Eurographics/SIGGRAPH Symposium on Computer Animation (2003)
D. Breen, M. Lin (Editors)

An Evaluation of a Cost Metric for Selecting Transitions between Motion Segments

Jing Wang† and Bobby Bodenheimer‡

Department of Computer Science and Electrical Engineering
Vanderbilt University
Nashville, TN 37235

Abstract

Designing a rich repertoire of behaviors for virtual humans is an important problem for virtual environments and computer games. One approach to designing such a repertoire is to collect motion capture data and pre-process it to form a structure that can be walked in various orders to re-sequence the data in new ways. In such an approach identifying the location of good transition points in the motion stream is critical. In this paper, we evaluate the cost function described by Lee et al.[15] for determining such transition points. Lee et al. proposed an original set of weights for their metric. We compute a set of optimal weights for the cost function using a constrained least-squares technique. The weights are then evaluated in two ways: first, through a cross-validation study and second, through a medium-scale user study. The cross-validation shows that the optimized weights are robust and work for a wide variety of behaviors. The user study demonstrates that the optimized weights select more appealing transition points than the original weights.

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

1. Introduction

Designing a rich repertoire of behaviors for virtual humans is a challenging task that has seen much work in recent years. Motion capture allows one to construct a large library of raw motion, but processing that motion into a finished product such as a video game character is still a labor intensive process. Minimizing the amount of manual intervention in processing motion capture data has been the focus of much recent research.

This research typically falls into two areas of effort. One area creates motion from an underlying probabilistic model.[17, 3, 19, 5] In this approach motion capture provides training data for the model. The second approach is motivated by the work of Schödl et al.[22] and creates new motion by re-ordering the original motion.[15, 10, 1] In this approach the qualities of the original data are emphasized.

The key to this second approach is the proper selection of transition points, points at which the motion will change from one segment of captured motion to another segment, either within the same motion or another motion. Because these transition points represent discontinuities in the motion stream, selection of good transition points can be crucial to the quality of the resulting motion. Each of the works cited above uses a different distance function to calculate the cost of transitioning from one frame to another. These distance functions are all parameterized through user-selected weights.

The current paper evaluates the cost function proposed by Lee et al.[15] for determining the transition cost. Studies of such cost functions have not been conducted, although their importance is readily acknowledged, e.g., see Lamouret and van de Panne.[13] This paper thus presents the first empirical evaluation of one component of a complete animation

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

system. The contribution of the paper lies in producing a set of optimized weights that select good transitions. These weights were evaluated first through a cross-validation study that shows the optimization is robust, and second through a user study that confirms that the weight selection is superior.

The paper is organized as follows. In Section 2 we provide background information on this area of animation and the cost function. In Section 3 we describe the process of optimizing weights for the cost functions. In Section 4 we describe both the process of cross-validation for the optimized weights and the user study that was conducted, and report their results. Finally, in Section 5 we discuss these results and provide directions for future work.

2. Background

Motion capture research has concentrated on studying ways of editing and modifying existing motions. See Gleicher[7] for a survey of work in this area. As mentioned previously, work in minimizing the amount of manual editing needed for using motion capture has been approached in two primary ways: through the use of probabilistic methods that synthesize new motion and through methods that re-use the original data for synthesis.

Probabilistic methods for motion synthesis build a model based on aggregate or statistical qualities of a set of training examples.[3, 2, 5, 19, 17] These techniques are very powerful, but may eliminate subtleties of the motion during synthesis that give the motion a sense of richness.

In contrast, other researchers have drawn inspiration from the work of Schödl et al.[22] on video textures to retain the original motion sequences but play them back in non-repetitive streams. Sidenbladh et al.[23] employ a probabilistic search method to find the next pose in a motion stream and obtain it from a motion database. Arikan and Forsyth[1] construct a hierarchy of graphs connecting a motion database and use randomized search to extract motion satisfying specified constraints. Kovar et al.[10] use a similar idea to construct a directed graph of motion that can be traversed to generate different styles of motion. Lee et al.[15] model motion as a first-order Markov process and also construct a graph of motion. They demonstrate three interfaces for controlling the traversal of their graph.

One of the distinguishing features among these papers is that they employ different underlying cost metrics for evaluating transition points in the graph. Lee et al. use a cost function based on joint orientations and velocities. Kovar et al. use a cost function based on the distance between point samples of the mesh representation of the character. Arikan and Forsyth use a hybrid method similar to that of Lee but involving joint accelerations as well. Sidenbladh et al. use a probabilistic search method. A principled understanding of the best cost function is not known.

In this paper we evaluate the cost metric based on joint orientations because of the fact that motion capture data is typically represented by joint orientations of the skeleton. Therefore, we chose the cost metric described by Lee et al. Compared to other cost metrics, the cost metric used by Lee et al. can be computed directly and quickly.

Our evaluation consists, in part, of a user study. Such studies have a tradition in the psychological literature dating from the original light point studies of Johansson.[9] In these studies, participants can identify such things as the gender[12] and emotional state of the actor.[24] However, these studies are after a much coarser evaluation of motion than we are interested in, typically evaluating whether the motion is biologically produced or not. Oesker et al.[18] perform a study methodologically similar to our own but trying to assess the effects of level of detail in animation. Their experimental design differs from ours in that it is within-subjects as opposed to between-subjects. They conclude, in part, that variations in animation style influence observer's evaluation of animated character "skill." Reitsma and Pollard [20] conducted user studies on perception of errors in ballistic motion and proposed an empirical metric based on their findings.

2.1. The Cost Function

In this section, the specifics of the cost function are reviewed. For further details, the reader is referred to the original paper describing the function.[15]

Lee et al. construct a matrix of probabilities for transitioning from one frame to another. The probabilities are constructed from a measure of similarity between the frames. We will study the cost in the paper, that is, we are not studying the probability of the transitions. The cost for transitioning from frame i to frame j is given by

$$D_{ij} = d(p_i, p_j) + \nu d(v_i, v_j) \tag{1}$$

where $d(v_i, v_j)$ is the weighted distance of joint velocities, ν weights the velocity difference with respect to $d(p_i, p_j)$, and $d(p_i, p_j)$ is the weighted difference of joint orientations. This term is given by

$$d(p_i, p_j) = \left\| p_{i,0} - p_{j,0} \right\|^2 + \sum_{k=1}^{m} w_k \left\| \log\left(q_{j,k}^{-1} q_{i,k} \right) \right\|^2. \tag{2}$$

In Equation 2, $p_{i,0}, p_{j,0} \in \mathbf{R}^3$ are the global translational positions of the figure at frames i and j, respectively; m is the number of joints in the figure; and $q_{i,k}, q_{j,k}$ are the orientations of joint k and 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.

For this work, our skeleton consisted of 16 joints and the complete figure had 54 degrees of freedom. Each joint was a three degree of freedom joint, and there were degrees of freedom for global position and orientation. All motion capture data was sampled at 30 frames per second.

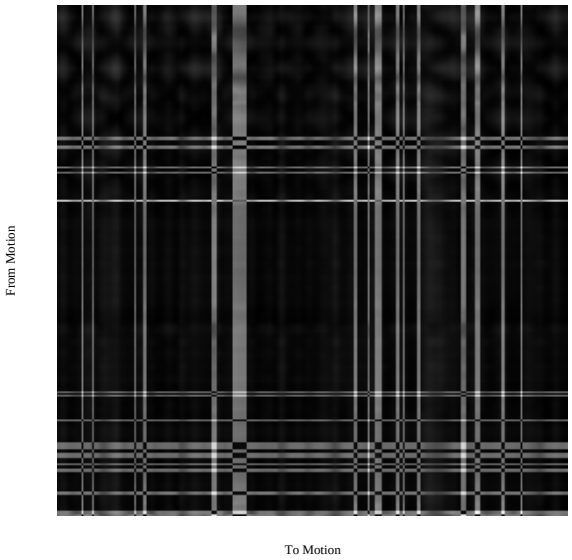

Figure 1: *The cost matrix for two clips of dance motion with original weights. Each motion is 10 seconds long. Darker values correspond to lower costs for transition.*

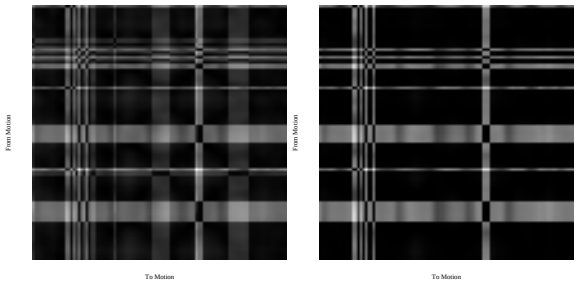

Figure 2: *The cost matrix for two clips of dancing motion using the original weights. In (a)* $\nu = 0.1$ *and in (b)* $\nu = 10$.

3. Optimizing the Weights

The cost function contains parameters to modify the transition cost. The parameters take the form of weights. The cost function weights both the geodesic norm between joint orientations and the joint velocities, and contains another parameter, ν, trading off the velocity and position distances. Lee et al. report setting the weights to one for the shoulders, elbows, hips, knees, pelvis, and spine; others are set to zero. No value for ν is given.

We would like to use motion capture to determine optimal values for the weights. We will contrast motions using optimized weights versus the weights Lee et al. report. We will refer to the sets of weights used by Lee et al. as the original weights. An example of the cost function for transitions from one motion to another for the original weights is shown in Figure 1 with $\nu = 1$. The figure is normalized so that an intensity of zero corresponds to the minimal value of the cost function for that motion and an intensity of 255 corresponds to the maximum cost. The minimum and maximum before normalization are 0.0993 and 20.0677. The cost function is reasonably uniformly distributed over its ranges, so the linear normalization gives an accurate picture of the function's variation.

To optimize the weights, we took a set of 16 different segments of captured motion, each several seconds long. These segments consisted of a variety of motions including walking in different styles, running in different styles, jogging, dancing, and gesturing. For these segments we manually selected 16 good transitions and 26 bad transitions. A good transition was one in which the visual discontinuity of the transition was minimal; a bad transition was one in which the visual discontinuity was disconcerting. The transitions were selected by a single person with animation experience and critically examined by two other experienced viewers for approval. Our optimization will depend on how well these transitions were selected, but in our experience it is not difficult to manually select good and bad transitions.

We then solved for the optimal values of the weights using a constrained least-squares minimization, that is,

$$\min_{w} ||Aw - b||_2^2 \tag{3}$$

where w is a vector of weights, A is a matrix of the position and velocity distances of Equation 1; b is a vector of ones and zeros—an entry is one if it corresponds to a bad transition, and zero if it corresponds to a good transition. The optimization was constrained such that the weights were non-negative and symmetric, i.e., the weight for the left shoulder must be identical to the right shoulder. The symmetry constraint makes intuitive sense but will generally not be the result of the optimization without this constraint. The optimization problem was solved using an active set method similar to that described in Gill et al.[6]

The weight ν enters the terms in the A matrix above non-linearly, thus requiring a more complicated optimization method. However, for motions in our database, our experience is that the velocity term makes little effective difference in the cost. Figure 2 shows this insensitivity to ν for transitioning from one dance motion to another dance motion, with $\nu = 0.1$ and $\nu = 10$. In fact, the global minimum for this motion was unchanged as ν was varied from 0 to 100. Thus, ν remained one in the optimization process.

The normalized weights (largest scaled to one) from the optimization process are shown in Table 1. The cost matrix for the motions using the optimal weights are shown in Figure 3. This figure is for the same motions that have the cost matrix shown in Figure 1. There is substantial difference between the original weights and the new weights. In general,

Right and Left Hip	1.0000
Right and Left Knee	0.0901
Right and Left Shoulder	0.7884
Right and Left Elbow	0.0247

Table 1: *Joints with non-zero weights and their associated weights when solved as described in the text. The optimization zeroed the weights for the remaining joints.*

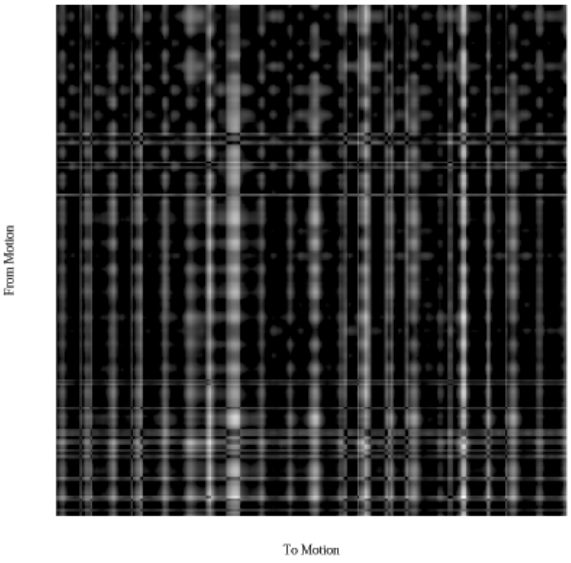

Figure 3: *The cost matrix for two clips of dance motion with optimized weights. Each motion is 10 seconds long. Darker values correspond to lower costs for transition.*

the cost with new weights has become more restrictive. Numerically, the optimization zeroed weights associated with joints found to be unimportant. Most of the weights were found to be unimportant: only the hips, knees, shoulders and elbows were important. The result is consistent with our expectation that those joints are the most important ones, but surprising nonetheless since they imply the rest are unimportant in selecting a reasonable transition.

4. Evaluation

4.1. Cross-Validation

To estimate the generalization rate of the optimized weights, we employed a full leave-one-out cross-validation study.[4] In this technique, the weights are optimized with one set of training data deleted, and the resulting weights are then used to compute the optimal value of a transition for the deleted data set. Recall that our training set contained a rich variety of motions. The results of this study were quite encouraging. The average deviation between the full optimization and that of the leave-one-out optimization was less than one frame in the animation sequences. The median deviation was zero frames. Additionally, we performed a similar test by again deleting one set of training data, re-computing the optimal weights, and then computing the optimal transition for a completely different motion than the weights were trained on, a dancing motion from a different performer containing different dynamics. For this study, the resulting weights computed the same optimal transition in 41 of 42 cases. For the case where the optimal transition was not computed identically to the other cases, it was one frame different. Based on these empirical approaches, we believe that the optimal weights we computed are both robust and generalize to pick reasonable transitions for a wide variety of motions. However, whether the optimal weights are necessarily better than the original weights cannot be verified by this technique. Instead, we must conduct a user study to determine the result.

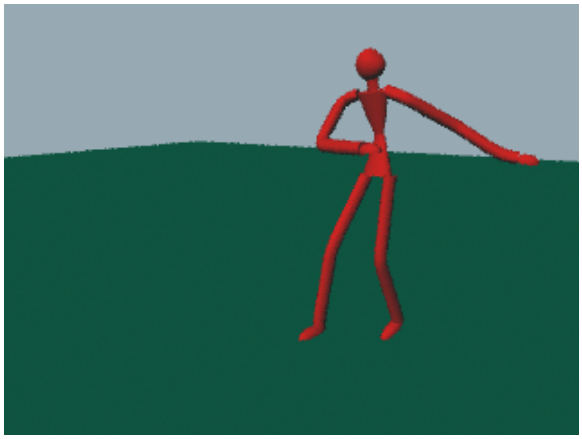

Figure 4: *The animated character used in the user study.*

4.2. User Study

A user study was conducted to evaluate the weighting determined by the optimization. A motion capture sequence of dancing was created by a performer different from the one used to capture the motions used in Section 3 for optimizing the weights. This practice was employed to eliminate the possibility of any performer-dependent effects on the weighting. A frame of an animated sequence used in the study is shown in Figure 4.

The participant group consisted of 26 adults with normal color vision who had no prior experience working with animation outside of computer games and the like. Participants in the study were told they would be viewing an animation of motion sequences and shown an example animation of a walking sequence using the model that would be used in

the experiment. They were then told the motions they would be viewing would have a discontinuity in the motion, and shown an example of an egregious discontinuity. Participants were told they would be asked to rate how noticeable and natural the discontinuities were, both individually and in comparison to another motion. Participants were shown two motions. Each motion was a six-second clip; the transition from the first motion sequence to the second occurred at $t = 3$s in the clip. The motions consisted of the globally optimum Lee cost transition with the weights used in Lee at al.[15], and the globally optimum Lee cost transition using weights determined in Section 3. These two motions were different, i.e., the lowest cost transition occurred at different points in the motion for each cost function. In particular, for the dancing motion used in the user study, the optimal transition for the original cost occurred from frame 91 to frame 281, whereas the optimal cost transition occurred from frame 232 to frame 280.

The sets of motions participants were asked to compare were the original cost versus the optimal cost. To best evaluate the transition, no interpolation or smoothing between the sequences was done. Since a complete animation system, such as that present in a video game, will employ some sort of motion transition mechanism, this decision may seem odd. However, we believe that this decision is necessary for the following reasons. Employing a motion transition mechanism involves making many engineering decisions about how the motions are to be blended. For example, the time over which the transition will occur must be specified. The method of blending must be determined, e.g., linear interpolation, ease-in ease-out, or employing a very sophisticated mechanism such as in Rose et al.[21] Additionally, some sort of inverse kinematics routine is usually required, because blending introduces the problem of foot-skate or foot-slide (see, for example, Kovar et al.[11]). There are a number of inverse kinematics routines available,[16, 11, 21] and each of them also makes engineering decisions that affect the quality of the resulting motion.

Additionally, blending works two ways. It can mask the selection of a bad transition (for example, it can smooth a discontinuity), or it can make a bad transition extremely obvious (for example, if the transition causes one part of the body to intersect with another). Which effect predominates depends on the quality of the transition selected. Our work tries to evaluate the quality of transitions in the absence of the other confounding factors such as the decisions made in creating smooth transitions. Once we have empirically established that a metric produces reasonable or good transitions, then we can begin to evaluate the engineering factors associated with producing visually compelling and optimal transitions. We note, however, for each transition the global position and orientation of the character was matched at the point of transition.

We controlled for order effects in the presentation by randomly dividing the participants into two equal-sized groups. The first group was presented the original and optimized weighted motion sequences first, the second was presented the optimally weighted motion sequences first. After viewing each sequence, participants completed a post-sequence questionnaire consisting of six questions asking them to compare and rate their impressions of the motions using a five-point Likert scale. Likert scale responses rather than forced choice responses were chosen to exploit the statistical power the former offers in distinguishing subtle differences. Users were asked which sequence seemed to have better quality, was more natural, and compare the sequences based on the noticeability of the transitions. Where applicable, the results of the user study were analyzed using two-way between-subjects analyses of variance (ANOVAs). The ANOVAs were 2×2 with the independent variables being cost function and presentation order. Results were considered significant if $p < 0.05$.

Comparison	Mean	Std. Dev.
Optimized Weight vs. Original Weight		
Looks Better?	2.73	1.00
More Natural?	2.69	0.84

Table 2: *Summary of results for direct comparisons of optimized versus original functions. Preferences were rated on a scale of zero to four where zero corresponded to looking much worse (or very unnatural) and four corresponded to looking much better (or very natural). For example, participants were asked if they thought that the optimized motion "looked better" than the original motion.*

In general, participants considered the optimally weighted better and more natural than the original weighted. These results are summarized in Table 2. To support these results, participants also found that the optimal transition for the weighted cost to be significantly more realistic than the original cost ($F = 6.39$, $MS = 6.94$, $p = 0.01$). No significant interaction effects were present in the ordering of the motions for any of the comparisons.

5. Discussion

This paper produced weights for the cost function described by Lee et al. that led to the determination of superior transitions. We cross-validated the weights to assess the generalizeability and robustness of the optimization procedure. We compared the optimized weights with the original weights by running a user study. The user study showed that optimized weights produce perceptually better transition points.

However, there are several limitations and possible sources of bias in our results. When optimizing the weights,

our motion data did not contain highly dynamic motions such as would be typical from a gymnastic floor exercise or other sources. The weights may not be a good predictor of good transitions for such motions. It was also limited to 16 different sequences of motion for the optimization. We would like to repeat our experiments using a larger library with more dynamic motion. Rendering style can affect the quality of the perceived motion as shown by Hodgins et al.[8] Our weight optimization used only one performer, although our motions were tested in the cross-validation and the user study on motion generated by a different performer. Finally, our motion data did not contain a large repertoire of "backward" motions, which may have resulted in the position-velocity weight ν having marginal impact for the Lee cost. Our data suggest that the velocity component of the Lee cost is not significant for a wide variety of motions. Removing these limitations or better understanding their necessity is an on-going project.

We are interested in examining other cost metrics.[10, 1] These cost metrics are based on different features of motions and how they are represented. We want to evaluate which cost metric could pick the best transition point sets for motion capture reusing.

We plan to investigate the effect of blending on these transitions as well. As noted earlier, blending smooths some of the discontinuities in the motion. At least for linear interpolation, however, noticeable differences are still apparent. There are three methods of smoothing the transitions that we would like to understand better from a perceptual viewpoint: linear interpolation (possibly with ease-in ease-out), filtered smoothing using the machinery developed in Lee and Shin,[14] and dynamic generation of transitions as done in Rose et al.[21] The use of a transition mechanism in conjunction with a cost metric may have some effect on the weights, and we would like to determine what that effect is, if it exists.

Additional topics for further research include extending the parameters of the user study to better assess how people perceive the transitions and what characteristics are important to them in determining the quality of the motions. Most participants in the study reported that the transition break was noticeable. Assessing how an interpolation scheme impacts the perception of the transition is likely to be a difficult task, but could yield important insights into the characteristics of visually compelling motion.

The increasing richness of human characters in three-dimensional computer games should make transitions in motion data an increasingly important problem. We believe that these results give significant guidance to those concerned with creating virtual humans with a rich repertoire of behaviors, and may help in the re-use of large motion capture data-sets.

6. Acknowledgments

The authors thank Christina de Juan for useful discussions throughout the project, and to the anonymous reviewers for constructive comments on the presentation. The authors are grateful to Steve Park and the Graphics, Visualization, and Usability Center at the Georgia Institute of Technology for supplying some of the motion capture data used in this study.

References

1. Okan Arikan and D. A. Forsyth. Synthesizing constrained motions from examples. *ACM Transactions on Graphics*, 21(3):483–490, July 2002. ISSN 0730-0301 (Proceedings of ACM SIGGRAPH 2002).

2. Richard Bowden. Learning statistical models of human motion. In *IEEE Workshop on Human Modelling, Analysis, and Synthesis*, 2000. CVPR 2000.

3. Matthew Brand and Aaron Hertzmann. Style machines. In *Proceedings of ACM SIGGRAPH 2000*, Computer Graphics Proceedings, Annual Conference Series, pages 183–192. ACM Press / ACM SIGGRAPH / Addison Wesley Longman, July 2000. ISBN 1-58113-208-5.

4. Richard O. Duda, Peter E. Hart, and David G. Stork. *Pattern Classification*. John Wiley and Sons, second edition, 2001.

5. A. Galata, N. Johnson, and D. Hogg. Learning variable length markov models of behaviour. *Computer Vision and Image Understanding Journal*, 81(3):398–413, March 2001.

6. Phillip E. Gill, Walter Murray, and Margaret H. Wright. *Practical Optimization*. Academic Press, 1981.

7. Michael Gleicher. Comparing constraint-based motion editing methods. *Graphical Models*, 63(2):107–134, March 2001. ISSN 1524-0703.

8. Jessica K. Hodgins, James F. O'Brien, and Jack F. Tumblin. Judgments of human motion with different geometric models. *IEEE Transactions on Visualization and Computer Graphics*, 4(4), 1998.

9. G. Johansson. Visual perception of biological motion and a model for its analysis. *Perception and Psychophysics*, 14:201–211, 1973.

10. Lucas Kovar, Michael Gleicher, and Frédéric Pighin. Motion graphs. *ACM Transactions on Graphics*, 21(3):473–482, July 2002. ISSN 0730-0301 (Proceedings of ACM SIGGRAPH 2002).

11. Lucas Kovar, John Schreiner, and Michael Gleicher. Footskate cleanup for motion capture editing. In *ACM SIGGRAPH Symposium on Computer Animation*, pages 97–104, July 2002.

12. L. T. Kozlowski and J. E. Cutting. Recognizing the gender of walkers from dynamic point-light displays. *Perception and Psychophysics*, 21:575–580, 1977.

13. A. Lamouret and M. van de Panne. Motion synthesis by example. In *EGCAS '96: Seventh International Workshop on Computer Animation and Simulation*. Eurographics, 1996. ISBN 3-211-82885-0.

14. J. Lee and S. Y. Shin. General construction of time-domain filters for orientation data. *IEEE Transactions on Visualization and Computer Graphics*, 8(2):119–128, April - June 2002. ISSN 1077-2626.

15. Jehee Lee, Jinxiang Chai, Paul S. A. Reitsma, Jessica K. Hodgins, and Nancy S. Pollard. Interactive control of avatars animated with human motion data. *ACM Transactions on Graphics*, 21(3):491–500, July 2002. ISSN 0730-0301 (Proceedings of ACM SIGGRAPH 2002).

16. Jehee Lee and Sung Yong Shin. A hierarchical approach to interactive motion editing for human-like figures. In *Proceedings of SIGGRAPH 99*, Computer Graphics Proceedings, Annual Conference Series, pages 39–48, August 1999.

17. Yan Li, Tianshu Wang, and Heung-Yeung Shum. Motion texture: A two-level statistical model for character motion synthesis. *ACM Transactions on Graphics*, 21(3):465–472, July 2002. ISSN 0730-0301 (Proceedings of ACM SIGGRAPH 2002).

18. Markus Oesker, Heiko Hecht, and Bernhard Jung. Psychological evidence for unconscious processing of detail in real-time animation of multiple characters. *The Journal of Visualization and Computer Animation*, 11(2):105–112, June 2000.

19. Katherine Pullen and Christoph Bregler. Animating by multi-level sampling. In *Computer Animation 2000*, pages 36–42. IEEE CS Press, May 2000. ISBN 0-7695-0683-6.

20. Paul S. A. Reitsma and Nancy S. Pollard. Perceptual metrics for character animation: Sensitivity to errors in ballistic motion. *ACM Transactions on Graphics*, July 2003. Proceedings of SIGGRAPH 2003, to appear.

21. Charles F. Rose, Brian Guenter, Bobby Bodenheimer, and Michael F. Cohen. Efficient generation of motion transitions using spacetime constraints. In *Proceedings of SIGGRAPH 96*, Computer Graphics Proceedings, Annual Conference Series, pages 147–154, New Orleans, Louisiana, August 1996. ACM SIGGRAPH / Addison Wesley. ISBN 0-201-94800-1.

22. Arno Schödl, Richard Szeliski, David H. Salesin, and Irfan Essa. Video textures. In *Proceedings of ACM SIGGRAPH 2000*, Computer Graphics Proceedings, Annual Conference Series, pages 489–498. ACM Press / ACM SIGGRAPH / Addison Wesley Longman, July 2000. ISBN 1-58113-208-5.

23. Hedvig Sidenbladh, Michael J. Black, and L. Sigal. Implicit probabilistic models of human motion for synthesis and tracking. In A. Heyden, G. Sparr, M. Nielsen, and P. Johansen, editors, *Computer Vistion — ECCV 2002 (1)*, Lecture Notes in Computer Science, pages 784–800, Copenhagen, Denmark, May 2002. Springer-Verlag. 7th European Conference on Computer Vision.

24. S. Sogon and C. B. Izard. Sex differences in emotion recognition by observing body movements. *Psychological Research*, 29:89–93, 1987.