications, when a particular motion is required, a common approach is to employ an actor to perform the desired motion, recording the trajectories of markers attached to the body. The motion is then mapped onto the animated character. Sometimes, the motion needs to be edited because the animated character has a different anthropometry than the actor or the motion needs to be retargetted, for example, a hand needs to reach to a somewhat different location. Techniques exist for doing this — for example see [21] who describe *motion warping* or *retargetting*, see [9], or *motion editing*, see [14]. See also [20].

It is expensive and impractical in many applications to employ an actor every time a new motion is needed. Instead one might assemble a database of motions that could be used to predict new motions when they are needed. In the example presented below, one wishes to predict how a worker will move a box to a shelf — one wants to know the positions of all parts of the body as the motion occurs. We collected a database of people moving boxes to shelves at a range of locations. We used this data to construct a model that takes the anthropometry and other features of the subject along with the desired position of the shelf as inputs and outputs the predicted motion. One of these models uses the notion of local averaging. The motions selected in the database, that are most similar to the one we want to predict, are then averaged to make the prediction. The advantage of averaging several motions compared to just a single motion is that the effect of individual variability is reduced and the user can be sure that a more typical motion is being viewed. Of course, the individual variability is still interesting and can be modeled — see [7].

An alternative to local averaging of motions is to regress the motion on the inputs. In this manner, all motions relatively similar to the desired motion can contribute information to the prediction according to the closeness of their inputs. This functional regression style of modeling is explained below.

The difficult part of this is to construct an appropriate method for averaging that satisfies constraints such as the fixed segment lengths of the human frame or placing the

box exactly on the shelf. In the sections below, the heart of the problem, the methodology and an application are described and a comparison to existing methods is made.

DATA

In this section, the data to which the methods were applied is described. The methodology presented here could be applied to data collected in different ways for different linkages, so this data serves as an example and not a boundary to what can be achieved.

In 1999, the Human Motion Simulation Laboratory (HUMOSIM) at the University of Michigan conducted a set of experiments concentrated on the motions of standing people performing reaches to a dispersed set of targets. In one experiment, subjects were required to move a small box with both hands from a position in front of the body at waist height to a shelf at the specified target. The subject then released the box and returned to the rest position, then reached back to the box, grasped it and returned it to the rest position. The left foot was held fixed, but the rest of the body was free to move.

A total of about 3000 motions were performed by a group of 20 subjects reaching to 30 targets where the four parts of the motion described above are counted separately. The subjects were selected to provide a means to assess the effects of anthropometry (height in particular), gender, and age on the motions. The subjects ranged from very short to very tall and from 20 to 60 years of age. Two different motion capture systems were simultaneously used to estimate joint center locations throughout the motions: an optical reflective marker system and an electromagnetic one. The observed markers were external to the body but were projected to joint centers within the body. The whole posture was described by 23 joint center locations located at the left and right feet, ankles, knees, hips, shoulders, elbows, wrist, front and back of hand together with the L5/S1 (small of the back), the C7/T1 (back of the neck), the sternoclavicular joint (supersternale), the tragon (center of the head) and the nasion (between the eyes).

The portions of the recorded motion where the subject was reaching were extracted based on the motion of the right hand. Because the data was collected at 20Hz and each motion lasted around 1 second on average, each motion was represented by the 3D coordinates of the 23 markers at around 20 timepoints on average. Due to occlusion or other data collection problems, some markers are missing during some parts of some motions. Where the amount of missing data was not substantial, the gaps were interpolated, otherwise the motion was discarded.

INVERSE KINEMATICS AND THE STRETCH PIVOT

Consider a chain of l jointed links in three dimensions as depicted in Figure 1. Suppose that one end of the chain is fixed at the origin, that the joints have full flexibility and

that the segments are of known lengths. One wishes to position the other end of the chain at some target T . For example, the chain might represent the arm and torso of a person who must reach to some target. The *inverse kinematics* problem is to position the rest of the chain to satisfy this endpoint constraint. There is no unique solution and so the challenge is to select the “best” solution.

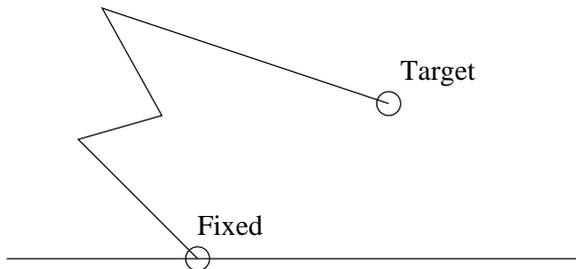


Figure 1: Inverse kinematics: Chain must be positioned to meet endpoint constraints

Many authors have proposed different criteria by which the solutions should be judged — see for example, [22] or [11]. Measures of comfort, effort, stress, etc. have been developed. Finding such solutions involves optimization subject to non-linear constraints. Such problems can be difficult and time consuming to solve. Furthermore, it seems doubtful that any complex human behavior can be encapsulated by the optimization of any single criterion. Other approaches, like motion warping, are based on data showing a posture (or postures) similar to the desired one. The task is then to modify the observed posture *as little as possible* to meet the constraints of desired target and segment lengths. This again involves a constrained optimization with the same practical problems with computation.

The key advance in this paper is to present a parameterization of the posture such that the constraints are always implicitly satisfied. This allows for rapid and simple computation of postures. We call this parameterization, *Stretch Pivot* coordinates. Alternative approaches that model the posture using the joint coordinates or angles describing the orientation of the links of the chain, will require further conditions to meet the link length and endpoint constraints and thus cannot be conveniently averaged.

The stretch pivot coordinates introduced here have the advantage that they can be averaged and still produce a valid configuration of the chain for any segment lengths and endpoints (provided the total segment length exceeds the distance between the endpoints). This allows the direct application of statistical methods.

Only $2l - 3$ parameters are necessary to describe a closed (i.e. endpoints fixed) l -link kinematic chain (two parameters for each segment minus three for the endpoint constraint). The key to success is selecting these parameters in a suitable way. Consider first a closed 2-link chain in three dimensions, like the shoulder, elbow and wrist linkage, where the endpoints, the shoulder and wrist, are in

fixed positions. Only one parameter is needed to describe this linkage, since the midpoint (the elbow) of the chain is constrained to lie on a circle whose center lies on and is orthogonal to an axis joining the endpoints. One needs only to specify the angle on this circle. We call this midpoint the *pivot* and we call this angle the *pivot angle*. Such an angle was used by [13], [18] and [19]. The angle is illustrated in Figure 2.

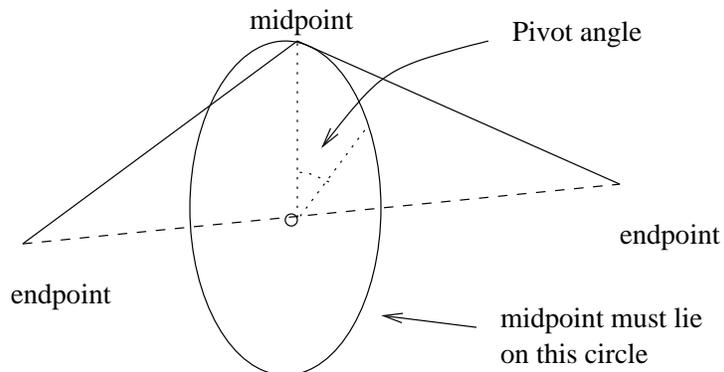


Figure 2: The pivot angle describes the location of the midpoint on the circle of its possible positions (3D view)

Now consider an l link chain and pick a marker in the middle of this chain. The selected marker does not need to be the middle in the sense of median, but assume, for now, that there are at least two links on each side of the selected marker. Let us arbitrarily call the two endpoint markers the proximal and the distal and the selected midpoint, the medial. Let l_p and l_d be the distances between the proximal and the medial and the distal and the medial respectively. If l_p and l_d are considered fixed, then the position of the medial relative to the proximal and distal may be described in terms of a pivot angle, θ_m lying on the circle orthogonal to, and whose center lies on, the axis joining the proximal and distal.

Let m_p and m_d respectively represent the total length of all the links joining the proximal and the medial and the distal and the medial. So

$$0 \leq l_p \leq m_p \quad \text{and} \quad 0 \leq l_d \leq m_d$$

Define $p_p = l_p/m_p$ and $p_d = l_d/m_d$. We call the p 's the *stretch* parameters. The position of the medial may be described in terms of the three parameters (θ_m, p_p, p_d) . Hence the name *stretch pivot*. See Figure 3.

Once the position of the pivot has been determined, the problem is reduced to two smaller problems. The same procedure may be repeated on the two halves of the chain recursively until the problem is reduced to a collection of two or three link chains. For the two link chains, only a pivot angle is needed to describe the middle marker. For a three link chain, the position of one of the midpoints may be described in terms of one pivot angle and one stretch parameter thereby reducing the problem to a two link chain requiring only one more pivot angle for description. This parameterization uses exactly $2l - 3$ parameters equal to the number of degrees of freedom in the chain.

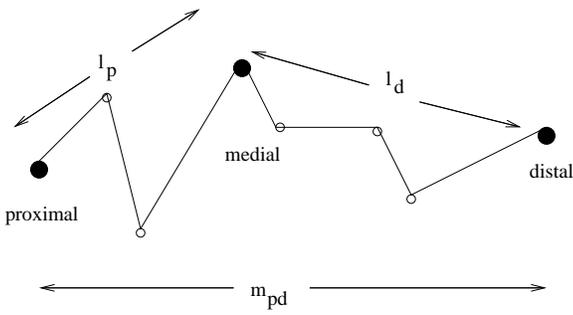


Figure 3: Stretch parameters illustrated. The distance between the proximal and medial if that part of the chain were fully extended is m_p while the corresponding distance for the medial to the distal is m_d . We define $p_p = l_p/m_p$ and $p_d = l_d/m_d$.

Depending on the particular configuration, further restrictions will apply to the pair (p_p, p_d) because the two halves of the chain must neither be too stretched or not stretched enough so that the halves might not join in the middle. Let m_{pd} be the distance from the proximal to the distal. The triangle inequality and being proportions means that (p_p, p_d) must satisfy the following constraints:

$$\begin{aligned}
 p_d m_d + p_p m_p &\geq m_{pd} \\
 p_d m_d + m_{pd} &\geq p_p m_p \\
 m_{pd} + p_p m_p &\geq p_d m_d \\
 0 \leq p_p \leq 1 &\quad 0 \leq p_d \leq 1
 \end{aligned}$$

Depending on the relative values of (m_p, m_d, m_{pd}) , this results in a range of potential values for (p_p, p_d) some of which are depicted in Figure 4.

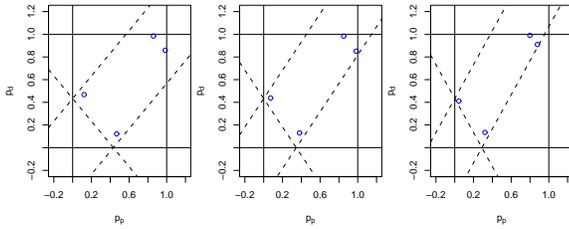


Figure 4: Allowable range for (p_p, p_d) lies within the unit square further restricted by the triangle inequality constraints shown as dotted lines. In all three panels, the proximal is at $(0,0,0)$ and the distal is at $(1,1,1)$. In the first panel, $m_p = 4, m_d = 4$, in the second $m_p = 5, m_d = 4$ and in the third $m_p = 6, m_d = 4$. The points in all three plots are the values of (p_p, p_d) corresponding to four different values of (q_s, q_d) : $(0.1, 0.9)$, $(0.1, 0.1)$, $(0.9, 0.1)$ and $(0.9, 0.9)$.

One needs to model a bivariate response, (p_p, p_d) , but the irregular shape of the domain being dependent on the relative values of (m_p, m_d, m_{pd}) , makes it difficult to build a prediction equation that will always produce valid values. The following solution is proposed — work with $(p_p + p_d, p_d - p_p)$ where the first parameter measures how stretched out the configuration is while the second mea-

sures the extent to which the midpoint is towards the proximal or distal.

For any given (m_p, m_d, m_{pd}) , one may determine the maximum, q_{smax} , and the minimum, q_{smin} , values that can be taken by the sum, $s = p_p + p_d$. We then define a new parameter q_s as

$$q_s = (s - q_{smin}) / (q_{smax} - q_{smin})$$

So q_s is a number between 0 and 1 representing the sum of the stretch parameters scaled within its allowable range.

Now treating, $s = p_p + p_d$ as fixed, we compute the range of the difference $d = p_d - p_p$ as (q_{dmin}, q_{dmax}) , and define q_d as

$$q_d = (d - q_{dmin}) / (q_{dmax} - q_{dmin})$$

So q_d is again a number between 0 and 1 representing the difference of the stretch parameters conditional on their sum, scaled within its allowable range. So any combination of (q_s, q_d) within the unit square may be mapped to a (p_p, p_d) for any (m_p, m_d, m_{pd}) combination.

Importantly, the predicted configurations for a given (q_s, q_d) for different (m_p, m_d, m_{pd}) combinations will be roughly homologous in the sense of placing the medial in about the same location to the extent possible given the values of (m_p, m_d, m_{pd}) . This feature is important if one is to use (q_s, q_d) as responses that can be sensibly averaged.

The values of $q_{smin}, q_{smax}, q_{dmin}, q_{dmax}$ for given (m_p, m_d, m_{pd}) are given in the Appendix. The stretch parameters must lie between 0 and 1 so it is simpler to model $\log q / (1 - q)$ as the response so that the predicted q will always lie between 0 and 1.

There remains the problem of the three link chains which require only one stretch parameter to describe. Again the specific configuration of (m_p, m_d, m_{pd}) will determine the range of this stretch parameter which will be a subset of $[0,1]$. The problem is just a simpler version of the one solved above. One may compute the range of this stretch parameter and the compute an appropriate $q \in [0,1]$ as above.

In describing a pivot angle, two problems may occur. Firstly, when $m_p + m_d = m_{pd}$, i.e. the two links are collinear, the angle cannot be defined at all. When this condition is close to occurring, there will be some instability in the pivot angle in that small changes in the midpoint may result in large changes in the pivot angle. However, given that the ultimate aim is to predict the midpoint and not the pivot angle, this problem is not serious.

The second problem lies in defining an origin for the pivot angle and is more troublesome. One way to define zero is to pick a direction, say the vertical, and project this direction onto the circle formed by the range of the pivot angle to define a zero. However, if the circle lies in the

horizontal plane, the zero will be undefined. Of course, this is unlikely to occur exactly in practice, but more seriously, there will be much instability in the pivot angle when the motion is such that the axis joining proximal and distal passes close to vertical. (Zero could also be defined using a plane, but the same problems will arise).

Of course, one could just choose a different polar direction but this will simply move the problem elsewhere. For some combinations of body markers, it is possible to pick a good polar direction since one might know that any axis that is likely to occur will not be close to the polar direction. However, some combinations, such as the wrist, elbow and shoulder, can clearly have axes in all directions. Some adaptive choice is necessary.

The approach taken is to avoid this zero problem entirely by defining the angle in terms of the normal vector to the plane passing through the pivot and the two endpoints. The three coordinates of the normal vector are then modeled. Admittedly, three parameters are introduced where only one appears to be needed but the polar instabilities are avoided and the normal vector does contain relevant information about the orientation of the endpoints and the pivot which can be usefully modeled.

MODELING DYNAMIC POSTURE

We want to compute averages (and other statistics) of motion and use these to make predictions of motion that satisfy endpoint constraints such as placing a box on a shelf for a subject of specified anthropometry. We have already shown how to model the interior joints of a chain. We now describe how to scale and represent these parameters as they change over time during a motion. We also present a parameterization of the motion of the endpoint convenient for modeling motion to and from specified targets. The complete posture will be represented by a conjunction of chains as described later in the application section.

TIME SCALING AND B-SPLINE REPRESENTATION

Many individual components of the motion, such as angles between body segments as they change over time, can be described as functions. For example, consider an axis joining the initial and final location of the hand. We can compute the orthogonal distance of the hand from this axis during motion, which we will call the *radial deviation*. Because we observe the data only at discrete timepoints, we have a sequence of observed values from the start to the end of the motion. These sequences are of different lengths because some targets are further away than others and people reach at different speeds. We rescale all these motions so that $t = 0$ is the start of the motion and $t = 1$ is the end of the motion. We can save the actual time taken as a possible predictor of the motion and, perhaps, to be predicted itself.

Plots of this distance for 20 subjects reaching with the right hand to a location somewhat to the left and front of

the body and about the same height as the initial position of the hand are shown in Figure 5. Since there are 30 targets and some reaches were replicated, a plot of the complete data for this measure would show considerably more curves. It would be reasonable to average (or take

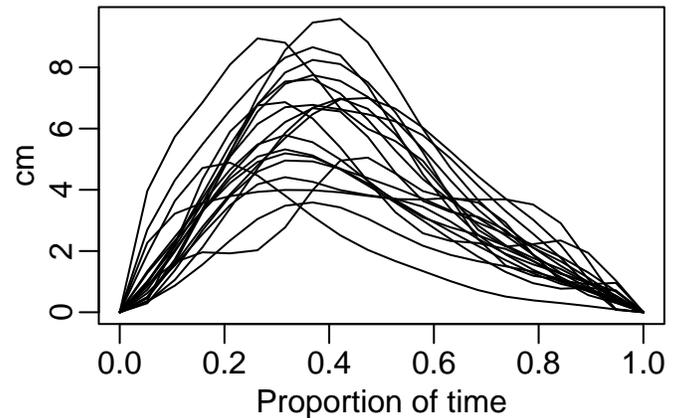


Figure 5: The radial deviation of the hand from a straight line path when reaching with the right hand to a location on the left for 20 subjects

averages of subsets of) the curves in Figure 5.

For reasons of compactness and ease of manipulation, we represent the curves as linear combinations of m cubic B-spline basis functions, $\psi_j(t)$. A curve $y_i(t)$ is represented as $\sum_{j=1}^m y_{ij} \psi_j(t)$ where the coefficients y_{ij} are estimated using least squares over the points at which $y_i(t)$ is observed.

Given that human motion is usually quite smooth, it is not necessary to have a large number of basis functions. In this particular application, we found that eight basis functions were sufficient. Any approximation error is dwarfed by the variation within and between individuals repeating the same motion so there is little value in using more basis functions. Using fewer basis functions is desirable especially when motion databases may be large. So each observed curve is represented by eight coefficients and the functional response is thereby converted into a multivariate response which is easier to work with. A parametric approach to modeling such functions in terms of the predictors may be found in [5].

TRAJECTORY REPRESENTATION In addition to predicting univariate functions, such as the radial deviation, one needs to predict the 3D curves formed by the trajectory of the hand or other body markers. One could simply model each of the three Cartesian coordinates of the trajectory, but this is unsatisfactory as it is not invariant to rotations of the coordinate system which one may well wish to make. A different parameterization has been chosen that has more interpretable components and assumes that one has specified the beginning and end of the trajectory (as would be the case for predicting the move-

ment of the hand for specified tasks).

We define $r(t)$ as the radial deviation at time t describing the orthogonal distance from the axis joining the endpoints, $p(t) \in [0, 1]$ as the proportionate progress along the axis at time t and $v(t) = p'(t)$ as the relative axial velocity. See the left panel of Figure 6 for a depiction of these quantities.

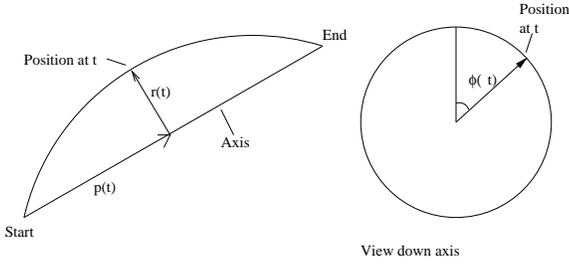


Figure 6: On the left, a side view of the reach is shown with the axis drawn as a straight line connecting the start and end of the reach. On the right, the view is down the axis, looking from the start towards the end.

Let $\phi(t)$ be the angle describing the position of the hand at time t on the circle orthogonal to the axis of the reach and whose center lies on this axis. We define $\phi(t) = 0$ to be the projection of the unit vertical vector onto this circle.

The trajectory is modeled using the triplet

$$(v(t), r(t), \phi(t)).$$

Since $v(0) = v(1) = r(0) = r(1) = 0$ by definition, one can accommodate this by omitting the first and last cubic B-spline basis function in the representation which, since these are, respectively, the only non-zero basis functions at $t = 0$ and $t = 1$, will ensure the desired property. Furthermore, one should not directly model $\phi(t)$ because it is an angle. Angles $2\pi - \varepsilon$ and ε are only 2ε apart but averaging their numerical values produces π which is diametrically opposite. For this reason, one models the responses $\cos \phi(t)$ and $\sin \phi(t)$ and then use the relation $\phi = \tan^{-1}(\sin \phi / \cos \phi)$ to predict ϕ which does respect the appropriate continuity properties of an angle.

Note that this representation contains no explicit information about the endpoints. This is convenient because when one predicts the hand motion one expects that the endpoints will be specified by the requirements of the application. One can directly impose these endpoints and then use the representation to reconstruct the rest of the trajectory.

A parametric approach to trajectory prediction may be found in [6].

So one is now able to describe the dynamic posture in terms of endpoint trajectories, pivot angles and transformed stretch parameters, which can each take any

value within their allowable range *independent* of the value of any of the other parameters. This *crucial property* allows us to independently average each parameter. Given any predicted posture parameters, one can construct a motion for any choice of link lengths long enough to reach the target.

MOTION WARPING OR RETARGETTING Suppose one observes a motion by one individual moving or reaching to some target and want to map it into a predicted motion for another individual moving or reaching to some other target. We describe the observed motion by

$$(\mathbf{e}, \mathbf{v}(t), \mathbf{r}(t), \phi(t), \theta(t), \mathbf{q}(t), \alpha(t))$$

where \mathbf{e} represents the endpoint information such as where the hand should begin and end, $\mathbf{v}, \mathbf{r}, \phi$ represents the trajectory information θ, \mathbf{q} represent the stretch pivot coordinates and α represents the unconstrained angular information (used to model the head motion — see later). All these elements have been written in bold to indicate that they may be vector quantities. For example, one may wish to model the trajectories of both hands independently. One may have more than one closed kinematic chain modeled by stretch pivot coordinates. There are a wealth of different possibilities that might be included within this framework. The individual for whom the motion was collected also has set of segment lengths \mathbf{l} or $\mathbf{l}(t)$ since some segment lengths (such as the torso) might be allowed to vary in time.

Now consider a new individual with segment lengths \mathbf{l}' with specified endpoints \mathbf{e}' . One may now take the $\mathbf{v}, \mathbf{r}, \phi, \theta, \mathbf{q}, \alpha$ from the original individual and use them to predict the motion for the new individual. Because of the properties of stretch pivot coordinates, one can be sure that such a motion will exist provided the segment lengths \mathbf{l}' of

the new individual are sufficiently long to stretch between the required endpoints \mathbf{e}' . The quality of the prediction will depend on how close \mathbf{e}' is to \mathbf{e} and how close \mathbf{l}' is to \mathbf{l} .

NEAREST NEIGHBOR METHODS

There is no reason to restrict ourselves to using a single motion for constructing the prediction. Suppose one has motion data with various endpoints, segment lengths and other characteristics which are similar to the case one wishes to predict. One may average the stretch pivot and trajectory parameters describing these motions to construct the prediction. This prediction will tend to be superior to the prediction based on a single motion for the usual reason that averages possess lower variances. One can consider more than just simple local averages — many of the well known ideas of nonparametric regression can be employed. See [16] for a review.

There is also the question of which motions in a database should be considered comparable in the sense of having

characteristics similar to those required for the prediction. For example, one may wish to predict the motion of a 48-year old, 50kg, 170cm tall woman of average strength lifting a box from specified shelf position. What characteristics should be used to judge similarity and how should they be combined? How many neighbors should be used or more generally how should the smoothing parameters be selected? Clearly there are a multitude of plausible schemes that could be used. Fortunately, one may evaluate any candidate scheme with respect to its predictive ability using crossvalidation. We will describe how this was done in practice below.

FUNCTIONAL REGRESSION ANALYSIS

An alternative approach to simply averaging motions is to use regression modeling. The motions are described a collection of functions over time. These can be modeled using functional regression analysis. The potential advantage of this method over the nearest neighbor method is that it is possible to model the marginal contribution of subject factors such as age and gender in a transparent manner and test their statistical significance.

Suppose the rescaled functions for some chosen quantity are given by $\mathbf{y}(t) = (y_1(t), \dots, y_n(t))$ where the subscript runs over the n collected motions. These functions might be expected to depend on certain covariates such as the location of the target being reached, the age and anthropometry of the subject and other factors. For the i^{th} curve, these predictors are collected in a vector x_i . Typically, the first term in this vector is one. We then propose a functional linear model:

$$y_i(t) = x_i^T \beta(t) + \varepsilon_i(t)$$

Notice that this is similar to a standard regression model but the response is now a function as is the error term $\varepsilon_i(t)$. The regression coefficients $\beta(t)$ are now a vector of functions. A general introduction to the area of functional data analysis may be found in [15]. The particular coefficient function for a given covariate will now represent the effect on the response of that covariate over the duration of the reach. One can now estimate $\beta(t)$ using least squares by

$$\hat{\beta}(t) = (X^T X)^{-1} X^T \mathbf{y}(t)$$

where X is the matrix whose rows are given by the x_i 's.

This formula cannot be directly applied since one cannot observe a $y_i(t)$ at all possible t . One approach is to approximate the functions on a grid of values. This was done in [4]. A fine grid of values is necessary for accurate representation which is somewhat inefficient. So instead, the curves are represented as linear combinations of m cubic B-spline basis functions, $\psi_j(t)$. A curve $y_i(t)$ is represented as $y_i(t) \approx \sum_{j=1}^m y_{ij} \psi_j(t)$ where the coefficients y_{ij} are estimated using least squares over the points at which $y_i(t)$ is observed.

Given that human motion is usually quite smooth, it is not necessary to have a large number of basis functions. In this particular application, one finds that eight basis functions were sufficient. So each observed curve is represented by eight coefficients and the functional response is thereby converted into a multivariate response

Thus the model can be written in the form

$$Y_{n \times m} \Psi_{m \times 1}(t) = X_{n \times p} B_{p \times m} \Psi_{m \times 1}(t) + \varepsilon_{n \times 1}(t)$$

or factoring out the $\Psi(t)$, one can write it in the simpler form:

$$Y_{n \times m} = X_{n \times p} B_{p \times m} + \varepsilon_{n \times m}$$

which is now a multivariate multiple regression model where the coefficient matrix B may be estimated using least squares:

$$\hat{B} = (X^T X)^{-1} X^T Y$$

One may then use the standard methods of statistical inference using this modeling approach. Details of such methods may be found in texts such as [12]. For prediction and interpretation purposes, it is necessary to transform back from this basis function representation to the original form by

$$\hat{\beta}_{p \times 1}(t) = \hat{B}_{p \times m} \Psi_{m \times 1}(t)$$

One can predict future responses given a new predictor value x_0 by

$$\hat{y}_0(t) = x_{01 \times p} \hat{B}_{p \times m} \Psi_{m \times 1}(t)$$

More details on this approach may be found in [5].

IMPLEMENTATION

In this section, we describe the particular implementation used for modeling the data described above and how well the models performed. These choices would need to be reconsidered and recombined for other types of motion but it is worth describing the choices here to understand the considerations involved.

We identified a chain stretching from the left foot to the right hand and containing 10 links. Other linkages branch off this chain. In this example, these would be the right leg, left arm and the head. The main linkage is modeled first. These secondary linkages are then modeled in a similar fashion once the main linkage has been predicted. See Figure 7.

The selection of pivots is essential to the implementation of the stretch pivot coordinates. First consider the primary kinematic chain linking the left ball of foot to the right hand. The pivots selected were the c7/t1, the left hip, the right shoulder, the right elbow and the left knee where the order indicates the order of precedence. The remaining markers on this chain required only a pivot angle to describe.

The right ball of foot was free to move in this experiment. We modeled the Cartesian coordinates $(x(t), y(t), z(t))$ directly for this marker. Once the position of the right foot was determined, the right knee was used as a pivot with

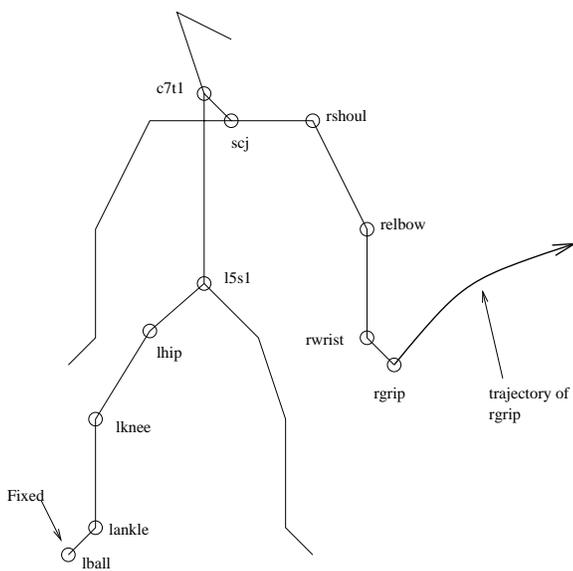


Figure 7: Schematic of human figure. The critical kinematic chain linking the endpoints at the left ball of foot which, in this example, does not move and the right grip, which moves along a trajectory that we have already predicted. One needs to predict the locations of the nine interior markers of the ten link chain as labeled in the diagram. When this is done, the three smaller chains consisting of the right leg, left arm and head can be predicted.

the remaining markers on this chain required only a pivot angle to describe.

When the left hand was holding the box, the angles describing the orientation of the box were modeled. Thus the position of the left grip was derived from the right grip together with the predicted box angles. For the motions where the left grip was empty (returning to the rest position or reaching to grasp the box on the shelf), the trajectory of the left hand was independently modeled in the same way as the right hand.

The two link head chain did not have a fixed endpoint so that the angles describing the orientation were simply modeled using local angles.

In order to make a prediction, it is necessary to provide the segment lengths (the distances between adjacent markers). Various studies have been made concerning these lengths and how they vary as a function of anthropometry. In this case, we fit simple regression models to describe these lengths as a linear function of stature. For this small number of subjects, there was no evidence that more than stature alone would be helpful in predicting these lengths. Other segment length specifications can easily be used.

Various schemes for selecting the nearest neighbors were contemplated and then evaluated using crossvalidation. Each motion in the database was left out in turn. We predicted this left out motion using all the other motions and compared this prediction to the motion derived from the parameters of the left out motion. Because link lengths

based on those predicted for a person of that height and the B-spline based representation of the parameters was used, the comparison was not exactly between prediction and observed motion data. Such a comparison would be possible but it would require using the actual link lengths which would typically not be known in the usual predictive scenario.

One obstacle to constructing a scheme for measuring similarity is that the characteristics such as the target location and age of the subject are on quite different scales. For this reason, we considered only information based on the ranks of $\|x_{data} - x_{predict}\|$ where the norm $\|\cdot\|$ is the absolute value for univariate quantities like height and Euclidean distance for the target coordinates. The particular scheme that gave the lowest crossvalidated error was based on the rank of sum of the target ranks and the stature ranks. When k nearest neighbors were considered, the nearest neighbor was given weight k in the average, the second nearest weight $k - 1$ i.e. triangular weights. Other schemes that used age, gender, weight, strength, time taken for the move and various weighting schemes for the average were inferior to that chosen. For the data considered here, 5 neighbors gave the best results. It's likely that the best scheme may differ for other databases but it would be possible to evaluate a variety of schemes in new settings to find the best choice. For this data, for the optimal scheme and number of neighbors, the median of the average difference between the predicted and observed motions over the markers and the complete motion was 8.1cm.

There were 100 duplicated motions in the database where the same subject reached to the same target. The median distance between these pairs of motions averaged over the markers and the complete motion was 6.8cm. This figure represents the natural variation in reaching motions — it would not be reasonable to expect any prediction model to predict with greater accuracy than this. This suggests it would not be possible to make substantial improvements on the method we have described above for this dataset. Clearly, one can't say how well the method will perform for other datasets, but this does give some cause for optimism.

One important advantage of using stretch pivot coordinates over other methods that require optimization is that it is fast. To predict all four phases of the motion required only 0.01 seconds on an 850Mhz Pentium III PC. The algorithm was implemented in C but without any special efforts at optimization. This speed makes it practical for applications where rapid animation is important.

We have also applied the functional regression method to model the functions that describe the dynamic posture. We have found that it is impractical to fit models to the entire target region. The patterns of motion change with the target and it is difficult to devise a single functional form to describe this. Instead, we subdivide the target area into regions determined by proximity to an experimental target

and fit a regression model to the data within this region. When a new prediction is needed, the sub-model corresponding to the nearest experimental target will be used.

Motion capture technology is far from perfect. Problems with the data collection occasionally result in damage to the data. This includes missing and improper measurements. With a large database such as this, it is difficult to detect all such problems. Movements which are clearly incorrect are easily detected but those that have milder problems can be hard to find particularly by automatic screening methods. For these reasons, we used robust regression methodology in the functional regressions and used the median rather than the mean in the nearest neighbor averaging.

COMPARISON

There are many competing methods of human animation. A large class of such methods is not based directly on data, but on theories about how people move. Others are content to directly use motion capture data for animation, perhaps hand-editing the motion for their requirements. However, my method is most reasonably compared to other methods that use motion capture data to predict motion with specified constraints.

Two recent papers provide a review of the current state of research in this type of problem. [17] present a parameterization of a popular seven degree of freedom model for the shoulder, elbow and wrist with the hand constrained to a target as in this case. Many previous approaches to this problem have specified the angles describing the orientation of the limb segments and then resolved the redundancy caused by the over-parameterization by optimizing some criterion to fix the position of the arm. In contrast, the advantage of their approach, and my own, is that the position of the arm can be explicitly and easily calculated without optimization. [17] describe the competing methods and show that their explicit method is typically orders of magnitude faster than those based on optimization.

My method also allows the explicit computation of posture without optimization and will thus enjoy a similar speed advantage over optimization based approaches. Furthermore, it applies to a kinematic chain of any number of links and the amount of computation is linear in the number of links. In contrast, the difficulties with optimization based methods rapidly increase with the number of links — complexity, local minima and sensitivity to initial values all become more problematic. This approach was developed after failing to find a workable optimization-based inverse kinematics solution to a full body linkage.

[8] addresses motion prediction rather than just static posture. He reviews the existing methods for editing a single motion to meet new objectives or constraints. Problems arise in editing successive frames of the motion to meet constraints while maintaining continuity. My method inherits the smoothness and characteristics of motion capture

data while adjusting to new constraints. It also does this much faster than methods that require some optimization.

This representation has some weaknesses. For example, most joints have mobility restrictions but this method does not explicitly respect these restrictions. Since the method is based on empirical data, these limitations will tend to be exceeded only on larger extrapolations. Nonetheless, the problem may occur. In such cases, the dynamic posture predicted by this method may serve as an initial estimate requiring some correction. For example, this standing figure occasionally has a predicted ankle position that would put the heel through the floor. This can be corrected by moving the ankle upwards to the first valid position while simultaneously moving the knee outwards in the originally predicted plane of the ankle, knee and hip. Some other methods explicitly respect these restrictions, but at substantial cost in speed of computation and complexity. One finds it substantially faster and simpler to make local corrections to the posture *if* they are needed.

Furthermore, more complex constraints may be needed. In this example of the two-handed box lift, it might happen that the predicted left hand position is too far for the left arm to reach. Although this hasn't been a problem in practice, one would prefer to have a parameterization analogous to the stretch pivot for more complex linkages than simple chains. Such parameterizations that avoid optimization should be the topic of future work.

DISCUSSION

Most of the weaknesses in the methodology described above lie in the fact that it uses no biomechanical knowledge. There are limits on motion imposed by considerations of balance and joint mobility while other considerations such as comfort mean that some postures, while not forbidden, are discouraged. There are other biomechanical and psychological considerations that affect motion. Naturally, these topics have been the subject of considerable research and yet it is clear that a complex combination of these considerations would be necessary to construct a predictive model of any versatility.

Our model is empirical and is only so good as the data it is based upon. In our case, the predictive performance declines as one specifies inputs that are far from those observed in the experiment. For example, if one inputs a target far to the left of the subject, one cannot expect a good prediction since all the targets in the experiment were the right of the sagittal plane. Such failings are inevitable and can be rectified by collecting more data with the required input conditions. Alternatively, biomechanical laws can help by detecting violations and suggesting the appropriate correction.

The current method also contains no notion of obstacle avoidance. If the hand must avoid some obstacle then the trajectory prediction must be suitably modified. This seems feasible enough but more problematic are obsta-

cles that intersect with other parts of the body. An even more difficult problem occurs when the predicted position of the body intersects with itself. For example, on reaches to targets near the feet, one needs avoid predictions where the hand passes through the knees.

Another difficulty is that DHM models, such as *Jack*, *Delmia* and others, all use different representation of the posture. We have used 23 markers to represent the posture but this ignores various details. Furthermore, different posture models use different linkages and impose different constraints. To make practical use of our models requires a mapping to and from the our linkage to the DHM linkage. There is no straightforward solution to this as it requires an in depth understanding of the chosen DHM linkage which differ from vendor to vendor. We have made such a mapping for *Jack* and our industrial partners have made a mapping to the *Delmia* figure.

So to conclude, our method is not the complete solution to the problem of human motion prediction. It enables some extrapolation from a database of motions but will begin to fail for more extensive extrapolations. The scope of the method depends on the quality and range of the database of motions.

ACKNOWLEDGMENTS

Thank you to Don Chaffin and the other members of the HUMOSIM laboratory. The work was made possible by the funding of Ford, General Motors, International Truck, Daimler Chrysler, Lockheed Martin, the US Postal Service and US Army TACOM.

APPENDIX

Range of Sum and Difference of Stretch Parameters

q_{smax} :

Set $q_{smax} = 2$ but

- if $m_d - m_p \geq m_{pd}$ set $q_{smax} = (m_{pd} + m_p + m_d) / m_d$
- if $m_p - m_d \geq m_{pd}$ set $q_{smax} = (m_{pd} + m_p + m_d) / m_p$

q_{smin} :

- If $m_{pd} < m_d$ then set $a = m_{pd} / m_d$ else set $a = (m_{pd} - m_d + m_p) / m_p$
- If $m_{pd} < m_p$ then set $b = m_{pd} / m_p$ else set $b = (m_{pd} - m_p + m_d) / m_d$
- Set $q_{smin} = \min(a, b)$.

Let $x = p_p + p_d$.

q_{dmax} :

- Set $a = 2(m_{pd} + xm_p) / (m_p + m_d) - x$
- If $m_p > m_d$ set $b = x - 2(m_{pd} - xm_d) / (m_p - m_d)$ else set $b = 2$
- Set $q_{dmax} = \min(x, 2 - x, a, b)$

q_{dmin} :

- Set $a = x - 2(m_{pd} + xm_d) / (m_p + m_d)$
- If $m_p < m_d$, set $b = x - 2(m_{pd} - xm_d) / (m_p - m_d)$ else set $b = -2$.
- Set $q_{dmin} = \min(-x, x - 2, a, b)$

REFERENCES

- [1] K. Abdel-Malek, W. Yu, M. Jaber, and J. Duncan. Realistic posture prediction for maximum dexterity. In *2001 SAE Digital Human Modeling for Design and Engineering Conference*, number 01-2110, 2001.
- [2] N. Badler, C. Phillips, and B. Webber. *Simulating Humans: Computer Graphics Animation and Control*. Oxford Univ. Press, New York, 1993.
- [3] D. Chaffin. Simulation of human reach motions for ergonomics analyses. In *2002 SAE Digital Human Modeling for Design and Engineering Conference*, pages 9–24, 2002.
- [4] J. Faraway. Regression analysis for a functional response. *Technometrics*, 39:254–261, 1997.
- [5] J. Faraway. Modeling reach motions using functional regression analysis. *SAE Technical Paper 2000-01-2175*, 2000.
- [6] J. Faraway. Modeling hand trajectories during reaching motions. submitted for publication to *Journal of the American Statistical Association*, 2001.
- [7] J. Faraway and J. Hu. Modeling variability in reaching motions. In *2001 SAE Digital Human Modeling for Design and Engineering Conference*, number 01-2094, 2001.
- [8] M. Gleicher. Comparing constraint-based motion editing methods. *Graphical Models*, 63:107–134, 2001.
- [9] M. Gleicher and P. Litwinowicz. Constraint-based motion adaptation. *The Journal of Visualization and Computer Animation*, 9(2):65–94, 1998.
- [10] E. Gu, Y. Teng, and L. Oriet. A study of human joint torque distribution for optimal posture auto-generation. In *2002 SAE Digital Human Modeling for Design and Engineering Conference*, pages 421–430, 2002.

- [11] S. Hsiang and M. Ayoub. Development of methodology in biomechanical simulation of manual lifting. *International Journal of Industrial Ergonomics*, 17:271–288, 1995.
- [12] R. Johnson and D. Wichern. *Applied Multivariate Statistical Analysis*. Prentice Hall, New Jersey, 3rd edition, 1992.
- [13] J. Korein. *A Geometric Investigation of Reach*. MIT, Cambridge, MA, 1985.
- [14] J. Lee and S. Y. Shin. A hierarchical approach to interactive motion editing for human-like figures. In *ACM SIGGRAPH 99, Annual Conference Series*, pages 39–48, 1999.
- [15] J. Ramsay and B. Silverman. *Functional Data Analysis*. Springer, New York, 1997.
- [16] J. Simonoff. *Smoothing Methods*. Springer-Verlag, New York, 1996.
- [17] D. Tolani, A. Goswami, and N. Badler. Real-time inverse kinematic techniques for anthropomorphic limbs. *Graphical Models*, 62:353–388, 2000.
- [18] X. Wang. A behavior-based inverse kinematics algorithm to predict arm prehension postures for computer-aided ergonomic evaluation. *Journal of Biomechanics*, 32:453–460, 1999.
- [19] X. Wang. Three-dimensional kinematic analysis of influence of hand orientation and joint limits on the control of arm postures and movements. *Biological Cybernetics*, 80:449–463, 1999.
- [20] D. Wiley and J. Hahn. Interpolation synthesis of articulated figure motion. *IEEE Computer Graphics and Applications*, 17:39–45, 1997.
- [21] A. Witkin and Z. Popovic. Motion warping. In *Computer Graphics (Proc. SIGGRAPH '95)*, 1995.
- [22] X. Zhang, A. Kuo, and D. Chaffin. An optimization-based differential inverse kinematics approach for modeling moderately complex three-dimensional dynamic seated postures. *Journal of Biomechanics*, 31:1035–1042, 1998.

CONTACT

Julian Faraway, Department of Statistics, University of Michigan, Ann Arbor, MI 48104, USA. Telephone (734)-763-5238, Fax (734)-576-2173, Email faraway@umich.edu