

Human Animation Using Nonparametric Regression

Julian FARAWAY

This article describes how to build a nonparametric regression model for the prediction of human motions using motion capture data. The method works by selecting a few similar motions from a database and then averaging them. An appropriate parameterization is developed for representing the motions composed of the trajectories of the endpoints such as the hands and stretch pivot coordinates to represent the interior joints such as the shoulders or knees. Within this representation, local averaging may sensibly occur while preserving specified constraints like the lengths of the body segments and final location of the endpoints. A nearest neighbor regression like method is developed and a cross-validation procedure implemented for selecting the best number of neighbors and appropriate notions of similarity in the motions selected for averaging. The methodology is illustrated with the development of a prediction model for two-handed standing lifts with applications to ergonomics.

Key Words: Ergonomics; Inverse kinematics; Motion warping; Trajectory modeling.

1. INTRODUCTION

Accurate predictions of human motion are valuable in many fields, for example, neuroscience, kinesiology, sports science, and entertainment applications in animation and gaming. The particular application that motivated this work comes from ergonomics. Vehicle interiors and industrial work spaces are usually designed using a CAD system. When evaluating a design, it is important to pay special attention to ergonomics. Can a smaller person reach a control without colliding with an intervening object? What stresses would be placed on the shoulder and back when an older person lifts a box from a shelf at a given position? Answers to these and similar questions require an understanding of how people move within a given environment. Traditionally, physical prototypes of the design would be constructed and tested with real people. However, sequentially improving the design in response to the problems discovered is very time consuming. Our aim is to produce a model

Julian Faraway is Professor, Department of Statistics, University of Michigan, 460 West Hall, Ann Arbor, MI 48109 (E-mail: faraway@umich.edu).

©2004 American Statistical Association, Institute of Mathematical Statistics,
and Interface Foundation of North America
Journal of Computational and Graphical Statistics, Volume 13, Number 3, Pages 337-353
DOI: 10.1198/106186004X2507

that predicts how people will move within a space that can be incorporated into the design package. This will enable the designer to discover many deficiencies in the design and make improvements before any physical prototypes are constructed. In the automotive industry, it is important to bring new products to market quickly, so speeding up the design process is valuable. Also, poor design is a common cause of workplace injuries so tools that can improve work area layouts are desirable.

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. In recent years, however, 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 presented here use relatively little knowledge of biomechanics, but rely on observations of real people.

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 retargeted; for example, a hand needs to reach to a somewhat different location. Techniques exist for doing this—for example, see Witkin and Popovic (1995), who described *motion warping*; see Gleicher and Litwinowicz (1998), who discussed *retargeting*; and see Lee and Shin (1999) for *motion editing*.

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 following example, we wish to predict how a worker will move a box to a shelf—we want to know the positions of all parts of the worker's body as the motion occurs. We collected a database of people moving boxes to shelves at a range of locations. We used these 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. The model uses the notion of local averaging. We select the motions in our database that are most similar to the one we want to predict and average them to make the prediction. So we have a nonparametric regression model for human animation.

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. Section 2 describes how the position of the joints of the body can be parameterized once the positions of the endpoints, such as the hands and the feet, are known. Section 3 explains how the curves that describe the motion are represented, how the trajectories of endpoints, such as the hand, are modeled and how motions can be combined using nearest neighbor nonparametric regression to make new predictions. Section 4 introduces the application and Section 5 describes the implementation of the method on these data. Section 6 closes with a comparison to other methods.

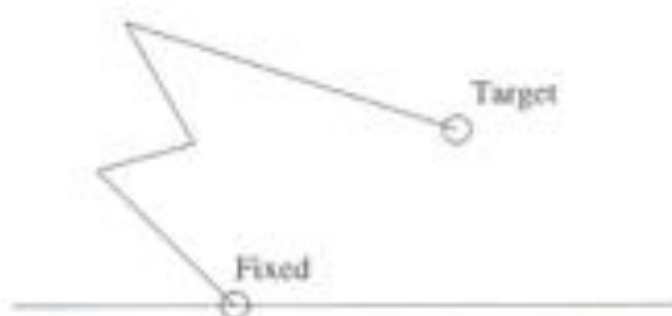


Figure 1. Inverse kinematics: Chain must be positioned to meet endpoint constraints.

2. INVERSE KINEMATICS AND THE STRETCH PIVOT

We start with the heart of the problem. 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. We wish 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.

Many authors have proposed different criteria by which the solutions should be judged—see, for example, Zhang, Kuo, and Chaffin (1998) or Hsiang and Ayoub (1995). Measures of comfort, effort, stress, and so on have been developed. Finding such solutions involves optimization subject to nonlinear constraints. Such problems can be difficult and time consuming to solve. Furthermore, it seems doubtful that 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 article 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. I 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 needed to describe a closed (i.e., endpoints fixed) l -link kinematic chain (two parameters for each segment minus three for the endpoint constraint).

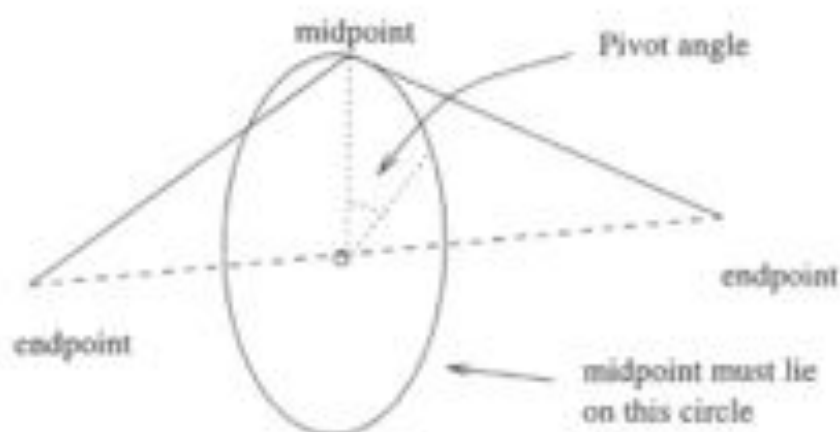


Figure 2. The pivot angle describes the location of the midpoint on the circle of its possible positions (2D view).

Consider first a closed two-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, because 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. We need only 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 Korein (1985) and Wang (1999a, b). The angle is illustrated in Figure 2.

Now consider an l -link chain and pick a joint in the middle of this chain. The selected joint 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 joint. Let us arbitrarily call the two endpoint joints 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

$$0 \leq l_p \leq m_p \quad \text{and} \quad 0 \leq l_d \leq m_d. \quad (2.1)$$

Define $p_p = l_p/m_p$ and $p_d = l_d/m_d$. We call the p 's the *stretch parameters*. We may describe the position of the medial 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 parameterization problem is reduced to two smaller problems. We may repeat the same procedure on the two halves of the chain recursively until we are reduced to a collection two- or three-link chains. For the two-link chains, only a pivot angle is needed to describe the middle joint. For a three-link chain, we may describe the position of one of the midpoints 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

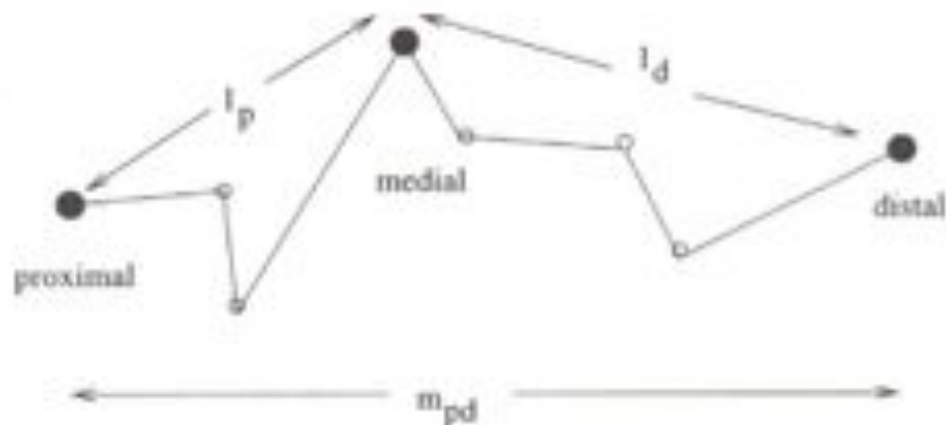


Figure 3. Stretch parameters illustrated. The distance between the proximal and medial if that part of the chain were fully extended is m_p while the corresponding distance for the medial to the distal is m_d . We define $p_p = l_p/m_p$ and $p_d = l_d/m_d$.

equal to the number of degrees of freedom in the chain.

There is some choice in the sequence in which joints are designated as pivots. Some experimentation may be necessary in given applications to determine the best choice of this sequence.

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:

$$p_d m_d + p_p m_p \geq m_{pd} \quad (2.2)$$

$$p_d m_d + m_{pd} \geq p_p m_p \quad (2.3)$$

$$m_{pd} + p_p m_p \geq p_d m_d \quad (2.4)$$

$$0 \leq p_p \leq 1 \quad 0 \leq p_d \leq 1 \quad (2.5)$$

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.

We need 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. I propose the following solution—we will work with $(p_p + p_d, p_d - p_p)$ where the first parameter measures how stretched out the configuration is and the second parameter measures the extent to which the midpoint is toward the proximal or distal.

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

$$q_s = (s - q_{s,min}) / (q_{s,max} - q_{s,min}) \quad (2.6)$$

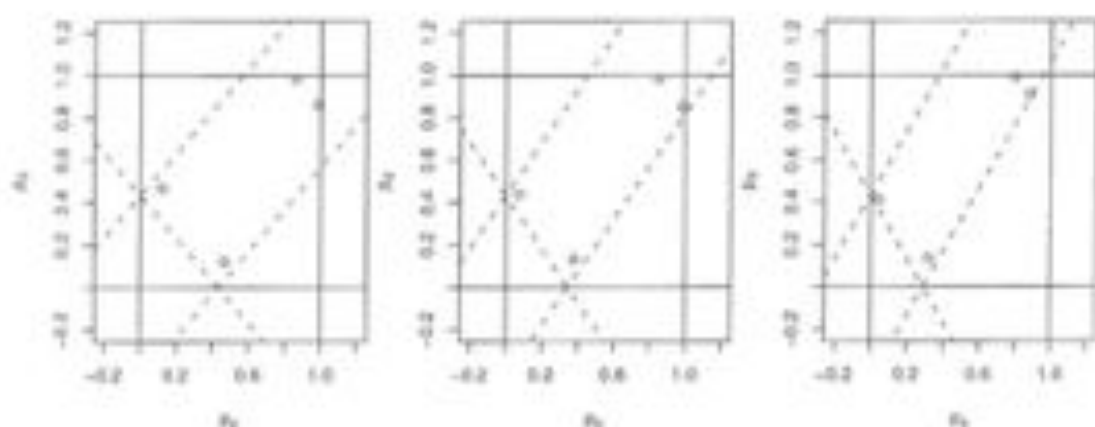


Figure 4. Allowable range for (p_y, p_d) lies within the unit square further restricted by the triangle inequality constraints shown as dotted lines. In all three panels, the proximal is at $(0,0,0)$ and the distal is at $(1,1,1)$. In the first panel, $m_y = 4$, $m_d = 4$, in the second $m_y = 5$, $m_d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

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

Now treating $s = p_y + p_d$ as fixed, we compute the range of the difference $d = p_d - p_y$ as $(q_{d\min}, q_{d\max})$, and define q_d as

$$q_d = (d - q_{d\min}) / (q_{d\max} - q_{d\min}). \quad (2.7)$$

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_y, q_d) within the unit square may be mapped to a (p_y, p_d) for any (m_y, m_d, m_{pd}) combination.

Importantly, the predicted configurations for a given (q_y, q_d) for different (m_y, 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_y, m_d, m_{pd}) . This feature is important if we are to use (q_y, q_d) as responses that can be sensibly averaged.

The values of $q_{y\min}, q_{y\max}, q_{d\min}, q_{d\max}$ for given (m_y, 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_y, 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. We may compute the range of this stretch parameter and then compute an appropriate $q \in [0, 1]$ as above.

In describing a pivot angle, two problems may occur. First, when $m_y + m_d = m_{pd}$, that is, 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. If the circle lies in the horizontal plane, however, 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 joints, it is possible to pick a good polar direction since we 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.

Our approach 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. We then model the three coordinates of the normal vector. 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.

3. 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 we 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.

3.1 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 farther 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

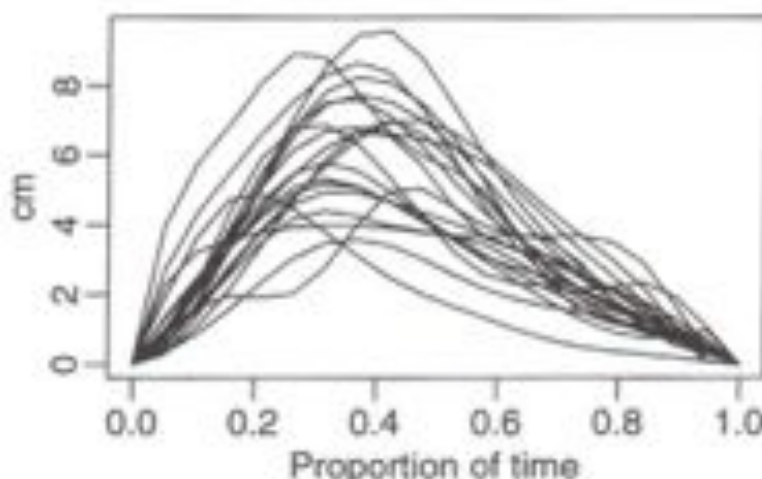


Figure 3. The radial deviation of the hand from a straight line path when reaching with the right hand to a location on the left for 20 subjects.

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. Because 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 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 $\sum_{j=1}^m \beta_{ij} \psi_j(t)$, where the coefficients β_{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 our 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 Faraway (2000).

3.2 TRAJECTORY REPRESENTATION

In addition to predicting univariate functions, such as the radial deviation, we need to predict the 3D curves formed by the trajectory of the hand or other body joints. We 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 we may well wish to make. We have chosen a different parameterization that has more interpretable components and assumes that we have specified the beginning and end of the trajectory (as would be the case for predicting the movement of the hand for specified tasks).

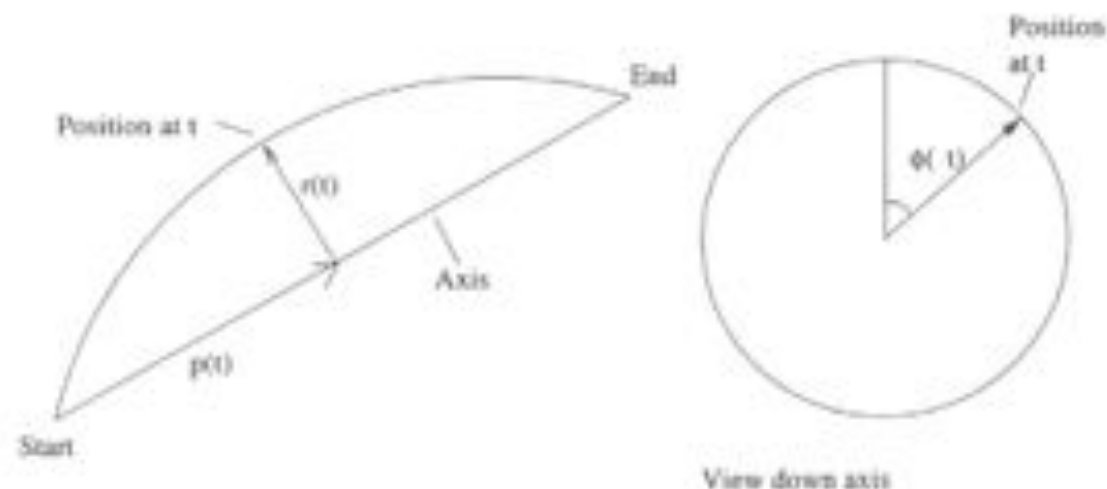


Figure 5. On the left, we show a side view of the reach with the axis drawn as a straight line connecting the start and end of the reach. On the right, the view is down the axis, looking from the start towards the end.

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.

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.

We model the trajectory using the triplet

$$\{v(t), r(t), \phi(t)\}. \quad (3.1)$$

Because $v(0) = v(1) = r(0) = r(1) = 0$ by definition, we can accommodate this by omitting the first and last cubic B-spline basis function in the representation. This omission will ensure the desired property because these functions are, respectively, the only nonzero basis functions at $t = 0$ and $t = 1$. Furthermore, we should not directly model $\phi(t)$ because it is an angle. Angles $2\pi - \epsilon$ and ϵ are only 2ϵ apart but averaging their numerical values produces π which is diametrically opposite. For this reason, we model 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 we predict the hand motion we expect that the endpoints will be specified by the requirements of the application. We 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 Faraway (2001).

So we are 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, we can construct a motion for any choice of link lengths long enough to reach the target.

3.3 MOTION WARPING OR RETARGETING

Suppose we observe a motion by one individual moving or reaching to some target and want to use it to predict the motion for another individual. We describe the observed motion by

$$\{\mathbf{e}, \mathbf{v}(t), \mathbf{r}(t), \phi(t), \boldsymbol{\theta}(t), \mathbf{q}(t), \boldsymbol{\alpha}(t)\}, \quad (3.2)$$

where \mathbf{e} represents the endpoint information such as where the hand should begin and end; \mathbf{v} , \mathbf{r} , and ϕ represent the trajectory information; $\boldsymbol{\theta}$, \mathbf{q} represent the stretch pivot coordinates; and $\boldsymbol{\alpha}$ 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, we may wish to model the trajectories of both hands independently. We 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 a set of segment lengths l or $l(t)$ since some segment lengths (such as the torso) might be allowed to vary in time.

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

3.4 NEAREST NEIGHBOR ANIMATION

There is no reason to restrict ourselves to using a single motion for constructing the prediction. Suppose we have motion data with various endpoints, segment lengths, and other characteristics which are similar to the case we wish to predict. We may average the stretch pivot and trajectory parameters describing these motions to construct our 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. We demonstrate this in our application below. We can consider more than just simple local averages—many of the well-known ideas of nonparametric regression can be employed. See Simonoff (1996) 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, we 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, we may evaluate any candidate scheme with respect to its predictive ability using cross-validation. I describe in the following how this was done in practice.

4. APPLICATION

This section describes the data to which the methods were applied. We emphasize that the methodology presented here could be applied to data collected in different ways for different linkages.

In 1999, the Human Motion Simulation Laboratory (HUMOSEM) 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 3,000 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 trignon (center of the head), and the nasion (between the eyes).

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

5. IMPLEMENTATION

This section describes the particular implementation we 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 our choices here to understand the considerations involved.

We identified a chain stretching from the left foot to the right hand and containing 10

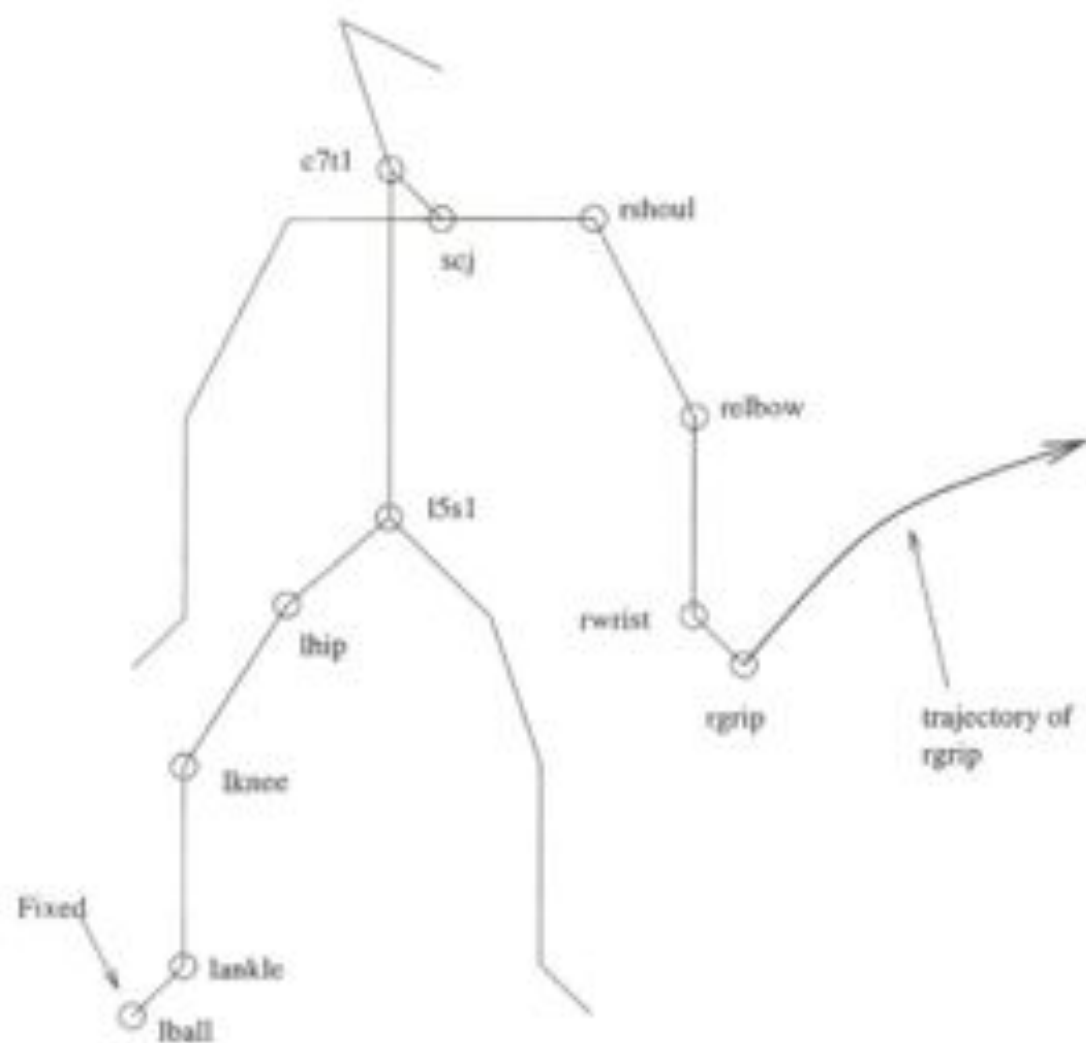


Figure 7. Schematic of human figure. The critical kinematic chain linking the endpoints at the left ball of foot which, in our example, does not move and the right grip, which moves along a trajectory that we have already predicted. We need to predict the locations of the nine interior joints of the 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.

links. Other linkages branch off this chain. In our example, these would be the right leg, left arm, and the head. I will model the main linkage 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 joints on this chain required only a pivot angle to describe. Other choices for the pivots were considered, but this particular choice produced the most stable results.

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

requiring 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 joints). Various studies have been made concerning these lengths and how they vary as a function of anthropometry. In our case, we fit simple regression models to describe these lengths as a linear function of stature. For our 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 cross-validation. 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. We used link lengths predicted using the height of the actual subject and not their actual link lengths. This choice makes the fit somewhat worse, but reflects what the practical user would normally do since height is easy to measure or specify while the link lengths are not.

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 cross-validated 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 $k - 1$, that is, 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, five neighbors gave the best results. It is 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 these data, for the optimal scheme and number of neighbors, the median of the average difference between the predicted and observed motions over the joints 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 joints 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, we cannot 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.

6. 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 articles provide a review of the current state of research in this type of problem. Tolani, Goswami, and Badler (2000) presented a parameterization of a popular seven degrees of freedom model for the shoulder, elbow, and wrist with the hand constrained to a target as in our 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. Tolani, Goswami, and Badler (2000) described the competing methods and showed 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. I developed this approach after failing to find a workable optimization-based inverse kinematics solution to a full body linkage.

Gleicher (2001) addressed 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.

Our representation has some weaknesses. For example, most joints have mobility re-

restrictions but our method does not explicitly respect these restrictions. Because 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 our method may serve as an initial estimate requiring some correction. For example, our standing figure occasionally has a predicted ankle position that would put the heel through the floor. We correct this 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. We find it substantially faster and simpler to make local corrections to the posture if they are needed.

Furthermore, more complex constraints may be needed. In our 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 has not been a problem in practice, we 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.

7. DISCUSSION

Most of the weaknesses in the methodology described here 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 we specify inputs that are far from those observed in the experiment. For example, if we input a target far to the left of the subject, we cannot expect a good prediction because all the targets in the experiment were to 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 obstacles 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, one reaches to targets near the feet, we need avoid predictions where the hand passes through the knees.

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.

APPENDIX

Range of Sum and Difference of Stretch Parameters

$q_{s,max}$:

Set $q_{s,max} = 2$, but

- If $m_d - m_p \geq m_{pd}$, set $q_{s,max} = (m_{pd} + m_p + m_d)/m_d$.
- If $m_p - m_d \geq m_{pd}$, set $q_{s,max} = (m_{pd} + m_p + m_d)/m_p$.

$q_{s,min}$:

- 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 $q_{s,min} = \min(a, b)$.

Let $x = p_p + p_d$.

$q_{d,max}$:

- 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 $q_{d,max} = \min(x, 2 - x, a, b)$.

$q_{d,min}$:

- 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 $q_{d,min} = \min(-x, x - 2, a, b)$.

ACKNOWLEDGMENTS

The author thanks the HableSim project and its industrial sponsors at the University of Michigan and Don Chaffin in particular. The author also thanks the Department of Mathematical Sciences at the University of Bath which hosted the author during the period in which this article was written.

[Received January 2002. Revised July 2003.]

REFERENCES

- Farway, J. (2000). "Modeling Reach Motions Using Functional Regression Analysis," SAE Technical Paper 2000-01-2175.
- (2001). "Modeling Hand Trajectories During Reaching Motions," University of Michigan, Department of Statistics, Technical Report No. 383.
- Gleicher, M. (2001). "Comparing Constraint-Based Motion Editing Methods," *Graphical Models*, 63, 107-134.
- Gleicher, M., and Litwinowicz, P. (1999). "Constraint-Based Motion Adaptation," *The Journal of Visualization and Computer Animation*, 9, 85-94.
- Hsiang, S. and Ayoub, M. (1995). "Development of Methodology in Biomechanical Simulation of Manual Lifting," *International Journal of Industrial Ergonomics*, 17, 271-288.
- Kovin, J. (1985). *A Geometric Investigation of Reach*. Cambridge, MA: MIT.
- Lee, J., and Shin, S. Y. (1999). "A Hierarchical Approach to Interactive Motion Editing for Human-Like Figures," in *ACM SIGGRAPH 99, Annual Conference Series*, pp. 39-48.
- Smootoff, J. (1996). *Smoother Methods*. New York: Springer-Verlag.
- Tokui, D., Gossami, A., and Badler, N. (2000). "Real-Time Inverse Kinematic Techniques for Anthropomorphic Limbs," *Graphical Models*, 62, 353-368.
- Wang, X. (1999a). "A Behavior-Based Inverse Kinematics Algorithm to Predict Arm Pushion Postures for Computer-Aided Ergonomic Evaluation," *Journal of Biomechanics*, 32, 453-460.
- (1999b). "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.
- Wilkin, A., and Popovic, Z. (1995). "Motion Warping," in *Computer Graphics (Proc. SIGGRAPH '95)*.
- Zhang, X., Kao, A., and Chaffin, D. (1999). "An Optimization-Based Differential Inverse Kinematics Approach for Modeling Moderately Complex Three-Dimensional Dynamic Seated Postures," *Journal of Biomechanics*, 31, 1035-1042.