Computing the Duration of Motion Transitions: An Empirical Approach

Jing Wang[†] and Bobby Bodenheimer[‡]

Department of Computer Science and Electrical Engineering
Vanderbilt University

Abstract

This paper develops methods for determining a visually appealing length for a motion transition, i.e., a segue between two sequences of character animation. Motion transitions are an important component in generating compelling animation streams in virtual environments and computer games. For reasons of efficiency and speed, linear interpolation is often used as the transition method, where the motion is blended between specified start and end frames. The blend length of a transition using this technique is critical to the visual appearance of the motion. Two methods for determining an optimal blend length for such transitions are presented. These methods are suited to different types of motion. They are empirically evaluated through user studies. For the motions tested, we find (1) that visually pleasing transitions can be generated using our optimal blend lengths without further tuning of the blending parameters; and (2), that users prefer these methods over a generic fixed-length blend.

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

1. Introduction

Motion capture is a popular way of obtaining realistic motions for games and films. New motions can be generated by concatenating existing motion clips in the motion library rather than capturing them. After determining a good transition point between motion clips, generating a seamless transition is critical to the visual appearance of the resulting motion.

Blending is one solution for creating smooth transitions among motion segments. However, generating high quality transitions using blending is still difficult and involves significant manual labor. An animator often needs to go back and forth to modify parameters for blending to obtain a pleasing transition. Some automatic systems simply pre-specify a fixed blend length for all motions.

An appropriate blend length is crucial for many transitions. Bad blend lengths can affect the timing between motions and introduce hops or other artifacts into the animation stream. Velocity mismatches may not be adequately smoothed by a short blend length. Target motions may be diluted by a longer blend length. In short, the blend interval is an important aspect of a good transition.

Our goal is to develop ways to produce visually compelling transitions without manual intervention. In this paper, we develop two methods for determining the blend length for a given transition. We also present the results of an empirical evaluation of these methods. We show that linear blending can create seamless transitions if a good blend length is used. A further goal is to have methods that are computationally inexpensive and that could be incorporated into interactive and real-time systems such as video games.

Section 2 of the paper places our work in context. Section 3 describes the methods we developed to determine an optimal blend length. Section 4 describes the empirical evaluation of the motions produced by using the methods and Section 5 presents the results of this evaluation. Section 6 discusses the results of the work.

[†] email: Jing.Wang@vanderbilt.edu
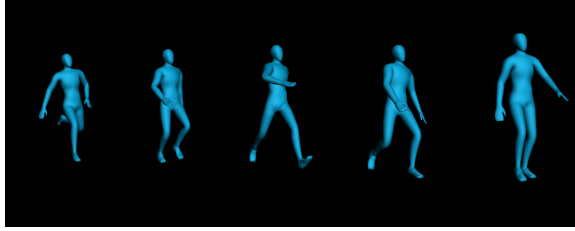[‡] email: bobbyb@vuse.vanderbilt.edu

Figure 1: *An example of a motion transition used in our study.*

2. Work in Context

The greatest drawback of motion capture data is the lack of flexibility. To conquer this problem, researchers have drawn inspiration from the work of Schödl et al. [SSSE00] on video textures to retain the original motion sequences but play them back in non-repetitive streams, creating new sequences [SBS02, AF02, KGP02, LCR*02]. Transitions are an essential component of these systems, but the emphasis of the above research was on selecting appropriate transition points rather than the durations of transitions. While it is true that transitions are less of a problem if the motions are similar, visual artifacts can still appear if the duration is too short or too long.

Blending is a basic way to create transitions. If incorrectly applied, simple blending can produce undesirable results for cases where the motions are not properly aligned, as noted by Perlin [Per95]. In contrast, Rose et al. [RGBC96] used dynamic simulation to generate transitions, using a combination of spacetime constraints and inverse kinematic constraints to create dynamically plausible transitions. Kovar et al. [KG03] introduced registration curves that automatically determined relationships involving the timing, local coordinate frame, and constraints of the input motions; they present a blending algorithm that exploits the information contained in the registration curves. Their work, like that of Bruderlin and Williams [BW95], uses dynamic programming and timewarping to create transitions. Rose et al. [RCB98] also use timewarping and linear interpolation with an ease-in ease-out component to blend motions. Park et al. [PSS02] employed a similar method. Several researchers [WP95, GSKJ03, LCR*02] use displacement mapping techniques to blend motions. Mizuguchi et al. [MBC01] designed a framework that allows animators to define transitions for use in interactive environments. They investigated simple transition techniques and had animators pre-specify the parameters for each of the transitions.

In addition to the variety of techniques being used to generate transitions, there is no generally accepted standard for generating or specifying a transition. Several researchers [RGBC96, RCB98, MBC01, KGP02] specify transitions using a start and end frame and blend be-

tween these marker points. Others [AF02, AFO03, KG03] use a center-aligned transition specification, i.e., the transition frames are the midpoint in the blending interval. Lee et al. [LCR*02] use a "left-aligned" method. Pullen and Bregler [PB02] and Park et al. [PSS02] have still different methods. In particular, Pullen and Bregler join motions directly, but then blend the motion with a smooth quadratic fitted to the curves. There is no direct blending of the two motions.

There are advantages and disadvantages to each of these methods. Start and end frames have the advantage that they are intuitive and easy to specify. They also work well if the transition points are at the end or beginning of motion segments. Their disadvantage is that they can change the alignment of the motions as they are changed. Center-aligned transitions have fixed alignment, which is both an advantage and a disadvantage. If the center-aligned poses are quite similar, then a center-aligned transition is more robust to variations in the blend length. On the other hand, if the poses are mismatched, then no amount of blending will make the transition look good. Center-aligned transitions also have the disadvantage that depending on the blend length there is a region at the beginning and end of each motion segment for which a true blended transition cannot be made.

In this work, we chose to specify transitions with start and end frames. Center-aligned transitions rely at present too heavily upon transition metrics. State of the art transition metrics in the literature, e.g., [LCR*02, KGP02], do not always choose points which represent visually convincing transitions. Kovar et al. [KGP02] use a threshold for generating transitions and any transition below this threshold is a candidate. They themselves note that transition inaccuracies may occur. Lee et al. [LCR*02] use a Markov process to select a transition, thus there is some probability that a bad transition will be selected *(see example in supporting video, where a "bad" transition with probability 0.2 is shown)*. Arikan and Forsyth [AF02] do a randomized search that may yield a bad transition. As noted in Wang and Bodenheimer [WB03], different motions behave differently under these transitions metrics, tuning is required, and there are no guarantees that a optimal transition selected by a method is visually appealing. Therefore, changing the transition points by changing the alignment, if it can be done in a computationally efficient way, represents a second-pass process that can improve the visual appeal of a transition.

Aside from the particular method of specifying a transition, the blend length or duration of the transition is a critical component in the visual fidelity of a spliced animation stream. Rose et al. [RGBC96] found that transition durations of 0.3s to 0.8s worked well, but left the exact specification of the duration to the operator. Lee et al. [LCR*02] found a transition duration of 1 to 2 seconds worked well, but again allowed the operator to select it based on the particular motions. Arikan and Forsyth [AF02] used a constant blend duration of 2 seconds. Pullen and Bre-

gler [PB02] found a duration of 0.2s to 0.8s to work well, based on user selection and depending on the particular motions. Wang and Bodenheimer [WB03] recognized the problem of transition duration but considered it a confounding factor for their experiments and simply concatenated motion segments. Mizuguchi et al. [MBC01] were explicitly concerned with the blend length for transitions, but used an *ad hoc* method of determining them. In their experience, 10 frames (0.33s) worked for a wide variety of motions. Kovar et al. [KGP02] also used this transition duration.

None of the prior work attempted to compute an optimal duration for their particular method of transition generation. When dynamics are known, transition durations can be computed based on the dynamics of the motions, e.g., Popović et al. [PSE*00] and Grassia [Gra00]. Unfortunately, this information is not available for most motion capture libraries and is difficult to estimate. To our knowledge, the present work is the first that is explicitly concerned with determining the optimal blend length of a transition.

Additionally, we validate our results empirically. Empirical evaluation has recently gained popularity in the graphics community, although there is considerable work on point light experiments in the psychophysics literature, e.g., [Joh73, SI87, Md98, PPBS01]. The relevance of these experiments is that users can discriminate between subtle effects given a coarse representation of human motion and absent other visual cues. Hodgins et al. [HOT98] showed that discrimination of motion artifacts depended on rendering style, with a full-body rendering being better than a stick figure. Oesker et al. [OHJ00] assessed the level of detail in animation by studying observer's discrimination of soccer motion. Wang and Bodenheimer [WB03] used empirical methods to optimize the weights for computing transition points between motions. The present work is different from theirs in that we assume the transition points are given and want to compute a blend duration. Reitsma and Pollard [RP03] studied observers' ability to discriminate when errors in ballistic motion were present and presented a metric to measure perceptual errors. O'Sullivan et al. [ODGK03] evaluated the visual quality of physically based simulation in which physical behaviors were degraded or distorted.

3. Transition Method

We chose to build our transition methods on top of linear blending because linear blending is the most common and widely used method for generating transitions between motion segments. However, linear blending violates the laws of physics because it distorts the real motions, being the weighted sum of two or more motions. Linear blending is nonetheless a popular method because it is simple and often generates visually pleasing results.

For two motions, spherical linear interpolation is used to blend between the quaternions of each joint using a linear

weight function. A sigmoidal weight function produces similar results with only subtle differences. The facing direction and the position of the figure on the floor plane are aligned during the blending. We assume in this work that a start frame in a *from* motion and an end frame in a *to* motion are specified. The start and end frames indicate the beginning and end of the blending, respectively.

Linear blending may introduce artifacts such as foot-slide. To fix such problems, inverse kinematics or other techniques [KSG02] are often used as a post-process. These methods may be automatic. We used the inverse kinematic solver provided by MotionBuilder 4.02 to constrain support limbs and correct foot-slide. Other than correcting foot-slide, it rarely affects the visual appearance of the motion.

3.1. Methods for Computing Blend Length

We develop two methods to compute blend length based on two hypotheses on the nature of blending.

3.1.1. Method I: Using the Geodesic Distance

One hypothesis for motion blending is that a transition will be smooth if two windows of the motions to be blended have strong correspondences, which implies that these two pieces of motion have consistent velocities. We compute the best blend length for blending between two arbitrary frames by calculating the cost for blending where the blend length normally ranges from 0.03 to 2 seconds (1 to 60 frames), and pick the blend length with minimum cost.

The cost for transitioning from frame i to frame j with blend length b is computed by averaging the difference of all pairs of corresponding frames within the blend window and is given by

$$D_{f_i f_j} = \sum_{t=1}^{b} \frac{d_{f_i f_j t}}{b}. \tag{1}$$

In this equation, $d_{f_i f_j t}$ is the difference between two corresponding frames given by

$$d_{f_i f_j t} = \sum_{k=1}^{m} w_k \left\| \log \left(q_{j-b+t,k}^{-1} q_{i+t-1,k} \right) \right\|^2 \tag{2}$$

where m is the number of non-global joints in the figure, and $q_{i,k}, q_{j,k}$ are the orientations of joint k at frames i and j, respectively, expressed as quaternions. The log-norm term represents the geodesic norm in quaternion space, and each term is weighted by w_k. The weights were those determined by Wang and Bodenheimer [WB03] in their work. Global degrees of freedom were blended using the method described in Rose et al. [RCB98].

Figure 2 illustrates how the cost as a function of the blend length b is calculated. The cost is the sum of the difference of corresponding frames. Once the costs for a blend length from 0.03 to 2 seconds are computed, the minimum cost can
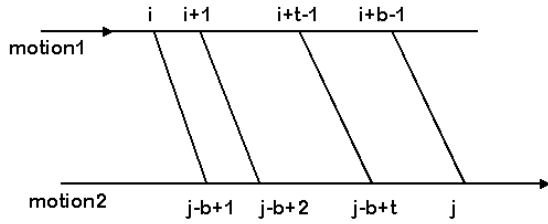
Figure 2: *The cost for blend length b is the average of the difference of corresponding frames. The transition is from frame i to frame j.*
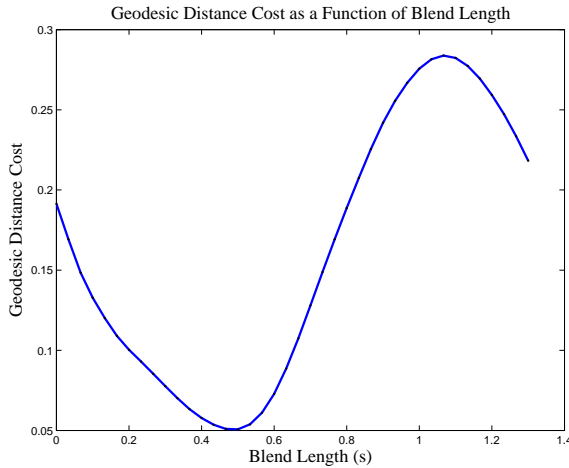

Figure 3: *An example of cost as a function of blend length. The optimal blend length is 0.5s for this example. The transition is from a walking motion to a walking motion.*

be computed; the optimal blend length is given by this minimum. An example of the cost for a walking to walking transition versus different blend lengths is shown in Figure 3. The optimal blend length is 0.5s.

3.1.2. Method II: Using Joint Velocities

Another approach for predicting optimal blend length is inspired by the idea that the rate of change for any joint in the pose should not change radically for a smooth transition. We calculate the joint difference between the start frame and the end frame for each major joint, i.e., shoulders, elbows, hips and knees. We then compute the optimal blend length based on the velocity of the joint that has the maximum difference between the start and end frames.

The difference between frame i and frame j for joint k is

$$d_{f_i f_j}^k = \left\| \log\left(q_{j,k}^{-1} q_{i,k} \right) \right\|. \tag{3}$$

Right Hip	0.0352
Right Knee	0.0481
Right Shoulder	**0.3029**
Right Elbow	0.0167
Left Hip	0.0044
Left Knee	0.0288
Left Shoulder	0.1192
Left Elbow	0.0178

Table 1: *An example of the joint differences between two frames of motions. The maximum difference is on the right shoulder.*

The optimal blend length is then

$$d_{f_i f_j}^p / \|v\| \tag{4}$$

where p is the joint with the maximum joint difference, and v is the average of the joint velocity for the beginning frame and the end frame of joint p, respectively. Table 1 shows an example of the joint differences between two frames of motions. The maximum difference is on the right shoulder.

3.1.3. *Ad Hoc* Comparison

In informal tests of these methods, our experience is that the geodesic distance is more suitable for cyclic locomotion such as walking and running for which correspondences in the blend are critical. One possible reason is that cyclic motion has a fixed pattern and people are sensitive to movements that are out of phase. The geodesic distance method aims at the phase requirement of these motions and finds the best correspondence of frames of motion for blending.

On the other hand, the velocity method is more suitable for physical activities such as boxing and free-style dancing, etc. For motions like these, people do not have strict perceptual predictions for the next move. However, a longer blend length does not necessarily mean a better transition. For example, a rather quick punch by a boxer might become a slow punch after a long blending. Therefore, finding the optimal blend length that produces smoothness and still preserves the quality of the target motion becomes important. The velocity method meets these requirements by smoothing the movement of every joint and does not unnecessarily stretch the resulting motion.

3.1.4. Alternative Methods

We studied alternative techniques for computing a good blend length, more complex and computationally expensive than the previous two. As noted in Section 2, timewarping
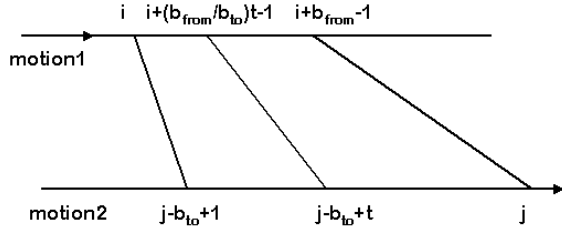
Figure 4: *An illustration showing the method for computing the cost of a blend length with timewarping. The blend length for the* from *motion is b_{from} and the blend length for the* to *motion is b_{to}. In this example, $b_{from} < b_{to}$.*

Figure 5: *The geodesic distance cost for each of the joints of the motion transition shown in Figure 2 (walking to walking). The joints that do* not *have minima at 0.5s are the left shoulder (black), right shoulder (yellow) and right elbow (magenta).*

has been used for generating transitions. Thus, we modified the geodesic distance method to compute a blend length where the *from* and *to* motions can be timewarped. Given a transition from frame *i* to frame *j*, a timewarped blend length is calculated by computing a cost matrix of blend lengths in the *from* motion versus blend lengths in the *to* motion. Each entry in this cost matrix C_{ij} is given by

$$C_{ij}(b_{from}, b_{to}) = \sum_{t=1}^{b_{to}} \sum_{k=1}^{m} w_k \left\| \log \left(q^{-1}_{j-b_{to}+t,k} q_{i+\frac{b_{from}}{b_{to}}t-1,k} \right) \right\|^2$$
(5)

if $b_{from} \le b_{to}$, and

$$C_{ij}(b_{from}, b_{to}) = \sum_{t=1}^{b_{from}} \sum_{k=1}^{m} w_k \left\| \log \left(q^{-1}_{j-b_{to}+\frac{b_{to}}{b_{from}}t,k} q_{i+t-1,k} \right) \right\|^2$$
(6)

if $b_{from} > b_{to}$, where b_{from} and b_{to} are the blend lengths in the *from* and *to* motions, respectively, and the other terms are defined as in Equation 2. Figure 4 illustrates how this cost is computed. The minimal cost from the cost matrix C_{ij} then gives the appropriate motion intervals with which to perform timewarped blending. When a computed frame time is not an integer, joint values of the pose are obtained by spherical linear interpolation between the two adjacent frames. Note that computing the optimal blend length using the geodesic distance method is $O(b)$ whereas the timewarped blend method is $O(b^2)$.

The second alternative we explored is the idea of using a non-uniform blend schedule on the degrees of freedom to produce a transition. We could, for example, transition a shoulder degree of freedom over 10 frames and a hip degree of freedom over 20. There are two drawbacks to this method, both related to the physical properties of the motion. First, as shown in Figure 5, the physical coherence of the individual joints indicates that the optimal blend length as computed by the geodesic distance occurs at the same value for most important degrees of freedom. This figure shows the geodesic cost for individual joints of the motion shown in Figure 2. The joints that do not have minima at 0.5s are the left shoul-
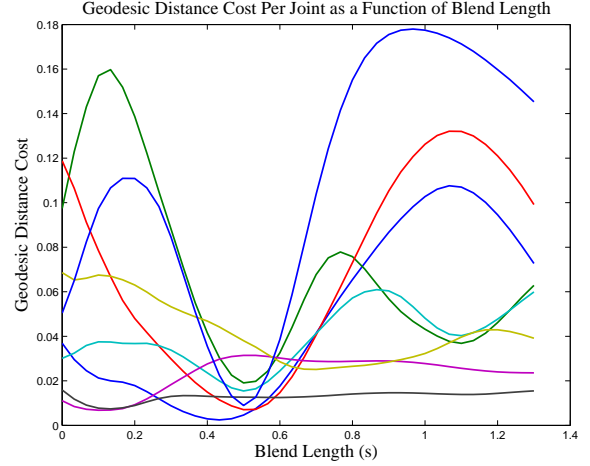
der, right shoulder, and right elbow, although the cost for the left shoulder is nearly constant. The second, more important drawback is that different blending schemes destroy the physical coherence of the degrees of freedom. When trying to adapt the methods to different blend schedules, visual artifacts were apparent.

We additionally tried to modify the velocity method according to the methods implemented by Grassia [Gra00]. We found no improvement from the basic method described above.

4. Experimental Evaluation

There are a number of interesting psychophysical evaluations that could be conducted on the methods described previously. In this study, the methods described above were compared to a typical transition scheme that employed a fixed blend length. The goal of these experiments was to leverage the results of our *ad hoc* comparison (Section 3.1.3) and determine user preferences for the methods over a wide repertoire of motions. Additionally, we conducted experiments to determine the "just noticeable difference" or differential threshold of observers to changes in blend lengths within a neighborhood of the optimal blend length determined by the geodesic distance. If users are largely insensitive to changes in blend length, then the methods used to determine those lengths may be unimportant.

4.1. Procedure

The fixed blend length we chose was 0.33s (10 frames). This value was chosen because it is the value suggested by Mizuguchi et al. [MBC01], used by [KGP02], and in the range specified by both [RGBC96] and [PB02]. Our experience from conducting pilot studies prior to these experiments leads us to believe that the results described here will hold for any fixed blend length.

All experiments were run in a single session consisting of four distinct studies. The participants were volunteers from our institution with no prior animation experience beyond exposure to video games and film. Thirty-five people volunteered: 20 male and 15 female, aged 22 to 40 years. All participants had normal or corrected-to-normal vision. Additionally, participants were naïve as to the purpose of these experiments.

Motion transitions were created from a variety of motion capture data and shown in the same rendering style. Groups of motions were shown from the same camera position. Consistent with the point light experiments mentioned in Section 2, we chose to omit rendering a ground plane. While the ground plane can provide important visual cues for some perceptual studies, e.g., [RP03], we judged it unnecessary for our purposes.

4.2. Study One: Just Noticeable Difference

In this experiment, we studied how noticeable the blend length of a transition was. We asked users to make a series of two-alternative forced choice responses. For a transition between two different motions, the optimal transition length k was computed using the geodesic distance method, and transitions of lengths $k-15$, $k-10$, $k-5$, k, $k+5$, $k+10$, and $k+20$ were generated. These motion transitions were grouped in pairs: one member of the pair was the length k transition and the other member was any of the other lengths (including length k). Each participant was shown each transition pair twice within a set of motions, one where the length k transition was presented first and one in which the length k transition was presented second. The presentation of the transitions was randomized, both pairwise and among pairs (that is, the length of the transition compared to length k was randomized and also whether the length k transition was presented first or second). These transition sets were created for three different types of motions: walking, standing to walking, and running to turning. Thus, users were shown a total of 42 motion pairs ($7 \times 2 \times 3$). Each motion pair consisted of two seconds of entry time before the transition occurred and two seconds of exit time after the transition completed.

Participants were presented with these motions and asked to determine whether the motion pair they were being shown consisted of identical or different motions. They were instructed that sometimes the motion pair would be identical and sometimes different. They were given five seconds between each motion pair to make their determination.

This experiment is of a kind known as the "method of constant stimuli" in psychophysics [GS66]. In this type of experiment, the just noticeable difference is the stimulus difference that is reported correctly 75% of the time; an accuracy rate of 50% represents chance guessing. To determine a very fine differential threshold, it is typical to use a large number of stimuli pairs. Our goal was a simpler and more coarse analysis to determine with what accuracy a blending method should operate. A different experimental design was used to detect *absolute* thresholds in [ODGK03]. Their design was similar to the classical method of limits.

4.3. Study Two: Geodesic distance method versus fixed blend-length

In this experiment, we studied whether participants judged that motions containing a transition generated by the geodesic distance method appeared more natural than motions containing a transition using a fixed blend-length of 10 frames. We selected eight different motion transitions consisting of such motions as standing and idling to walking or running (of different speeds), walking to running (of different speeds), and various turning motions. Transition points were selected randomly to make the study independent of a particular transition metric. As discussed previously, these motion types are those we believed most suitable for the geodesic distance method. None of the optimal blend lengths were close to 10 frames. Optimal blend lengths for the motions tested ranged from 5 frames to 35 frames. Motion pairs were generated, one containing the optimal blend length and one containing the 10 frame blend length. The order of these was randomized.

Participants were presented eight motion pairs and asked to determine whether the first or second motion of a pair was more natural. They were again given five seconds between each motion pair to make their determination.

4.4. Study Three: Velocity method versus fixed blend-length

This experiment was conducted to determine whether participants judged motions containing a transition generated by the velocity method to appear more natural than motions containing a transition generated by a fixed blend-length. The experimental procedures and preparation of stimuli were identical to Study Two above, except that the eight motions chosen for study consisted of boxing, dancing, and tai-chi motions, motions of a type we believed most suitable for the velocity method. Participants were again presented with eight motion pairs.

4.5. Study Four: Geodesic distance method versus timewarping method

Finally, in this experiment we studied whether participants judged that motions containing a transition generated by the geodesic distance method appeared more natural than motions containing a transition generated using the timewarping strategy discussed in Section 3.1.4. The experimental procedures and stimuli were identical to those of Study Two. The timewarped transitions averaged a warp of 10 frames; for example, one motion blended 18 frames in the *from* motion to 28 frames in the *to* motion.

5. Results and Analysis

5.1. Just Noticeable Difference

The aggregate probability of a subject determining correctly whether a pair of motions was different is plotted against difference in blend length and shown in Figure 6. The motions were identical for the zero value on the *x*-axis, which corresponds to a comparison of the length k transition versus the length k transition; otherwise, the motions were different. Also shown are standard errors of the mean for each blend length interval. Notice that the probability of saying that identical motions were different (0 on the *x*-axis) is low, as it should be. The 75% threshold or just noticeable difference occurs at $k + 7$ frames when the transition is made longer or $k - 8$ frames when the transition is made shorter. We discuss this finding further below, but the result we conclude from this study is that people can differentiate between transition lengths that differ by seven or eight frames. There was no statistically significant difference in the performance of the test across motions.

5.2. Comparison of Methods

Table 2 shows the percentage of study participants preferring various methods over the others as tested in Studies Two, Three, and Four. In particular, we see that 96.4% of the participants thought that the geodesic distance looked more natural when compared to a 10-frame blend for the motions studied, 65.7% of the participants favored the velocity method over the 10-frame blend when asked which produced more natural motion, and 55.7% of participants favored the geodesic distance method over the timewarping method.

Also shown in Table 2 is the χ^2 test statistic [Ric95] applied to these studies. An alpha level of .01 was used for all statistical tests (the critical value of χ^2 for this alpha is 6.64). There is one degree of freedom and a sample size of 280 (35×8) for each study.

The observed percentages of users preferring the geodesic method and velocity methods over 10-frame blending is statistically significant. However, it is not clear that users can successfully distinguish between the geodesic method and our timewarping method. This result is supported by the
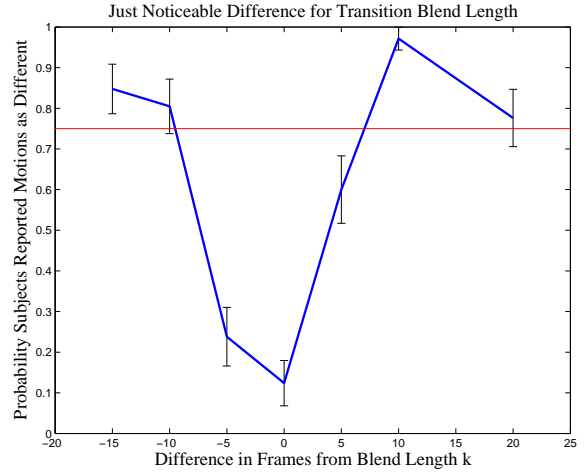

Just Noticeable Difference for Transition Blend Length

Figure 6: *Results of the just noticeable difference study for different blend lengths. The x-axis shows the difference in blend length from the optimal value k computed using the geodesic distance method. The y-axis shows the probability of successfully reporting that the motions were different. The 75% threshhold is also drawn. The error bars indicate standard errors of the mean.*

Method	% favoring	$\chi^2(1, N = 280)$
Geo. dist. over 10-frame	96.4	241.4, $p < .01$
Velocity over 10-frame	65.7	27.7, $p < .01$
Geo. dist. over Timewarping	55.7	3.66

Table 2: *The percentage of users that preferred the indicated methods as described in Studies Two, Three, and Four. The first column gives the percentage favoring the first method listed over the second, and the second column gives the χ^2 test statistic for the experiment.*

comments of many of the participants who noted that the motions seemed very similar. The same results are found when the data is analyzed on a per motion basis. Also, for Studies Two and Three, there were no individual motions for which users preferred the 10-frame blend.

6. Discussion

In this work, we developed two methods for determining the best blend lengths for generating a transition between two motions using linear blending. Visually appealing transitions are critical in the re-use of large motion data-sets, and the transition duration is one of the most important factors in creating a compelling transition. Human motions are highly varied, and developing a universal method for gen-

erating compelling transitions may not be possible. Thus, investigating methods that work for categories of motions seems reasonable.

The first method, which we call the geodesic distance method, determines the best blend lengths for motions that have a cyclic nature, such as running and walking. The second method, which we call the velocity method, determines the best blend lengths for motions that are non-repetitive, activities such as free-form dancing and boxing. These methods automatically generate a blend length for linear blending given two motions and the frames in those motions to transition between. This information is readily available from such systems as [LCR*02, KGP02, AF02]. There is no need for further modification of blend parameters by a user or animator.

The methods and experimental evaluation described in this paper give guidance to designers of animation systems who wish to incorporate varying blend lengths into their system. However, there are a number of other interesting psychophysical experiments that could be conducted on these methods. For example, we could conduct experiments using the geodesic and velocity methods to categorize motions, and determine how much, if any, overlap there is. We could assess the robustness of the methods across many categories of motion, and better compare them to each other. We could determine classes of motions and transitions for which people are sensitive or insensitive to a transition method. A more fine-grained and robust assessment of the just noticeable difference would also be interesting. These are the subject of future investigations.

In our experience, our methods work on a wide variety of motions and transition points. However, we also performed a quantitative evaluation of these methods through a user study. Users were shown transitions between motions appropriate to the particular methods. These motions consisted of running at different speeds, walking at different speeds, standing, idling, boxing, dancing, and tai-chi. When compared against a fixed blend length, users preferred both the geodesic distance and velocity methods for calculating blend lengths. The geodesic distance method was strongly preferred. We also found that users can differentiate blend lengths that differ by approximately eight frames (0.27s). This result is generally consistent with differentiation and recognition of motion in point light experiments, e.g., [Joh73], but reinforces the point that a fixed blend length will not produce visually appealing results for a large repertoire of motion transitions. In our user study, there were no motion transitions for which users consistently preferred the fixed blend length.

A surprising result of our study was that users showed no preference for our timewarping method over the geodesic distance method. Timewarping was found to be helpful for generating transitions by a number of researchers, e.g., [BW95, RCB98, KG03]. We conjecture that there are two major reasons for this contradictory finding. First, the method of timewarping used by [BW95, KG03] is more sophisticated and powerful than our technique, involving dynamic programming. We avoided employing dynamic programming because its computational cost precludes its use in a system where performance demands are interactive and high, e.g., a video game. However, there are quite likely advantages to the more expensive approach. Another reason may be that timewarping has been found to be useful when the motions transitioned between have very different timings. While we included such motions in our user study, we may not have included motions with significant enough time variations to make timewarping necessary. More investigation of this area is ongoing.

The methods described in this paper could be easily integrated into the systems described by [LCR*02, KGP02, AF02, AFO03]. These systems determine transition points as part of their function. The only additional information needed would be the category of the motions, so that either the geodesic distance method or the velocity method could be chosen. The same holds true for integration into a video game. In particular, the computational cost of these methods is minimal and well within the performance bounds set for animation by most rendering engines.

The use of the algorithms described here implies that the total length of a transition will vary depending on the motions transitioned from and to. In our experience, the blend length is normally between 0.03 and 2 seconds. Additional constraints could be added to the method to further control the blend length if an animation system required them.

An important issue for any automatic technique for re-using motion data is its applicability to motions for which it has not been tested. While our motion capture library is reasonably extensive, it does not contain highly specific motions that would be needed in, for example, a video game dealing with hockey. We may find new categories of motion for which we require different methods. Moreover, the perception of visual artifacts depends upon the task [OHJ00] and upon the rendering style [HOT98]. Insofar as task differs from the motion itself we assume that these effects are not significant, but have not tested this assumption.

Also, linear blending often exhibits artifacts when footslide occurs. In this work, a support limb is constrained to prevent foot-slide using a particular inverse kinematics formulation. There are other solutions to this problem [KSG02, LS99]. We believe that the artifacts present in the motion are dominated by the relative velocities of the two motions and differences in pose, and thus *not* by the mechanism of support limb constraint. Testing this assumption remains future work.

Additionally, we would like to test and incorporate our methods into dynamical simulation systems that produce long streams of animation, e.g., [DYP03, ZH02, RGBC96].

© The Eurographics Association 2004.

These types of systems typically impose torque constraints that will affect the duration of transitions, but it is likely that leeway exists in picking the transition. Also, linear blending may produce self-intersection. Determining a fast way of detecting this situation would extend its utility.

Acknowledgments

References

[AF02] ARIKAN O., FORSYTH D. A.: Interactive motion generation from examples. *ACM Transactions on Graphics 21*, 3 (July 2002), 483–490. ISSN 0730-0301 (Proceedings of ACM SIGGRAPH 2002). 2, 8

[AFO03] ARIKAN O., FORSYTH D. A., O'BRIEN J. F.: Motion synthesis from annotations. *ACM Transactions on Graphics 22*, 3 (July 2003), 402–408. 2, 8

[BW95] BRUDERLIN A., WILLIAMS L.: Motion signal processing. In *Computer Graphics* (Aug. 1995), pp. 97–104. Proceedings of SIGGRAPH 95. Held in Los Angeles, CA. 2, 8

[DYP03] DONTCHEVA M., YNGVE G., POPOVIĆ Z.: Layered acting for character animation. *ACM Transactions on Graphics 22*, 3 (July 2003), 409–416. 8

[Gra00] GRASSIA F. S.: *Believable Automatically Synthesized Motion by Knowledge-Enhanced Motion Transformation*. Ph.d. dissertation, Carnegie Mellon University, Aug. 2000. CMU-CS-00-163. 3, 5

[GS66] GREEN D. M., SWETS J. A.: *Signal Detection Theory and Psychophysics*. John Wiley and Sons, New York, 1966. 6

[GSKJ03] GLEICHER M., SHIN H. J., KOVAR L., JEPSEN A.: Snap-together motion: Assembling run-time animation. *ACM Transactions on Graphics 22*, 3 (July 2003), 702–702. 2

[HOT98] HODGINS J. K., O'BRIEN J. F., TUMBLIN J. F.: Judgments of human motion with different geometric models. *IEEE Transactions on Visualization and Computer Graphics 4*, 4 (1998). 3, 8

[Joh73] JOHANSSON G.: Visual perception of biological motion and a model for its analysis. *Perception and Psychophysics 14* (1973), 201–211. 3, 8

[KG03] KOVAR L., GLEICHER M.: Flexible automatic motion blending with registration curves. In *Symposium on Computer Animation 2003* (San Diego, CA, July 2003), Breen D., Lin M., (Eds.), ACM SIGGRAPH/Eurographics, pp. 214–224. 2, 8

[KGP02] KOVAR L., GLEICHER M., PIGHIN F.: Motion graphs. *ACM Transactions on Graphics 21*, 3 (July 2002), 473–482. ISSN 0730-0301 (Proceedings of ACM SIGGRAPH 2002). 2, 3, 6, 8

[KSG02] KOVAR L., SCHREINER J., GLEICHER M.: Footskate cleanup for motion capture editing. In *ACM SIGGRAPH Symposium on Computer Animation* (July 2002), pp. 97–104. 3, 8

[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. *ACM Transactions on Graphics 21*, 3 (July 2002), 491–500. ISSN 0730-0301 (Proceedings of ACM SIGGRAPH 2002). 2, 8

[LS99] LEE J., SHIN S. Y.: A hierarchical approach to interactive motion editing for human-like figures. In *Proceedings of SIGGRAPH 99* (Aug. 1999), Computer Graphics Proceedings, Annual Conference Series, pp. 39–48. 8

[MBC01] MIZUGUCHI M., BUCHANAN J., CALVERT T.: Data driven motion transitions for interactive games. *Eurographics 2001 Short Presentations* (2001). 2, 3, 6

[Md98] MICHAELS C. F., DE VRIES M. M.: Higher order and lower order variables in the visual perception of relative pulling force. *Journal of Experimental Psychology: Human Perception and Performance 24*, 2 (1998), 526–546. 3

[ODGK03] O'SULLIVAN C., DINGLIANA J., GIANG T., KAISER M. K.: Evaluating the visual fidelity of physically based animations. *ACM Transactions on Graphics 22*, 3 (July 2003), 527–536. 3, 6

[OHJ00] OESKER M., HECHT H., JUNG B.: Psychological evidence for unconscious processing of detail in real-time animation of multiple characters. *The Journal of Visualization and Com-*

puter Animation 11, 2 (June 2000), 105–112. 3, 8

[PB02]   PULLEN K., BREGLER C.: Motion capture assisted animation: Texturing and synthesis. *ACM Transactions on Graphics 21*, 3 (July 2002), 501–508. 2, 3, 6

[Per95]   PERLIN K.: Real time responsive animation with personality. *IEEE Transactions on Visualization and Computer Graphics 1*, 1 (Mar. 1995), 5–15. 2

[PPBS01]   POLLICK F., PATERSON H. M., BRUDERLIN A., SANFORD A. J.: Perceiving affect from arm movement. *Cognition 82*, 2 (2001), B51–B61. 3

[PSE*00]   POPOVIĆ J., SEITZ S. M., ERDMANN M., POPOVIĆ Z., WITKIN A.: Interactive manipulation of rigid body simulations. In *Proceedings of SIGGRAPH 2000* (Aug. 2000), ACM SIGGRAPH, pp. 209–218. Held in New Orleans, Louisiana. 3

[PSS02]   PARK S. I., SHIN H. J., SHIN S. Y.: On-line locomotion generation based on motion blending. In *ACM SIGGRAPH Symposium on Computer Animation* (July 2002), pp. 105–112. 2

[RCB98]   ROSE C., COHEN M., BODENHEIMER B.: Verbs and adverbs: Multidimensional motion interpolation. *IEEE Computer Graphics and Applications 18*, 5 (1998), 32–40. 2, 3, 8

[RGBC96]   ROSE C. F., GUENTER B., BODENHEIMER B., COHEN M. F.: Efficient generation of motion transitions using spacetime constraints. In *Proceedings of SIGGRAPH 96* (New Orleans, Louisiana, August 1996), Computer Graphics Proceedings, Annual Conference Series, ACM SIGGRAPH / Addison Wesley, pp. 147–154. ISBN 0-201-94800-1. 2, 6, 8

[Ric95]   RICE J. A.: *Mathematical Statistics and Data Analysis*, 2nd ed. Duxbury Press, Belmont, CA, 1995. 7

[RP03]   REITSMA P. S. A., POLLARD N. S.: Perceptual metrics for character animation: Sensitivity to errors in ballistic motion. *ACM Transactions on Graphics* (July 2003). Proceedings of SIGGRAPH 2003, to appear. 3, 6

[SBS02]   SIDENBLADH H., BLACK M. J., SIGAL L.: Implicit probabilistic models of human motion for synthesis and tracking. In *Computer Vision — ECCV 2002 (1)* (Copenhagen, Denmark, May 2002), Heyden A., Sparr G., Nielsen M.,, Johansen P., (Eds.), Lecture Notes in Computer Science, Springer-Verlag, pp. 784–800. 7th European Conference on Computer Vision. 2

[SI87]   SOGON S., IZARD C. B.: Sex differences in emotion recognition by observing body movements. *Psychological Research 29* (1987), 89–93. 3

[SSSE00]   SCHÖDL A., SZELISKI R., SALESIN D. H., ESSA I.: Video textures. In *Proceedings of ACM SIGGRAPH 2000* (July 2000), Computer Graphics Proceedings, Annual Conference Series, ACM Press / ACM SIGGRAPH / Addison Wesley Longman, pp. 489–498. ISBN 1-58113-208-5. 2

[WB03]   WANG J., BODENHEIMER B.: An evaluation of a cost metric for selecting transitions between motion segments. In *Symposium on Computer Animation 2003* (San Diego, CA, July 2003), Breen D., Lin M., (Eds.), ACM SIGGRAPH/Eurographics, pp. 232–238. 2, 3

[WP95]   WITKIN A. P., POPOVIĆ Z.: Motion warping. In *Proceedings of SIGGRAPH 95* (Aug. 1995), Computer Graphics Proceedings, Annual Conference Series, pp. 105–108. 2

[ZH02]   ZORDAN V. B., HODGINS J. K.: Motion capture-driven simulations that hit and react. In *ACM SIGGRAPH Symposium on Computer Animation* (July 2002), pp. 89–96. 8