Trajectory Synthesis by Hierarchical Spatio-temporal Correspondence: Comparison of Different Methods

Johannes Mezger*
WSI/GRIS
University of Tübingen, Germany

Winfried Ilg† Martin A. Giese‡
Laboratory of Action Representation and Learning (ARL)
Hertie Inst. for Clinical Brain Research, Tübingen, Germany

Figure 1: Rendering of synthesized Karate form ('kata').

Abstract

We present several methods for the generation of complex human motion trajectories by linear combination of prototypical example trajectories with well-defined styles. These methods decompose longer trajectories automatically into movement primitives by robust matching with stored templates. To synthesize movement primitives with new style properties, segments from the prototype trajectories are linearly combined. These linear combinations are based on the computation of spatio-temporal correspondence between trajectory segments. The synthesized new movement primitives are automatically concatenated into longer action sequences, trying to minimize artifacts at the transition points. The proposed methods are evaluated by synthesizing movement sequences from martial arts ("karate katas") that include movements primitives with different styles. For assessing the physical correctness of the generated movements we employ a zero-moment-point criterion. This physical measure was very similar for real human movement trajectories and trajectories synthesized by linear combination. In addition, we evaluated the perceptual quality of the synthesized movement sequences in a psychophysical study, involving naive subjects and computer graphics experts. We found significant differences between the different methods. For complex movements methods based on space-time correspondence seem to outperform algorithms without time-warping. In addition, computer graphics experts seem to be more sensitive to artifacts in the trajectories than normal observers.

CR Categories: I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Animation

Keywords: Character Animation, Motion Morphing, Motion Capture, Linear Combination, Movement Primitive, Realism, Perceptual Dimension

*e-mail: mezger@gris.uni-tuebingen.de
†e-mail: winfried.ilg@uni-tuebingen.de
‡e-mail: martin.giese@uni-tuebingen.de

1 Introduction

For many technical applications, e.g. in the entertainment industry or for the design of man-machine interfaces, the synthesis of virtual characters moving in a realistic human-like way is an important problem. Even though the rendering quality of animated films and computer games has continuously increased in the last decade, it remains a challenging problem to animate such characters with movements that induce the percept of highly realistic human motion. Even synthetic movements that are physically correct and smooth often appear unnatural and mechanical, or "robot-like". The synthesis of flexible models for human-like complex movements is thus a highly relevant and challenging research problem, which needs to be addressed by combining computer graphics with perception research.

Present methods for computer animation are based on three major approaches: keyframe editing, physics based simulation, and motion capture. *Keyframe editing* demands great skill and experience of the animator. Additionally, it is quite difficult to synthesize in particular the higher frequency components of complex human movements that carry important information, e.g. about different movement styles. *Physics-based simulation* approaches suffer from the problem that detailed dynamical models for human full-body movements are very complex. The design of such models, which produce highly realistic human movements, has turned out to be very difficult. In addition, the parameters of physical models are often only very indirectly linked to perceived style properties of human movements. Because of these limitations of other approaches, *motion capture* has become a very popular approach for the synthesis of high-quality animations of human movements. The importance of this approach is also continuously growing because of the increasing availability of 3D motion capture systems. Animations by motion capture preserve all important details of human movements. However, they require that all desired movements have to be precisely predefined and executed by an actor. This provides relatively low flexibility for the animator. In addition, the recording and pre-processing of motion capture data is very time-consuming. This motivates the development of algorithms for the synthesis of novel highly realistic movement trajectories by interpolation between a small number of recorded prototypical trajectories.

Synthesis of movement trajectories by interpolation is a common approach in computer animation. Some approaches in the context of keyframe animation require the animator to specify key poses and actions providing "spacetime constraints" for an automatic synthesis of trajectories that satisfy Newton's laws [Witkin and Kass 1988; Gleicher 1997; Faloutsos et al. 2001; Fang and Pollard 2003]. For the determination of trajectories that are consistent with physical laws of motion criteria based on the zero-moment-point (ZMP, i.e. the point where all moments acting on the body cancel out each other) have been shown to be particularly useful [Shin et al. 2003; Tak and Ko 2003]. Other approaches synthesize movements by interpolation or blending between captured trajectories [Unuma et al. 1995; Bruderlin and Williams 1995; Witkin and Popović 1995; Brand and Herzman 2000]. One class of these methods approximates trajectories by linear combinations of example trajectories, defining *spatio-temporal morphable models (STMM)* [Giese and Poggio 2000; Lim and Thalmann 2001]. Interpolation between trajectories has also been realized with radial basis functions [Rose et al. 1998] or hierarchical splines [Lee and Shin 1999]. For the interpolation of joint rotations linearization methods have been developed [Grassia 1998; Lee and Shin 2002]. In addition, algorithms for the registration of the coordinate frames of interpolated trajectories before the morphing have been proposed [Kovar and Gleicher 2003]. Interpolation techniques have been combined with inverse kinematics in order to enforce the compatibility with specific end-effector positions [Park et al. 2002]. Several recent approaches synthesize long motion sequences by concatenation of shorter trajectory segments with different styles, e.g. using statistical learning methods [Li et al. 2002], or by searching appropriate transitions between segments in a motion database [Choi et al. 2003; Arikan et al. 2003; Kovar and Gleicher 2004]. The work in this paper is based on a method that generates long movements sequences by establishing spatio-temporal correspondence between example movement sequences using a hierarchical representation [Giese et al. 2002; Ilg et al. 2004]. The method extracts movement primitives and interpolates between them by linear combination in space-time. Recently several authors assessed the perception of computer-generated human motion, e.g. by evaluating physically inaccurate movements [Reitsma and Pollard 2003] or transitions between individual motion segments [Wang and Bodenheimer 2003; Wang and Bodenheimer 2004].

This paper is structured as follows: We first briefly describe a method for the synthesis of movement sequences that is based on hierarchical spatio-temporal correspondence in section 2. We then extend this method by introducing a new morphing algorithm for the linear combination of movements with time-continuous weights (section 3). In section 4 we discuss the evaluation of the proposed method using a physical measure that characterizes the compatibility of the executed movements with a mechanical stability criterion. Finally, in section 5 we present a psychophysical study with naïve and expert observers that evaluates the perceptual quality of animations generated with the different methods. Conclusions are presented in the last section.

2 Hierarchical Spatio-temporal Morphable Models (HSTMM)

One method for the synthesis of trajectories for action sequences that contain movements with multiple styles is based on establishing hierarchical space-time correspondence between prototype trajectories with different styles. For this purpose, the prototypical trajectories are first segmented into shorter meaningful segments, called *movement primitives*. The trajectory segments from different prototypes that correspond to the same movement primitive are then brought into spatio-temporal correspondence, and linearly combined. The synthesized movements for the individual primitives are then concatenated into a new longer action sequences, trying to avoid artifacts at the transition points between different primitives. In the following, we give a brief description of the individual steps of the algorithm, which is described in more detail in [Giese et al. 2002; Ilg et al. 2004].

2.1 Space-Time Correspondence

Spatio-temporal correspondence defines a field of spatio-temporal displacements that transforms (warps) a trajectory $x_p(t)$ (with $x_p : \mathbb{R} \rightarrow \mathbb{R}^d$) into a reference trajectory $x_r(t)$. The reference trajectory can be defined as an appropriately-chosen average of prototype trajectories that are linearly combined. The chosen definition of space-time correspondence assumes that the trajectory x_p is related to the reference trajectory x_r by time-warping, defined by a temporal displacement function $T_p(t)$, and by adding a spatial displacement function $X_p(t)$:

$$x_p(t) = x_r(t + T_p(t)) + X_p(t), \tag{1}$$

The determination of the spatial and temporal displacement functions is an ill-posed problem [Giese and Poggio 2000]. However, it can be solved by appropriately constraining the set of admissible displacement functions. Typically, it is assumed that these functions are smooth. In addition, the temporal displacement function $T_p(t)$ must define a unique relationship between the warped time $t' = t + T_p(t)$ and the original time t. This is obviously fulfilled for $T'_p(t) > -1$. A unique solution can be found by minimizing the weighted sum of the L_2-norms of the spatial and temporal shifts, resulting in the error functional:

$$E[X_p, T_p] = \int \left[|X_p(t)|^2 + \lambda\ T_p(t)^2 \right] dt \tag{2}$$

The positive constant λ controls the trade-off between time-alignment and size of the spatial shifts. The minimization of this error functional under the above constraints can be accomplished most efficiently by dynamic programming (see [Bruderlin and Williams 1995; Giese and Poggio 2000] for further details.)

2.2 Linear Combination of Trajectories

If space-time correspondence between a set of prototypical trajectories x_p, $1 \leq p \leq N$, and a reference trajectory x_r has been established, the linear combinations of the corresponding spatio-temporal displacement fields is given by

$$X(t) = \sum_{p=1}^{N} w_p X_p(t) \quad \text{and} \quad T(t) = \sum_{p=1}^{N} w_p T_p(t). \tag{3}$$

We always assume convex linear combinations with weights that fulfill $\sum_p w_p = 1$. The movement trajectory that corresponds to these linearly combined displacement fields can be recovered by space-time warping of the reference trajectory with the linearly combined displacement fields:

$$x(t) = x_r(t + T(t)) + X(t). \tag{4}$$

Unit weight vectors specify trajectories that closely approximate one of the prototypes. Weight vectors with components of intermediate size ($0 < w_p < 1$) specify trajectories that interpolate between the spatio-temporal properties of the prototypes.

2.3 Segmentation of Movement Primitives

To apply spatio-temporal correspondence to longer trajectories that contain movements with different style, it is useful to decompose action sequences into meaningful shorter elements with constant style properties. One possibility for the automatic segmentation into such movement primitives has been proposed in [Ilg et al. 2004]. Trajectories and templates for individual primitives are characterized by robust trajectory features (key events) that are defined by the zeros of the velocities in selected degrees of freedom. Subsequences of key events of the long trajectory are matched with templates derived from hand-segmented trajectories by dynamic programming. This matching is robust against additional or missing key events. The automatic segmentation algorithm determines the trajectory segments of all prototype trajectories, which correspond to the individual movement primitives. The trajectory segment of the p-th prototype and primitive i is defined by the time interval $[t_{p,i}, t_{p,i+1}]$. (Further details about the segmentation algorithm are given in [Ilg et al. 2004].)

2.4 Normalization and Concatenation

The start and end points of the trajectory segments derived by linear combination are typically dependent on the weights w_p. If synthesized trajectories of movement primitives with different styles (weight vectors) are concatenated into longer trajectories changes of the weight vector can induce discontinuities in the concatenated trajectory. This problem can be solved by appropriate normalization of the trajectories before the computation of correspondence. One normalization technique [Giese et al. 2002] transforms the trajectory segments $x_{p,i}$, belonging to prototype p and movement primitive i, by subtracting a linear function that ensures that the normalized trajectory $\tilde{x}_{p,i}(t)$ has zero values at the segment boundaries:

$$\begin{aligned} \tilde{x}_{p,i}(t) &= x_{p,i}(t) - x_p(t_{p,i}) \\ &\quad - (t - t_{p,i}) \, \frac{x_p(t_{p,i+1}) - x_p(t_{p,i})}{t_{p,i+1} - t_{p,i}} \end{aligned} \tag{5}$$

All normalized trajectories are resampled with the same number of time steps. Independent from the chosen linear weights for individual elements, these normalized trajectory segments can be concatenated without inducing discontinuities in the resulting trajectories.

In addition to the normalized trajectory segments, also the start and end positions of the individual movement primitives ($x_p(t_{p,i})$ and $x_p(t_{p,i+1})$), and their temporal durations are linearly combined. The linearly combined normalized trajectories are finally "unnormalized" by adding a piecewise linear function, which is defined by the linearly combined start and endpoints, and by rescaling of the durations of the individual movement primitives to match the linearly combined element durations. (See [Giese et al. 2002] for further details.)

3 Time-continuous Motion Morphing

The previously described method assumes that the style properties of the movements, and thus the weights of the linear combinations remain constant during the individual movement primitives. Even though this method has been applied successfully for the analysis of complex movements, e.g. in sports [Ilg et al. 2003], and for the synthesis of movements in robots [Ilg et al. 2004], it produces not always optimal results in the context of character animation. In some situations the proposed concatenation method induces foot sliding, which is not acceptable for high-quality animations. We present in the following an extension of the method that avoids this problem. This extension is based on the introduction of *time-dependent weights* for the linear combination. When the weight vector changes continuously with time, discontinuities at the transition points between different movements primitives can be avoided, even for non-normalized trajectories.

3.1 Time-dependent Weights

Replacing the constant weights w_p in (3) by time-dependent weights $w_p(t)$, results in the linearly combined trajectory:

$$x(t) = x_r\left(t + \sum_{p=1}^{N} w_p(t)\, T_p(t)\right) + \sum_{p=1}^{N} w_p(t)\, X_p(t) \tag{6}$$

When the weights change over time, this simple extension often results in artifacts, like foot sliding, in particular in presence of substantial spatial deviations between the prototype trajectories. This can be easily understood: Assuming that the prototype trajectories are not time-warped relative to each other, one obtains with $T_p(t) \equiv 0$ for $1 \le p \le N$ for the velocity of the synthesized trajectory

$$\dot{x}(t) = \dot{x}_r(t) + \sum_{p=1}^{N} \left[\dot{w}_p(t) X_p(t) + w_p(t) \dot{X}_p(t)\right].$$

This implies that in presence of dynamic weight changes ($\dot{w}_p \neq 0$) an additional velocity component is added that scales with the size of the spatial displacements. This problem can be avoided by redefining the morph in a way that ensures that the velocity of the linearly combined trajectory does not depend on the temporal derivative of the weight vector $w(t)$. This is condition is fulfilled if the modified morph $x^C(t)$ satisfies

$$\dot{x}^C(t) = \dot{x}_r(t) + \sum_{p=1}^{N} w_p(t) \dot{X}_p(t).$$

For general time-warps $T_p(t) \neq 0$ the velocity of the morphed trajectory has to fulfill

$$\dot{x}^C(t) = \dot{x}_r(t + \sum_{p=1}^{N} w_p(t) T_p(t)) + \sum_{p=1}^{N} w_p(t) \dot{X}_p(t). \tag{7}$$

The morphed trajectory can be computed by integration of this velocity over time with the initial position $x_0^C = x^C(t_0)$

$$\begin{aligned} x^C(t) &= \int_{t_0}^{t} \dot{x}^C(u) du \, + \, x_0^C \\ &= x_r(t + T(t)) + X^C(t) + \, x_0^C \end{aligned} \tag{8}$$

with the linearly combined temporal displacement according to (3) and the modified spatial displacement

$$X^C(t) = \int_{t_0}^{t} \sum_{p=1}^{N} w_p(u) \dot{X}_p(u) \, du.$$

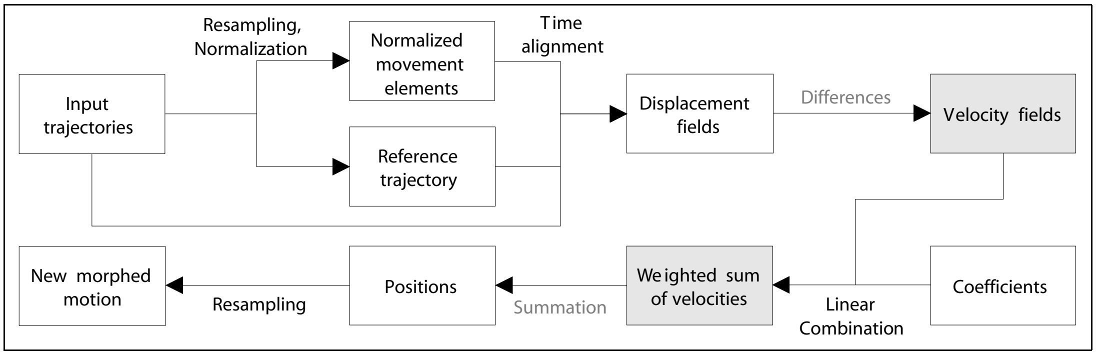

Figure 2: Overview of the motion morphing algorithm with time-continuous weights. Gray elements indicate the steps that have been added to the HSTMM algorithm discussed in section 2.

The starting point is given by

$$x_0^C = \sum_{p=1}^{N} w_p(t_0)x_p(t_0). \tag{9}$$

In the general case, this morphing method requires thus the differentiation of the spatial displacements $X_p(t)$, a time warping of the velocity of the reference trajectory x_r, and an additional integration. The steps of the modified morphing algorithm are illustrated 2. (The elements indicated in gray show the new steps that have been introduced for the algorithm with time-continuous weights.)

If the time-dependent weights fulfill the assumption $\sum_p w_p(t) = 1$ the above algorithm can be further simplified. In this case, exploiting the relationship $\dot{X}_p(t) = \frac{d}{dt}\left(x_p(t) - x_r(t)\right)$ equation (7) can be rewritten

$$\dot{x}^C(t) = \dot{x}_r(t + T(t)) - \dot{x}_r(t) + \sum_{p=1}^{N} w_p(t)\dot{x}_p(t). \tag{10}$$

In this case, the linear combination can be directly computed from the velocities of the prototype trajectories. If in addition, the linearly combined time-warping function is zero ($T(t) \equiv 0$) the last equation simplifies to a simple linear averaging of the prototype velocity in each time step.

3.2 STMM with Time-continuous Weights

Based on the steps discussed in section 2 we can now define the complete algorithm for time-continuous weights. As illustrated in figure 2, the algorithm has the following steps:

1) Segmentation of movement primitives: As in section 2.3.

2) Normalization of the trajectory segments: As in section 2.4.

3) Computation of space-time correspondence for the primitives: As in section 2.1.

4) Linear combination / integration: Linear combination of the velocities of the prototype trajectories (using (10)), of the start positions (using (9)), and of the temporal durations of each primitive. Integration of the linearly combined trajectories according to (8).

5) Concatenation and rescaling of time: The synthesized trajectory segments for each primitive are concatenated and temporally resampled in order to adjust them to the linearly combined element durations.

The proposed new algorithm produces excellent results, even for long integration times above 30 s. In our implementation, we sampled each normalized primitive with 200 time steps. Resampling was accomplished by linear interpolation.

4 Validation of Physical Properties

We compare in the following different algorithms for character animation by linear combination of prototypical example trajectories. We tried to assess, on one hand, the compatibility of the generated trajectories with physical laws of motion using a mechanical stability criterion. On the other hand, we tried in psychophysical experiments that are described in section 5 to evaluate how different motion morphs are perceived by human observers.

4.1 Motion Capture and Animation

Our evaluation experiments were based on a sequence of complex martial arts movements, a karate kata (heian shodan). The kata contains 20 techniques that define the movement primitives. Two actors with different skill levels, a white and a black belt, were recorded executing the same movement sequence. Data was acquired using a motion capture system (VICON 612, Oxford) with 11 cameras and 41 passive markers. The sampling frequency was 120 Hz. The trajectories were pre-processed using commercial software provided by VICON. Animations were generated using 3D Studio MAX using a commercially available avatar model. An example animation sequence is shown in figure 3.

4.2 Zero-moment-point

In several recent studies the physical correctness of character animation has been quantified and optimized using measures that are based on the zero-moment-point (ZMP). Such criteria quantify whether the simulated character moves in a dynamically stable way, similar to real human movements. If a character is nonsliding and in contact with the ground, dynamic stability requires that the zero moment point (the point on the ground for which the sum of all moments acting on the character disappears) lies within the support polygon that is defined as the convex hull of all contact points of the feet. (Figure 4 illustrates this support polygon and the ZMP as sphere on the ground.) When the character is not moving the ZMP is identical with the projection of the center of mass onto the ground. A measure for the incompatibility of a character with this dynamic stability criterion is given by the distance between the ZMP (for an assumed typical mass distribution) and the support

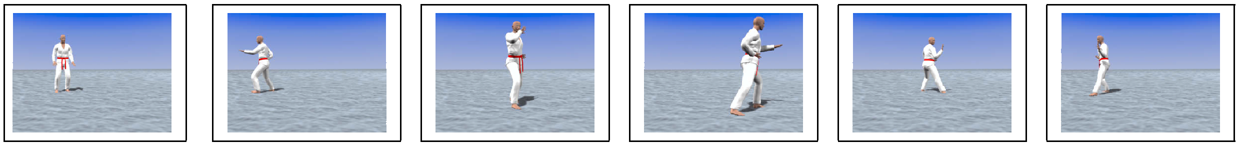

Figure 3: Example pictures from a morphed Karate sequence (duration 30s).

polygon during support phases. For our evaluation, we used a mass distribution and equations described in [Tak et al. 2000] for the calculation of the ZMP. We assumed a height of 1.75 m and a weight 79 kg of the character (similar to the values of the real actors).

4.3 Results

Figure 5 (a) shows the distance between the estimated ZMP and the support polygon for two motion-captured trajectories of a beginner and a master in karate. The dynamical stability criterion seems to be more frequently violated by the master than by the beginner. This is potentially consistent with the fact that skilled karateka are able to increase the power of individual techniques by temporarily giving up dynamic equilibrium (”leaning into the technique”) whereas beginners might be more focus on maintaining their mechanical stability.

Figure 5 (b) shows the distance between the estimated ZMP and the support polygon for trajectories that were synthesized with the two methods described in sections 2 and 3. The trajectories represent morphs between the beginner and the master. The style changed continuously from the beginner to the master in the course of the sequence. The distances of the ZMP from the support polygon have the same magnitudes as for the original human movements. This implies that both interpolation techniques do not induce significant violations of physical stability properties. In addition, the differences between the two morphing methods are minimal. The ZMP criterion seems thus not to be sensitive for foot sliding.

The relatively frequent violations of the ZMP criterion for real human movements raises the question in which situations physical criteria are really helpful for judging subtle differences between animations of complex human movements.

5 Psychophysical Validation

Beyond the application of physical criteria, we also conducted a psychophysical study to evaluate how naïve subjects and expert observers perceive the different animations. In this study, we compared five techniques for motion interpolation: STMM with and without time-warping, the new technique with time-variant weights, with and without time-warping, and a technique without time-warping that is based on PCA. For this purpose, the trajectories over all time steps and training examples were concatenated into a matrix, and this matrix was subject to a principle components analysis, resulting in ”eigen postures”. Morphing was accomplished by linearly combining the eigenvectors of derived from this analysis. In the psychophysical experiments, we used long animation sequences showing a karate kata, and short movies presenting single gait cycles from locomotion patterns. The rendered avatars and their virtual environment were kept simple to minimize the distraction of the subjects by other cues than motion. However, we did not present strongly impoverished stimuli, like stick figures or point light walkers, because such stimuli are not always suitable for judging animation quality [Hodgins et al. 1998].

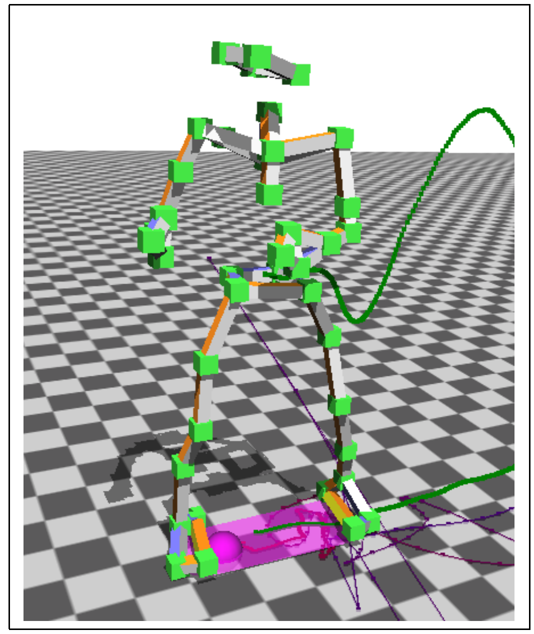

Figure 4: ZMP (sphere) and support polygon (magenta) compared to the center of mass (green trajectory).

5.1 Subjects and Procedure

We tested two groups of observers. The first group (5 females and 3 males; age 25 to 33 y) was completely naïve about computer graphics. All members of this group were payed for their participation. The other group (2 females and 9 males; age 27 to 29 y) consisted only of individuals working at least two years in the field of computer graphics. All these subjects had experience with character animation. From the 19 tested subjects 8 had previous training of at least one year in martial arts (including disciplines like judo, karate, and taekwondo). Both groups contained subjects with expertise in martial arts. Except reported otherwise we did not find significant influences of this variable. All subjects had normal or to-normal corrected vision. They were placed in front of a laptop (DELL Inspiron 8600, TFT screen with 1680 x 1050 pixels) at a distance of 40 cm in a dimly lit room.

Subjects were presented with two sets of animations. The first set showed a karate kata (20 techniques) generated by linear combination of the movements of a beginner and a master. During the movie the style changed continuously from beginner to master level. These movies had a duration of 30 s. The second set of movies showed prototypes and morphs (with equal weights of the prototypes) of walking and limping, presenting a single gait cycle. These movies had a duration of 2 s. Movies were presented always once and in the same order. Five different animation methods were tested: HSTMM with piecewise constant weights and time-continuous weights, with and without time-warping, and the PCA-based morphing technique. (Figure 3 shows an example sequence

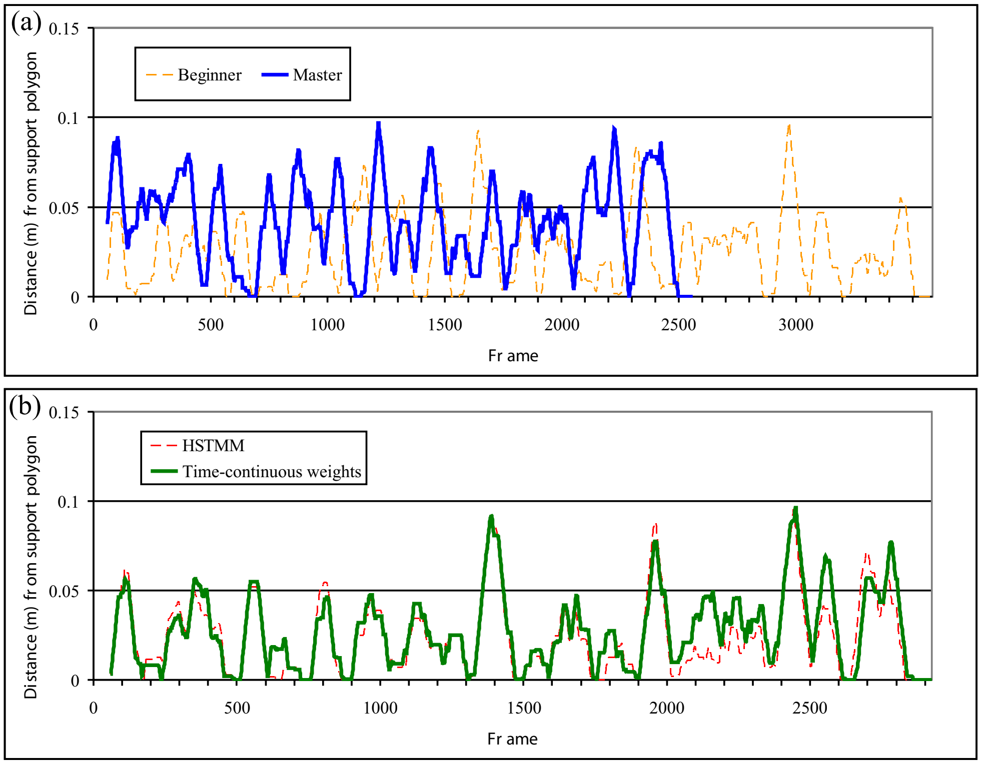

Figure 5: Distance of ZMP from support polygon (moving averages over 60 frames). (a) Results for real human movement trajectories from a beginner and a master in karate. (b) Results for motion morphs between these human trajectories generated with HSTMM piecewise constant weights, and with the new method with time-continuous weights.

from the long, and figure 6 examples from the short animations).

Subjects received a questionnaire for rating the avatars on 5 point scales along different dimensions. The first set of movies was rated according to the dimensions: 1) naturalness, 2) skill level at the beginning of the movie (white to black belt), 3) skill level at the end of the movie, 4) movement speed at the beginning of the movie, 5) speed at the end of the sequence, 6) animation quality (amateur level to Hollywood quality). The second set of movies was only rated along the dimensions naturalness, quality and walking vs. limping. Subjects were also instructed to give a free written report of specific artifacts that they noticed for each movie.

5.2 Results

The analysis of the free response data on *recognized artifacts* reveals a prominent result: experts seem much more skilled than naïve observers in the detection of errors in the animation sequences. About 40 % of the graphics experts reported sliding of the feet, whereas not a single naïve subject noticed this artifact. Sliding was observed, in particular, for the animations based on PCA, and for HSTMM with constant weights in presence of time-warping. Self-collisions, e.g. of the arms with the figure, were noticed by 40 % of the graphics experts, but only by a single naïve observer. 25 % of the naïve subjects recognized violations of kinematic constraints, compared to 50 % of the graphics experts. Few subjects (one or two) in both groups reported slow or jerky movements, and errors in the posture of the character.

In the following, first the results from the karate sequences will be discussed. The analysis of the *naturalness ratings* shows that the new morphing technique with time-variant weights and time warping produces morphs that are perceived as significantly more natural than the others. A two-way ANOVA with the factors Animation Type and Naïve vs. Experienced observers reveals a significant main effect of the first factor ($F(4,68) = 2.3$, $p < 0.1$), but no significant main effect and interaction with the computer graphics expertise ($p > 0.3$). A post-hoc analysis confirms that the new algorithm is perceived as more natural than the others ($p < 0.1$).

Subjects seem to be able to recognize the simulated skill levels. They perceive, on average, higher skill levels at the end of the movies than at the beginning. This is confirmed by a two-way ANOVA with the factors Animation Type and Beginning / End of the movies that reveals two significant main effects ($F(4,72) = 3.7$ resp. $F(1,72) = 5.2$, $p < 0.05$), but no interaction ($p > 0.5$). Subjects with expertise in martial arts tend to rate the skill level of the character lower than subjects without such expertise. This is confirmed by a three-way ANOVA with the additional factor Experience vs. No Experience in martial arts, which reveals a main effect of the last factor ($F(1,17) = 6$, $p < 0.05$), but no interactions ($p > 0.5$). In a similar way, subjects on average provide consistent ratings of the speed at the start and the end of the movies, indicated by a significant main effect of the factor Beginning / End of the sequence ($F(1,18) = 28.6$, $p < 0.01$). In addition, the perceived speed seems to depend significantly on the animation type (significant main effect of Animation Type ($F(4,72) = 9$, $p < 0.01$; but no interaction $p > 0.2$).

The results of the quality judgements are very mixed and do not reveal significant differences between the different animation methods. This result might reflect the problem that 'quality' is not a sufficiently constrained perceptual dimension. Subjects might have

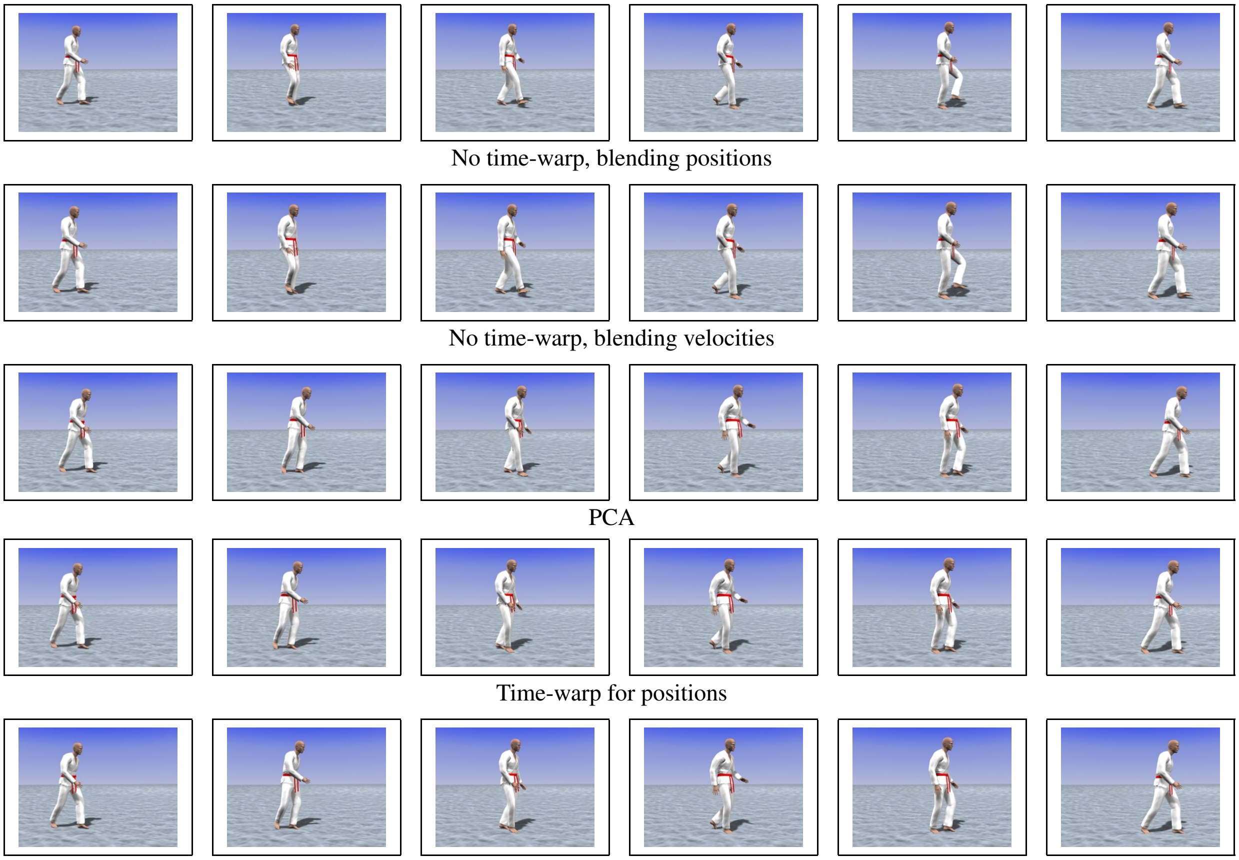

Figure 6: Morph sequences between walking-limping generated with different techniques.

included also the quality of non motion-related properties, like the avatar model or cloth simulation, in their judgements.

For the short movies showing locomotion patterns, the prototypes (walking and limping) were perceived as significantly more natural than the morphs. Methods including time-warping tend to produce morphs that are perceived as less natural than methods without time-warping (significant main effect of Animation Type, $F(6,102) = 4.8$, $p < 0.01$, but no influence of graphics expertise, $p > 0.3$). The ratings of animation quality are consistent with this result. Not surprisingly, the subjects give highly reliable ratings for the dimension locomotion type (scale walking to limping) with intermediate ratings for the morphs (significant main effect of Animation Type, $F(6,96) = 41$, $p < 0.01$ but no influence of graphics expertise, $p > 0.3$).

Summarizing, these results indicate an advantage of the correspondence-based methods, and in particular for the technique with time-continuous weights for the martial arts sequences. For the morphs between walking and limping, time-warping seems to reduce the animation quality. In addition, our animations conveyed reliable information about skill level.

6 Conclusions

We have presented a comparison between different techniques for the synthesis of trajectories for computer animation by linear combination of example movement sequences with different styles. All tested methods are suitable for integrating multiple movement styles within the same sequence. In addition, we have introduced a new technique for motion morphing with time-continuous weights that minimizes foot skating. The proposed method defines linear combinations based on velocities, and derives the position trajectories by subsequent temporal integration. The animation quality of several methods was tested in two different ways: (1) by evaluation of the compatibility of the movements with a mechanical stability criterion for articulated figures that is based on the zero moment point (ZMP); and (2) by a psychophysical study with naïve and expert observers who had to rate generated animation sequences along different perceptual dimensions.

The comparison of different interpolation methods using the mechanical criterion based on the ZMP did not reveal substantial differences between the methods, and between the morphs and real human movements. This result might be caused by a variety of reasons. The assumed mass distributions and mechanical models might not be accurate enough. Alternatively, in certain situations, real humans might violate the ZMP criterion. In such cases dynamic stability measure might not capture the relevant properties of realistic human movements.

Our psychophysical evaluation seems to support the advantage of the novel motion morphing method with time-continuous weights. However, some of the differences in our rating experiment were relatively small. In particular, the rating of animation quality might not be a well-defined perceptual measure, explaining the relatively inconsistent results for this perceptual dimension. However, we found consistent variations with respect to the perceived skill levels and movement speeds. Yet, the psychophysical measures seemed to be more sensitive and appropriate for quantifying slight quality differences between different animation techniques than the tested physical criterion.

In addition, we found interesting differences between naïve and expert observers. Experts in computer graphics seem to be very sensitive for artifacts in animated characters, whereas some artifacts were never be picked up by naïve observers. Experiments of this type might help to decide about aspects of computer animation methods that might be worth optimizing, in particular for applications targeting users outside the computer graphics community.

Acknowledgements

We thank W. Strasser for his support and interesting discussions. We thank Bernhard Eberhardt for his support during the motion capture, and Lucas Kovar and Hyun Joon Shin for the helpful discussions. This work is supported by the Volkswagen Foundation, DFG, and the Human Frontier Science Program. Additional support form the Max Planck Institute for Biological Cybernetics is gratefully acknowledged. M.G. is visiting research fellow at the Department of Biomedical Engineering, Boston University, USA.

References

ARIKAN, O., FORSYTH, D. A., AND O'BRIEN, J. F. 2003. Motion synthesis from annotations. In *Proceedings of ACM SIGGRAPH*.

BRAND, M., AND HERZMAN, A. 2000. Style machines. In *Proceedings of ACM SIGGRAPH*, 183–192.

BRUDERLIN, A., AND WILLIAMS, L. 1995. Motion signal processing. In *Proceedings of ACM SIGGRAPH*.

CHOI, M. G., LEE, J., AND SHIN, S. Y. 2003. Planning biped locomotion using motion capture data and probabilistic roadmaps. *ACM Transactions on Graphics 22*, 2.

FALOUTSOS, P., VAN DE PANNE, M., AND TERZOPOULOS, D. 2001. Composable controllers for physics-based character animation. In *Proceedings of ACM SIGGRAPH*.

FANG, A. C., AND POLLARD, N. S. 2003. Efficient synthesis of physically valid human motion. In *Proceedings of ACM SIGGRAPH*.

GIESE, M. A., AND POGGIO, T. 2000. Morphable models for motion patterns. *International Journal of Computer Vision 38*, 59–73.

GIESE, M., KNAPPMEYER, B., AND BÜLTHOFF, H. 2002. Automatic synthesis of sequences of human movements by linear combination of learned example patterns. In *Biologically motivated Computer Vision*, 538–547.

GLEICHER, M. 1997. Motion editing with spacetime constraints. In *Proceedings of the Symposium on Interactive 3D Graphics*.

GRASSIA, S. 1998. Practical parameterization of rotations using the exponential map. *The Journal of Graphics Tools 3*, 3.

HODGINS, J. K., O'BRIEN, J. F., AND TUMBLIN, J. 1998. Perception of human motion with different geometric models. *IEEE Transactions on Visualization and Computer Graphics 4*, 4, 307–316.

ILG, W., MEZGER, J., AND GIESE, M. 2003. Estimation of skill levels in sports based on hierarchical spatio-temporal correspondences. In *Proccedings of the 25th DAGM Pattern Recognition Symposium*.

ILG, W., BAKIR, G. H., MEZGER, J., AND GIESE, M. 2004. On the representation, learning and transfer of spatio-temporal movement characteristics. *International Journal of Humanoid Robotics 1*, 613–636.

KOVAR, L., AND GLEICHER, M. 2003. Flexible automatic motion blending with registration curves. In *Proceedings of Symposium on Computer Animation*, ACM SIGGRAPH.

KOVAR, L., AND GLEICHER, M. 2004. Automated extraction and parameterization of motions in large data sets. In *Proceedings of ACM SIGGRAPH*.

LEE, J., AND SHIN, S. Y. 1999. A hierarchical approach to interactive motion editing for human-like figures. In *Proceedings of ACM SIGGRAPH*, 39–48.

LEE, J., AND SHIN, S. 2002. General construction of timedomain filters for orientation data. *IEEE Transactions on Visualization and Computer Graphics 8*, 22.

LI, Y., WANG, T., AND SHUM, H. 2002. Motion texture: A two-level statistical model for character motion synthesis. In *Proceedings of ACM SIGGRAPH*.

LIM, I. S., AND THALMANN, D. 2001. A vector-space representation of motion data for example-based motion synthesis. In *Deformable Avatars*, 169–179.

PARK, S. I., SHIN, H. J., AND SHIN, S. Y. 2002. On-line locomotion generation based on motion blending. In *Proceedings of Symposium on Computer Animation*, ACM SIGGRAPH.

REITSMA, P. S. A., AND POLLARD, N. S. 2003. Perceptual metrics for character animation: Sensitivity to errors in ballistic motion. *ACM Transactions on Graphics 22*, 3, 537–542.

ROSE, C., COHEN, M. F., AND BODENHEIMER, B. 1998. Verbs and adverbs: Multidimensional motion interpolation. *IEEE Computer Graphics and Applications 18*, 5.

SHIN, H. J., KOVAR, L., AND GLEICHER, M. 2003. Physical touch-up of human motions. In *Proceedings of Pacific Graphics*.

TAK, S., AND KO, H. 2003. A physically-based motion retargeting filter. In *Proceedings of ACM SIGGRAPH*.

TAK, S., SONG, O., AND KO, H. 2000. Motion balance filtering. *Computer Graphics Forum 19*, 3.

UNUMA, M., ANJYO, K., AND TAKEUCHI, R. 1995. Fourier principles for emotion-based human figure animation. 91–96.

WANG, J., AND BODENHEIMER, B. 2003. An evaluation of a cost metric for selecting transitions between motion segments. In *Proceedings of Symposium on Computer Animation*, ACM SIGGRAPH, 232–238.

WANG, J., AND BODENHEIMER, B. 2004. Computing the duration of motion transitions: An empirical approach. In *Proceedings of Symposium on Computer Animation*, ACM SIGGRAPH, 335–344.

WITKIN, A., AND KASS, M. 1988. Spacetime constraints. *Computer Graphics 22*, 4.

WITKIN, A., AND POPOVIĆ, Z. 1995. Motion warping. In *Proceedings of ACM SIGGRAPH*.