Laziness is a virtue: Motion stitching using effort minimization

Lei Li, James McCann, Christos Faloutsos and Nancy Pollard

Carnegie Mellon University

Abstract

Given two motion-capture sequences that are to be stitched together, how can we assess the goodness of the stitching? The straightforward solution, Euclidean distance, permits counter-intuitive results because it ignores the effort required to actually make the stitch. The main contribution of our work is that we propose an intuitive, first-principles approach, by computing the effort that is needed to do the transition (laziness-effort, or 'L-score'). Our conjecture is that, the smaller the effort, the more natural the transition will seem to humans. Moreover, we propose the elastic L-score which allows for elongated stitching, to make a transition as natural as possible. We present preliminary experiments on both artificial and real motions which show that our L-score approach indeed agrees with human intuition, it chooses good stitching points, and generates natural transition paths.

Categories and Subject Descriptors (according to ACM CCS): I.3.7 [Computer Graphics]: Three Dimensional Graphics and Realism-Animation

1. Introduction

A good distance function is important for the generation of realistic character motion from motion capture databases. In this paper, we propose a novel distance function to pick natural stitching points between human motions. To motivate our work, we demonstrate that a a straightforward, ad-hoc approach may lead to poor stitchings. For example, Figure 1 shows a problem case for the often-used *windowed Euclidean distance* [WB03]. Other ad-hoc metrics like time-warping and geodesic joint-angle distance [WB04] may suffer from similar issues, because none of them tries to capture the dynamics of the stitching as explicitly as our upcoming proposal does.

How do we capture the "naturalness" of a stitching? Our approach is to go to first principles, informally expressed in our following conjecture:

Conjecture 1 (Laziness) Between two similar trajectories, the one that looks more "natural" is the one that implies less effort/work.

The rationale behind our conjecture is that humans and animals tend to minimize the work they spend during their motions, as captured in the "minimum jerk" [FH85] and "minimum torque change" [UKS89] hypotheses of motion, for example. Formally, we are focusing on the following problem (Figure 2):

Problem 1 (Stitching Naturalness) Given a query sequence $\mathcal{Q}$ of N points in m-dimensional space with take-off point $\mathbf{q}_a$, and a data sequence $\mathcal{X}$ of M points of the same dimensionality with landing point $\mathbf{x}_b$, **find** a function to assess the goodness of the resulting stitched sequence, i.e. $\mathbf{q}_1, \ldots, \mathbf{q}_a, \mathbf{x}_b, \ldots, \mathbf{x}_M$.

The goal is that the "goodness" metric should be low if humans consider the stitching to be natural. Once we obtain a qualified distance function, we can either do a sequential scan or use database indexing techniques to perform a fast search over the whole motion capture database to find the best stitching motions [LCR*02].

2. Proposed Method

Here, we describe our L-score for motion stitching. The idea is to exploit Conjecture 1, that humans tend to use as little work as possible and thus natural human motion transitions should be work-efficient. Our L-score relies on a fast, easy to compute estimate of the effort required to make a stitch. To

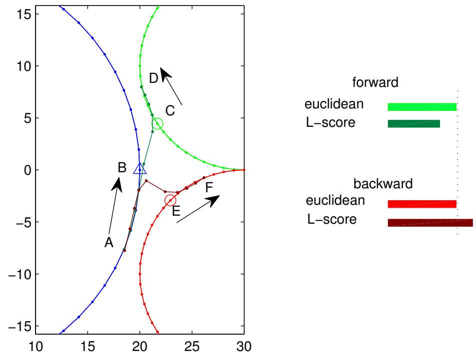

Figure 1: *Motivating example: The stitching from (AB)-to-(CD) ("forward") seems more natural than the stitching (AB)-to-(EF) ("backward"). The right part shows the corresponding "stitch-ability" scores. However, the Euclidean distance does not capture the awkwardness of the actual stitching and assigns the same cost (about 47) to both. The "forward" stitching (AB)-to-(CD) has a smoother, more natural-looking trajectory(darker lines).*

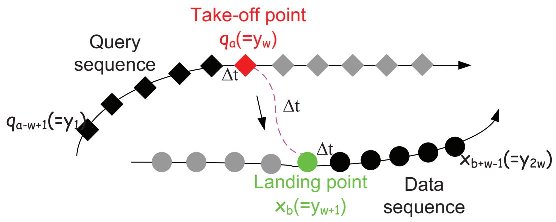

Figure 2: *Illustration of Problem 1. The query trajectory (with diamonds) is to be stitched with the data trajectory (with circles) at the two indicated points: the red diamond indicates the take-off point q_a, and the green circle marks the landing point x_b. Grayed out points indicate points that we ignore in our stitching.*

create the stitch, any on-the-shelf regression/fitting method (e.g. linear interpolation or spline) could be plugged into our L-score method in principle. However, these methods need manual tuning(e.g. order of spline). We recommend the Kalman filter to estimate the motion dynamics for the following reasons (a) it has explicit (Newtonian) dynamic equations consistent with first principles (b) it could reduce noise as well. Kalman filters have already been applied to human motion data for retargeting [TK05] and computer puppetry [SLSG01].

2.1. Estimation of Dynamics

Given the query sequence $\mathcal{Q}$ and the data sequence $\mathcal{X}$ in m dimensions, we create a new stitching sequence (within a certain window size w) $\mathcal{Y} = \mathbf{y}_1, \ldots, \mathbf{y}_{2w} = \mathbf{q}_{a-w+1}, \ldots, \mathbf{q}_a, \mathbf{x}_b, \ldots, \mathbf{x}_{b+w-1}$ (Figure 2). To estimate the trajectory in the stitching process, we try to find the hidden dynamics (the true position, velocity, and acceleration) at each time tick, while eliminating the observation noise. Given the observed position at every time tick, we build the following Kalman filter(Eq. 1) for each dimension of the stitching sequence. For the following, we assume the data sequence is one dimensional.

$$\begin{aligned} \mathbf{z}_1 &= \mu_0 + \omega_0 \\ \mathbf{z}_{n+1} &= \mathbf{A}\mathbf{z}_n + \omega_n, \\ y_n &= \mathbf{C}\mathbf{z}_n + \varepsilon_n \end{aligned} \qquad \mathbf{A} = \begin{pmatrix} 1 & \Delta t & \Delta t^2/2 \\ 0 & 1 & \Delta t \\ 0 & 0 & 1 \end{pmatrix} \tag{1}$$

where the hidden states consist of true position p_n, velocity v_n, acceleration a_n: $\mathbf{z}_n = (p_n, v_n, a_n)^T$. The transition matrix $\mathbf{A}$ is determined from Newtonian mechanics of a point mass, and the transmission matrix $\mathbf{C} = (1 \quad 0 \quad 0)$, with Gaussian noise terms $\omega_t \sim \mathcal{N}(0, \mathrm{diag}(\gamma_1, \gamma_2, \gamma_3))$ and $\varepsilon_t \sim \mathcal{N}(0, \sigma)$. We set the prior parameter $\mu_0 = (p_0, v_0, a_0)^T = (y_1, (y_2 - y_1)/\Delta t, (y_3 + y_1 - 2y_2)/\Delta t^2)^T$.

We use the forward-backward algorithm [GW01] to achieve our goal: to estimate the expected value of the hidden states $\hat{\mathbf{z}}_n = \mathrm{E}[\mathbf{z}_n \mid \mathcal{Y}](n = 1, \ldots, 2w)$. For motion stitching data in m dimensional space, we build the Kalman estimation for each dimension and estimate position, velocity and acceleration separately.

2.2. L-score

Now that the velocities and accelerations have been calculated, the next step is to calculate the effort during the stitching.

Given the estimated hidden states $\hat{\mathbf{z}}_n = (\hat{p}_n, \hat{v}_n, \hat{a}_n)^T (n = 1, \ldots, 2w)$, we can compute the energy spent as the product of force and displacement. Thus, we define the following L-score = $L(\mathcal{Q}, a, \mathcal{X}, b, w)$ for motion stitching:

$$L(\mathcal{Q}, a, \mathcal{X}, b, w) = \sum_{n=1}^{2w-1} |(\hat{p}_{n+1} - \hat{p}_n) \cdot \hat{a}_n| \tag{2}$$

2.3. Generalization: Elastic L-score

To remedy the abrupt transition in above method, we generalize the problem by allowing the injection of some intermediate frames between take-off and landing. To this end, we extend the above method and propose our *elastic* L-score (Figure 3).

For a given k, we again use the above Kalman filter to estimate the dynamics also for k injected frames(denoted as $z'_i = (p'_i, v'_i, a'_i)^T, i = 1 \ldots k$). To estimate the transition effort, we not only compute the work for real frames, but also compute the effort for injected frames in the same way as in Eq 2. We define the *elastic* L-score $L_*()$ for the optimal number of

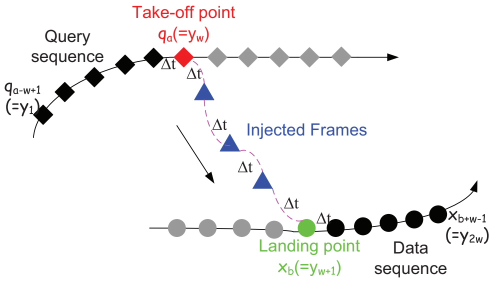

Figure 3: *Illustration of "take-off", "injected" and "landing" points. The trajectory of squares is to be stitched with the trajectory of circles at the two indicated points (red square for take-off, green circle for landing); the injected frames, shown as blue triangles, which are to be estimated. Grayed out points indicate points ignored in our stitching.*

injections k_{opt} that minimizes the L-score $L_k(\mathcal{Q},a,\mathcal{X},b,w)$ for fixed number of injections k.

$$\begin{aligned} L_k(\mathcal{Q},a,\mathcal{X},b,w) = & \sum_{n=1}^{w-1} |(\hat{p}_{n+1}-\hat{p}_n)\cdot\hat{a}_n| \\ & +|(\hat{p'}_1-\hat{p}_w)\cdot\hat{a}_w| + \sum_{i=1}^{k-1} |(\hat{p'}_{i+1}-\hat{p'}_i)\cdot\hat{a'}_i| \qquad (3) \\ & +|(\hat{p}_{w+1}-\hat{p'}_k)\cdot\hat{a'}_k| + \sum_{n=w+1}^{2w-1} |(\hat{p}_{n+1}-\hat{p}_n)\cdot\hat{a}_n| \end{aligned}$$

$$L_*(\mathcal{Q},a,\mathcal{X},b,w) = \min_{k\geq 0} \quad L_k(\mathcal{Q},a,\mathcal{X},b,w) \qquad (4)$$

The *elastic* L-score $L_*()$ not only gives an assessment of the stitching quality, but it also chooses the most suitable number k_{opt} of frames to inject - its goal is always to minimize the transition effort. Furthermore, once we decide the number k_{opt} of injected frames, we get a good transition trajectory for free: $\hat{p}_1,\ldots,\hat{p}_w,\hat{p'}_1,\ldots\hat{p'}_{k_{opt}},\hat{p}_{w+1},\ldots,\hat{p}_{2w}$.

3. Experiments

We have already illustrated (Figure 1) that the Euclidean distance may lead to counter-intuitive results. Next we present experiments with the *elastic* $L-score$ ($L_*()$) on (a) synthetic and (b) real motion capture data.

Synthetic Data We generated the *Three-Circles* dataset with a frame rate of $64/cycle$ in 2-dimensional space (m=2), as shown in Figure 1. The large, left circle has a radius of 20 units and is centered at $(0,0)$; the right circles both have radius 10 units and are centered symmetrically at $(30,10)$ and $(30,-10)$. We perform two experiments: the "forward" and the "backward" transition(Figure 1). In these experiments, we identify both the optimal landing point and the optimal number of injected frames. Figure 4 shows the results: the *elastic* L-score favors the forward transition, which agrees with human intuition. It also chooses a larger number of frames and a later landing point to ameliorate the effects of the awkward backward transition.

Real Human Motion: We capture a set of waving, walking, running and jumping motions at 30 frames per second. Motions are 300 to 2000 frames in length and have m=93 dimensional joint positions in body local coordinates. We use one Kalman filter for each of the m=93 features as described in Section 2, and set the parameters to be $\Delta t = 1$, $\gamma_1 = \gamma_2 = \gamma_3 = \sigma = 0.001$. We use the window of $2w = 10$. We have informally viewed a large variety of transitions within this database and find that our approach consistently performs as well or better than the Euclidean distance metric at generating pleasing transitions.

In order to assess the quality of the stitching found by our *elastic* L-score, we blank out a short interval (2 frames) and a long interval (11 frames) from the transition made by the human actor during 2 waving circle motions, and we compare the actual trajectory against the transition trajectories estimated by the *elastic* L-score. The processing time is around two and a half hours on a Pentium class machine. The observations (see Figure 5) are as follows:

- Our method computes the correct value of blanked-out frames, or gets very close to it.
- Our generated trajectories match very well the actual trajectories (please see the accompanying video).

4. Conclusions

The main contribution of this paper is the design of a new distance function for motion stitching, $L-score$, based on first principles. Motivated by the weaknesses of Euclidean distance (Figure 1), we wished to more accurately capture the perceived "naturalness" of a trajectory. This led to our Conjecture 1, stating that the most natural-looking motion trajectory is the laziest-looking one, that is, the one that requires the least effort. The specific contributions of this paper are the following:

- We show how to compute two dynamics-aware distance functions, the L-score and the *elastic* L-score. Among the many possible choices, we recommend the Kalman filter with Newtonian particle dynamics to estimate velocities and accelerations required to compute the L-score.
- Our technique allows for *elastic* stitching, where we automatically compute the optimal *duration* of a transition. Optimality, again, is judged by the total required effort.
- In experiments on both artificial and real motions, the dynamics-aware distance function chooses good stitching points and produces natural-looking trajectories.

Although our algorithm works well as is, and is simple to implement, it uses a rough (particle-based) approximation of character dynamics for state estimation. Improving this approximation, perhaps by using full-body dynamics and a nonlinear filter, is an interesting direction for future work.

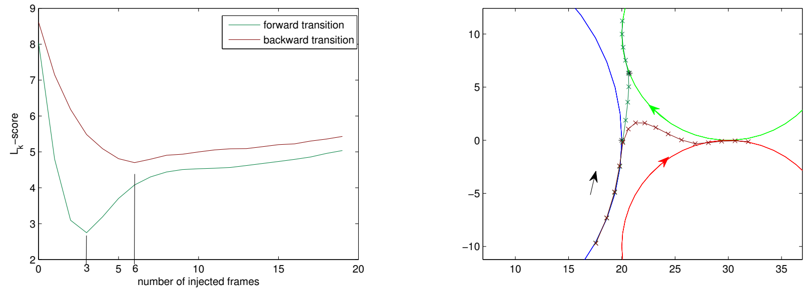

Figure 4: *Left shows the elastic L-score versus k (number of injected frames). The starting ("query") motion is the same, 'AB' (as in Figure 1), and the data motions correspond to the "forward" and the "backward" cases with both landing on optimal positions. Right shows the generated paths for the corresponding k_{opt} optimal number of injected frames. Notice the asymmetric landing positions and that the forward transition has lower elastic L-score, as well as it needs fewer injected frames ($k = 3$, vs $k = 6$), agreeing with human intuition.*

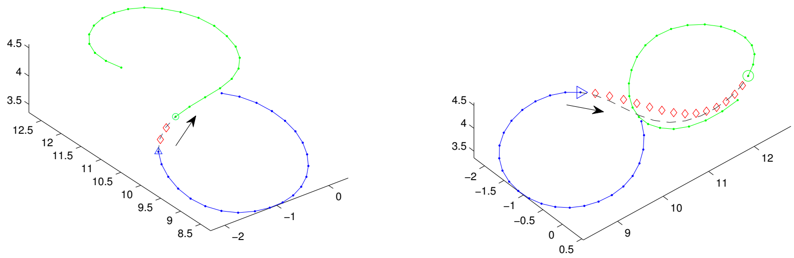

Figure 5: *Real motion stitching: Right-hand coordinates of a human transition motion, with the dashed part blanked out (2 blank-out frames for the left figure, 11 for the right). △/○ marks the take-off/landing frame, respectively. Red ◇ stand for our reconstructed path using* elastic *L-score; notice how close they are to the ground truth (gray dashed line). The* elastic *L-score either finds the correct k_{opt} (=2 in left) or gets very close (=14, vs 11, in right)*

Acknowledgement

The authors would like to thank Justin Macey for his assistance in the motion capture. This work was supported by NSF grants IIS-0326322 and ECS-0325383. The data used in this project was obtained from mocap.cs.cmu.edu supported by NSF EIA-0196217. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation, or other funding parties.

References

[FH85] FLASH T., HOGAN N.: The coordination of arm movements: an experimentally confirmed mathematical model. *J Neurosci 5*, 7 (July 1985), 1688–1703.

[GW01] GREGORY WELCH G. B.: An introduction to the kalman filter, 2001.

[LCR*02] LEE J., CHAI J., REITSMA P. S. A., HODGINS J. K., POLLARD N. S.: Interactive control of avatars animated with human motion data. In *SIGGRAPH '02* (2002), pp. 491–500.

[SLSG01] SHIN H. J., LEE J., SHIN S. Y., GLEICHER M.: Computer puppetry: An importance-based approach. *ACM Trans. Graph. 20*, 2 (2001), 67–94.

[TK05] TAK S., KO H.-S.: A physically-based motion retargeting filter. *ACM Trans. Graph. 24*, 1 (2005), 98–117.

[UKS89] UNO Y., KAWATO M., SUZUKI R.: Formation and control of optimal trajectory in human multijoint arm movement. *Biological Cybernetics 61*, 2 (June 1989), 89–101.

[WB03] WANG J., BODENHEIMER B.: An evaluation of a cost metric for selecting transitions between motion segments. In *SCA '03* (2003), pp. 232–238.

[WB04] WANG J., BODENHEIMER B.: Computing the duration of motion transitions: an empirical approach. In *SCA '04* (2004), pp. 335–344.