A Literature Review on Virtual Character Assessment

Author:
Elena Alexandra URSU

Supervisor:
Dr. Nicolas PRONOST

November 2012

UTRECHT UNIVERSITY

Universiteit Utrecht

Abstract

Validating character animation techniques has traditionally relied on human observers. However, more and more researchers have started to develop methods to automatically and reliably assess their results. This review tries to provide those interested a consistent selection of publications in the field that have addressed the issue of virtual character assessment, structured in a meaningful manner. At the end of the review, we identify two directions in virtual character assessment: proposing evaluation metrics and procedures against other methods or ground truth data or observing thresholds and patterns to bring up guidelines for developers who want to achieve a specific tradeoff between naturalness and computational requirements.

Contents

1 | Introduction

Character animation is used nowadays in various applications, for example: entertainment (movies, games), medicine (treatment evaluations on musculoskeletal models), civil engineering (safety and security in crowds behavior), and learning (interaction, sports, skills). Traditionally, a lot of work in this field is done by animators and artists. Researchers have developed numerous techniques during the last two decades to come to their assistance. The most popular means of validating these techniques was by asking human observers their opinion on the outcome. However, more and more researchers have started to develop methods to assess their results automatically and reliably. This literature review will include a number of publications in the field that have specifically addressed the issue of virtual character assessment, trying to provide unified frameworks for comparison across different methods.

We try to provide those interested a consistent selection of references structured in a meaningful manner. The reviewed papers were chosen both to be relevant and recent. We have structured the contents in seven chapters by trying to identify the main directions that have recently emerged from the wide topic of computer animation. From the assessment point of view, we would like to put forward a classification based on the following points:

- Works that use *visual perception* either as a means to provide assessment guidelines on naturalness and plausibility, in Chapter 2, or as a means to classify input data for further processing, in Sections 5.3, 5.4.

- Works that study *motion feasibility*, either by looking at kinematics, in Sections 6.1, or at the compliance with the laws of physics, in Chapter 7

- Works that formulate *mathematical rules and models* to describe animation, and assess the outcome in terms of how it complies with these rules and models in Chapter 4

- Works that *learn the models* that characterize different motions from real life data, and compare the outcome against ground-truth in Chapter 3, Sections 5.1, 5.2, 6.2, 6.3, and Chapter 8

For articles that proposed a method and evaluated it towards the end, the typical discussion flow in this review is first presenting the approach, then the means of evaluation and lastly the experimental results. For articles that directly address virtual character assessment, we present the approach and the results. Where possible, we try to comparatively present papers that focus on a similar purpose or use the same background, but decide to take different approaches.

At the end of this work, we will extract those papers that we considered to provide clear-cut assessment metrics or guidelines.

2 | Motion Naturalness

[Ren et al., 2005] distinguish three different approaches for quantifying natural motion. The first one is searching for the thresholds of perceptual plausibility, or when are humans able to perceive unnaturalness in virtual characters. We will review the works of [Vicovaro et al., 2012] and [Hoyet et al., 2012] that fall into this approach, in 2.1. The second one is proposing a set of heuristic rules (according to the laws of physics) to govern joint movement. We will review some of these approaches in 7. The third approach consists of employing learning algorithms, which can automatically determine if a motion looks natural, based on ground-truth data. [Ren et al., 2005] proposed an alternative to this approach that we will discuss next.

They built a model for natural motion that captured probabilistic dependencies between features across time. First, they selected a statistical model for data variation in time from three standard techniques: mixtures of Gaussians (MoG), hidden Markov models (HMM) and switching linear dynamic systems (SLDS).

The statistical models were formulated as ensembles of statistical models, each accounting for modeling dependencies between joints at different levels: at low level (joints) - 8D feature vectors consisting of joint angles and velocities for each of the body joints and one feature vector for the root consisting of the linear and angular velocity. At middle level (limbs) - previous features were grouped for each limb, to represent the aggregate motion of body parts. At top level, the full body pose (as joints rotation angles) comprised the last feature vector.

Each ensemble statistical model was associated with a set of parameters and a likelihood function which showed the probability of it generating an input motion. These parameters θ_i were used to compute a naturalness measure for a motion as $s_i = \frac{\log P(D|\theta_i)}{T}$, where D was the motion sequence and T its length.

To exemplify, the likelihood function, namely the distribution of the body poses and velocities, was represented with a mixture of Gaussians in the HMM. The parameters θ_i of the HMM included mixture weights for each hidden state and the mean vectors and covariance matrices of the Gaussians.

Then, the model parameters were fitted using natural human motions as training data (over one thousand trials consisting of locomotions, physical activities, environment interactions, subjects interacting and other common scenarios), by calculating the mean μ_i and standard deviation σ_i.

Last, each new input motion was attributed a score that measured its naturalness as $s = \min_i \left(\frac{(s_i - \mu_i)}{\sigma_i} \right)$, with i iterating through all the feature vectors.

In the experimental phase, the three statistical models were trained on the motion database and a number of natural and unnatural motions were tested against the method. The unnatural motions were obtained in a number of ways: by editing motion capture sequences in Maya, by keyframing motions by an experienced animator, by introducing noise, by introducing bad transition according to a commonly accepted metric and by using insufficiently cleaned motion capture data. The best performance was achieved by SLDS with 82% correct classification of natural motions and 84% correct classification of unnatural motions, then by HMM and MoG.

A user study was also conducted. Participants were asked to watch approximately half of the motions used in the previous experiment and decide their naturalness by answering with yes or no. The motions

were displayed in random order, for each of the participants. Results showed that the judgment of the human subjects outperformed all the statistical models, leaving room for improvement and allowing for a better insight of where the models performed poorly.

2.1 Perceptual Plausibility

[Hoyet et al., 2012] studied the level to which humans were sensitive to three types of anomalies in virtual character interactions (pushing, in this case): timing errors, force mismatches and angular distortions.

A first set of experiments, called baseline experiments, was designed to reveal whether participants could perceive five different force levels. The experiments showed that not only the five different force levels were distinguishable, but that also viewing only the character that pushed or the character that was being pushed was enough to classify the force intensity. The character that was being pushed, or the target, was shown to convey the most reliable cues in this respect.

Next, the three anomalies mentioned above were looked into separately. To study the timing errors, the motions were altered so that the reaction of the target would be early or late with respect to the original contact time. Both cases were found to be perceived in an equal manner, while timing errors of over 150 ms were found to be acceptable in less than 50% of the trials.

To test the effect of force mismatches on perceived naturalness of the pushing motion, two target motions and one source motion were selected for every force level and every push direction (seven in total). The motions were altered by introducing mismatching between the target and source force of zero to (plus/minus) four levels. It was found that the alterations were perceivable and that over-reactions were more tolerated than under-reactions.

Angular distortions were applied in four different steps from 0° to 67°, applied from clockwise directions on the right side of the target and counterclockwise directions on the left side of the target. One source motion and one target motion for each of the seven captured directions were displayed. It was found that the larger the distortion, the motions became less acceptable.

This study highlighted the importance of anomalies present in pushing motions over plausibility of the motions, and by detailed experiments provided a number of guidelines regarding the extent to which the respective anomalies were acceptable.

[Vicovaro et al., 2012] modified throwing motions and studied the thresholds at which human viewers could tell that the motions were modified. Experiments were designed according to the psychophysical approach and the staircase method [Cornsweet, 1962], which was a suitable procedure to identify thresholds by displaying the motions around the threshold for a particular observer. An ascending staircase meant that the displayed throws were natural at the beginning, then modified in small steps until the observers perceived them so. Then the inverse process took place, reducing the modifications until the motion looked natural again. This procedure was called "up-down". A descending staircase meant that the first throws were bluntly altered and then the anomalies were reduced until the motions looked natural. The same "up-down" procedure was applied. To avoid participants from anticipating the next throw, the trials from several staircases were interleaved, as to appear random.

Overarm and underarm throws of a tennis ball were displayed such as the viewers could clearly evaluate all the phases of the throws. The speed of the biological throwing motion and the ball release velocity were manipulated accordingly using dynamic time warping (DTW). The participants were asked which animations were natural or modified. The results showed that the subjects were more sensitive to slowing down throws than to speeding them up, especially for underarm throws. Also, it was indicated that DTW could be used to increase the throwing distance to a large extent by speeding up the throw, and to decrease the throwing distance of an underarm throw only by a small amount.

In another experiment, a physical mismatch was introduced by only modifying the ballistic motion. The preparatory motion remained unchanged. The horizontal and vertical components of the release velocity were altered in turn. It was found that the participants were sensitive to these physical mismatches, as

the preparatory motions provided enough information for the observers to anticipate the trajectory of the ball after release. Participants were more sensitive to modifications of the horizontal component in overarm throws and of the vertical component in underarm throws. An increase of the throwing distance of maximum 40% in underarm throws was found to be acceptable.

2.2   Rendering Acceleration Techniques

Another subject that we will review here is the level of naturalness in crowd simulation. Crowd simulation is a computationally expensive process, so efforts are being made to reduce the number of calculations, in order to achieve real time display rates. The most exploited area for this purpose is graphics. Research is being done on how to reduce the number of polygons sent to the graphics processing unit (GPU) to display crowds in real time, but to preserve realism. The main acceleration techniques for rendering are visibility culling methods, level-of-detail methods (LOD) and image-based rendering (IBR) [Tecchia et al., 2003, Rodriguez et al., 2010]. Some perceptual studies, [Hamill et al., 2005, McDonnell et al., 2005], have determined when and where these techniques are appropriate. We will review a number of articles that employ these techniques and the extent to which they preserve visual realism in the following.

To introduce IBR, we will briefly present the work of [Tecchia et al., 2003]. They used IBR to reduce the amount of rendered geometry in a large virtual environment populated with virtual humans. The principle of this method was to replace the polygonal representation of virtual humans with 2D images, called impostors, when the characters were far enough from the viewpoint. Impostors were precomputed in this case, by using a horizontally and vertically sampled hemisphere around the virtual character and by exploring the symmetry of the human body. In other words, 2D images of the character were taken with a camera whose field of view was restricted by the hemisphere sample. The best perspective to be displayed at runtime to match the viewpoint was chosen such as to minimize the popping effects. These effects were the main artifacts that appeared in this method, due to switching between perspectives.

To enhance realism, but to keep the number of models and of their respective impostors low, diversity among the characters present in the scene was ensured by coloring different parts of a model with different shades. Lighting and shadows were also addressed, to reduce artifacts and preserve visual fidelity.

Experimental results showed that the method was scalable with the number of virtual humans rendered in the scene. More, users reported that the visual quality was similar to normal polygonal models as long as they were not approaching the characters too much. However it was not mentioned what was the distance at which impostors played a convincing role.

To determine this distance threshold, [Hamill et al., 2005] conducted a perceptual study on impostor representations both for virtual humans and for buildings in large scene. They also evaluated how the model representation affected motion perception.

The experiments were designed according to the psychophysics approach, and implemented by using the staircase procedure. They set out to find the Point of Subjective Equality (PSE), that gave the threshold at which participants were able to tell two stimuli apart, and the Just Noticeable Difference (JND), that represented the smallest difference in intensity required for a subject to distinguish between two stimuli.

Virtual humans impostors were calculated similarly to [Tecchia et al., 2003]. Tests were carried on the authors' previous implementation of this technique in [Dobbyn et al., 2005], where the distance threshold at which impostors ought to be switched with the polygonal representations was formulated as a Pixel to Texel ratio (number of screen pixels occupied by an image).

First, the experimental results revealed that users were able to discriminate between impostors and geometric models when displayed side by side. However, they were not sensitive to small changes to the pixel to texel ratio at which the impostors were displayed. Next, switching between representation for a model facing the user and moving towards the screen was detectable when the distance was greater than the pixel to texel ratio of 1.4 : 1. Participants were also sensitive to subtle changes in the pixel to texel ratio at which the popping occurred.

Further, based on the belief that sensitivity to motion changes could be a good metric to evaluate visual fidelity, an experiment was designed to assess the effect of model representation on human motion perception. Arms, torso and legs motions were varied separately. Separate groups of participants viewed either the geometrical model, or the impostors. It was found that for leg motions, perceiving variation was similar in both cases. Perceiving small arm motion variations was easier when viewing impostors rather than the original models. Also, for the torso, participants noticed variation in motion faster with the impostor, than with the polygonal models.

With these results, the authors suggested that impostors were perceptually equivalent to geometrical models in the case of perception of human motion.

Building on the previous work, [McDonnell et al., 2005] brought into discussion the low level geometric meshes (LOD approach). Similar experiments were conducted to assess the efficiency of low level geometric meshes compared to that of the impostor representations.

The perception of motion test revealed that the motion variations were perceived similarly, regardless of the representation, but that the motion of the impostors was closer to that of the original model than that of the low resolution model.

The pixel to texel ratio for distinguishing between the impostor and polygonal model was refined in a subsequent experiment and found to be closer to one-on-one, namely 1.164 : 1. A similar experiment was carried out for the LOD approach, to determine the percentage of vertices in the low resolution model that triggered the observable difference between this and the original representation, for three different distances. At the closest distance, a mesh containing 36.4% of the vertices was equivalent to the original one. At the same pixel to texel threshold for impostors, a mesh represented with approximately 27.5% of the vertices was equivalent to the high resolution mesh.

The results of this research provide valuable guidelines of when realism can be achieved by using rendering acceleration techniques.

The level of detail approach can also be explored at different levels, other than the geometric level, like it was done for example in [Rodriguez et al., 2010] who managed the level of detail also on skeletal and behavioral levels, complementary to a modified culling method.

Detail management at skeletal level aimed at reducing the topological complexity of a virtual character when far from the camera. Each skeletal node from the character scenegraph was assigned a distance beyond which its degrees of freedom would not be evaluated. Three levels of details were defined corresponding to nine, four and zero nodes.

Detail management at behavioral level was closely connected to the geometric and skeletal level. For example, at a distance that would allow the simplification of the model by incorporating the geometry of the head into that of the thorax, there would be no need to perform any visual activity management.

[Rodriguez et al., 2010] demonstrated the utility of their method on a simulation of a town and its surroundings with up to 6000 characters displayed at 25 frames per second.

In contrast to previous work which used discrete solutions for geometry management, such as [Dobbyn et al., 2005] who used impostors beyond a certain distance and [Rodriguez et al., 2010] who used three skeletal levels of detail, [Ramos et al., 2012] proposed an approach based on continuous level of detail combined with mesh instancing and hardware palette skinning.

The advantages of this approach were the following: continuous level of detail offered better granularity, by exactly specifying how many polygons needed to be rendered, mesh instancing allowed applications to render a mesh multiple times in different positions with a single draw call, and hardware palette skinning codified all animation information on a texture.

The level of detail to display an entire character was decided based on the distance to the camera (no specific details were provided about this criterion). A performance analysis was conducted by measuring the triangle throughput when varying the number of characters in the scene with constant visual quality (as determined by the previously mentioned condition for the level of detail). This analysis showed that the continuous model of LOD could offer a balance between performance and perceived visual quality.

Motivated by the findings of [Hamill et al., 2005], [Yuksel et al., 2012] proposed an impostor rendering method using image morphing techniques. These techniques aimed at cutting the number of intermediate frames of animation in half and recreating them in the rendering phase, in order to save space from the texture memory of the GPU that could be used to enhance visual quality by different means.

The visual performance of the image morphing algorithm was statistically found to provide close results to the original images. A user study showed that the method provided 38% smoother animations and increased the appearance quality by 87% compared to a reference application built in OpenGL and GLSL.

In conclusion of this chapter, we would like to specify that although evaluating the extent to which different rendering acceleration techniques achieve realism in the same measure as the impractical high resolution representation still relies on perceptual studies, we can distinguish an obvious concern towards providing guidelines as to how these techniques perform. Yet, we can point out another reason for which such guidelines are hard to formulate in a robust manner, namely due to the nature of crowd simulation applications which heavily rely on ongoing hardware development that can make previous findings inapplicable to current implementations.

3 | Gestures

Gesture recognition is being investigated in various research fields (computer animation, computer vision, human-computer interaction) for the wide range of applications that it generates in real life (games, mobile interactions, surveillance, etc). Recognizing the gestures performed by a virtual character can help a sports trainer improve the student's performance [Majoe et al., 2009a], or mediate the interaction between a virtual character and a real person in a theatrical representation [Billon et al., 2008]. Hand gestures are given a particular attention and remain the basic input material for this research [Martin et al., 2010], [Oshita and Matsunaga, 2010], [Głomb et al., 2011].

Most commonly, gesture recognition consists of a feature extraction phase, a dimensionality reduction and training phase, and a classification phase. In the following, we will review a number of papers grouped on the methods used for the matching or classification phase.

3.1 Hidden Markov Model

The Hidden Markov Model is a popular technique for recognizing gestures. The model represents a network of nodes that produce symbols and are interconnected by transition represented by probabilities, hence the need to reduce the input signal to a discrete series of symbols [Oshita and Matsunaga, 2010].

[Majoe et al., 2009a,b] used HMM to classify Tai Chi movements. Their initial endeavor in this area was presented in [Kunze et al., 2006]. They studied the feasibility of using low-cost motion capture equipment such as gyroscopes and accelerometers worn on the body for whole-body dynamic movements' recognition. By placing sensors at the upper arms, lower legs, knees, neck and rear hip, the authors collected data from two Tai Chi experts and two amateurs performing three specific movements.

By analyzing the raw data, it was found that the signals collected from the experts showed peaks of approximately the same length, smoothness and periodicity, corresponding to consistency in execution. On the other hand, the amateurs showed a higher number of events where the absolute sum of foot gyroscopes was close to zero, corresponding to pauses and jerky movements. Also, the experts showed faster neck and hip movements. Next, the energy required to perform the exercises was examined. The squared angular velocity, ω^2 in $(rad/s)^2$, indicated the rotational energy, defined by $E_{rotational} = \frac{1}{2}I\omega^2$. It was found that the experts had the least rotational energy consumption.

Using a sliding window over the data, two features were analyzed closely, the 75th percentile and the frequency range power of the accelerometer x-axis at the neck. These showed clear separation between the amateurs and the experts. KNN clustering on these features resulted in a 76% correct classification using cross validation, while adding the RMS to a second KNN classifier resulted in an improved 85% correct classification.

Finding that the used sensors provided a good base for capturing data in this scope, [Majoe et al., 2009a,b] used forward kinematics to create a 3D avatar that provided positional data for the body. To use HMM for recognition, it was necessary to translate the sequences into observation codes, these being the features. In an attempt to see whether observing the action correlation (between two limbs for example) would improve the recognition process or not, three different approaches were designed for

features derivation in three dimensions: first the angle between any two limb end and the torso, second the vectors between any two limb ends and third, the 3D trajectory data of any limb. Combining the limbs, nine total feature extraction methods were tested. To keep the number of codes to a small value (12 in this case), k-means clustering was used to cluster the training data before feature generation. The corresponding HMMs were trained on five different Tai Chi movements. The best recognition rate, 99.7% was obtained for the two dimensional features.

[Xiang et al., 2006] used Isomap for dimensionality reduction and HMM for classification. A 16 joint skeleton was used and an initial 48 dimensional feature vector was generated from motion capture data. The eight bones that constituted the limbs and connected the root and chest were chosen as additional features, represented by angles and position. Further, joint velocities were added, resulting in a 72 dimensional feature vector. Driven by the fact that Isomap could find a meaningful low-dimensional structure behind the original observations, the authors utilized this approach to reduce dimensionality and obtained a 7 low-dimensional space with minimal residual error.

Next AdaBoost was used to combine an ensemble of HMMs (a set of classifiers whose decisions are combined). Experiments were run on more than 1000 motion clips including typical motions (walking, running, jumping, etc) and the best recognition rate, 93.2% was obtained for running.

[Głomb et al., 2011] also used HMMs, combined with Vector Quantization (VQ). Regarding human motion as a time sequence of vectors or features, a quantization function was used to transform the sequence into a series of discrete symbols. To first derive the number of symbols, the discrete symbol series were treated as strings and the Levenshtein distance (equal to the smallest number of deletions, insertions and reversals that will transform a string into the other [Levenshtein, 1966]) between them was calculated. Both distance between strings from the same gestures class and distance between strings from different gesture classes were computed. For each of the two distance types, normalized histograms h_W (within-class) and h_B (between-class) were computed. Then, the Bhattacharyya distance [Kailath, 1967] between the histograms (calculated as $d_B(h_B, h_W) = -\ln \sum_{i \in I} \sqrt{h_B(i)h_w(i)}$) was maximized in order to find the number of symbols that achieved the best separation between classes.

Next, the authors proposed a procedure to improve initial values for the HMM parameters, based on K-means clustering. The Baum-Welch algorithm [Baum et al., 1970] subsequently improved these parameters. The experiments showed that maximizing the Bhattacharyya distance to obtain the number of symbols gave the best results. Additionally, a guideline for choosing the number of states for the HMM as under 30 was put forward.

3.2   Dynamic Time Warping

Dynamic time warping (DTW) is used to compare similarity between two time-dependent sequences which can be discrete signals or feature sequences sampled at equidistant points in time [Müller, 2007]. In this section we will review a number of papers that use DTW to compute the similarities between feature sequences that correspond to different gestures.

[Billon et al., 2008] focused on recognizing gestures in a real-time flow of actions. The starting point was defining the gesture as a variation (in this case, angle or acceleration variation) between two rest states. Compared to the usual approach we have seen so far - recording multiple executions of various gestures, reducing dimensionality and classification for matching - an interesting idea of this research was to design a multiagent sytem where each agent (called an Observer) represented a gesture and was triggered when it started to recognize itself in the flow.

PCA was used similarly to feature selection to select a subset of relevant features, in the form of a projection matrix that allowed the mapping of the gesture in 2D space, as a curve. Given a posture in time, an Observer checked its signature (the 2D curve). This was done for all Observers corresponding to all gestures at the same time.

To accurately choose the right Observer, three so-called fidelity values were calculated: the distance between curves using Dynamic Time Warping [Berndt and Clifford, 1996], the noise toleration at beginning and end of the observation and the difference in time.

An initial experiment aimed at recognizing the random replays (played by an animation engine) of 22 pointing gestures. All gestures were recognized. For the second experiment, the Wii Remote was used to record seven participants executing four different gestures thrice. The first evaluation of the experiment aimed at comparing one execution of the same gesture to the other two and found 61% recognized gestures, meaning that humans' repetition of the same gesture was not accurate enough for the system. The second evaluation looked at recognizing all the gestures performed by one person and resulted in a better recognition rate of 83%. The third evaluation compared four executions for each gesture to all the other executions of all gestures and showed a recognition rate of 86%. The average recognition time was 0.7 seconds before the gesture ended.

[Billon et al., 2010] improved the previous work by investigating the overlapping between the ending and beginning phases of two consecutive gestures and by restricting a following set of gesture for a recognized gesture. This way, false positives were excluded from the recognition process.

The experiments were also improved by testing two different databases of gestures. The first one consisted of 8 closely similar gestures repeated a few times, leading to 22 motions. The second one consisted of 7 increasingly difficult to reproduce gestures repeated a few times to a total of 21 motions. The databases were tested separately, 100% recognition was achieved for the first one and only one unrecognized gesture for the second one. Recognition time was on average 0.5 seconds before the end of each gesture.

This research also resulted in a convincing public demonstration of a Capoeira fight between a real human and a virtual one.

[Martin et al., 2010] analyzed hand gestures using a combined approach of Vector Quantization (VQ) and DTW. Data was collected via a Vicon system with 11 markers on the hands. Nine features were selected for extraction and normalization, such as distance between thumb tip and index finger tip markers or angles between successive velocities.

VQ was used to map the feature vectors to codebook vectors. Distribution of these vectors in feature space was done using k-means clustering, self-organizing maps or growing neural gas. Then, for each feature vector, the closest codebook vector was computed, obtaining new feature sequences.

After that, DTW was used to compute the distances between sequences. Different distances between individual symbols were used for each of the three training procedures mentioned above: for k-means and GNG, the Euclidean distance between codebook vectors assigned to the two symbols, and the distance between two neurons assigned to the two symbols on the SOM grid for SOM.

In the experimentation phase, eleven every day actions were recorded to constitute the training data. The validation and test data were formed of seven long sequences containing new gestures or variations of the training set gestures. Each sequence was recorded three times, the first time for the validation data, and the next two times for the test data. To find the best feature sets (a combination of features that gave the best results), an evaluation algorithm was run over all possible combinations, but without VQ. Three feature sets were chosen and the test data recognition performance was noted for each of the three training approaches. The recognition rates were around $65 - 70\%$, with the best value of 72.23% obtained by SOM combined with one of the feature sets.

3.3   Recognition Paths

[Oshita and Matsunaga, 2010] developed a recognition model represented as a state machine, by applying SOM to all feature vectors from sample data, resulting into units that made up the states. This approach was based on the observation that each gesture could be decomposed in atomic actions that followed each other in a sequence. Each phase of the gesture was represented by a state and the continuity between the phases was modeled through transitions occurring with a given probability. Each state was linked to

the initial state as gestures could be interrupted any time (a feature that we haven't seen so far). Beside the state machine, the recognition model also contained an initial state and a recognition path.

The authors claimed to have made a step towards full automatization of the process, and illustrate this through a novel technique of automatically selecting the number of units of each state machine. For this, several state machines were built per gesture with varying number of units and the optimum one was chosen as the one that minimized the sum $e = e_1 + ke_2$, where e_1 was the error rate that correct inputs were not recognized and e_2 was the error rate that incorrect inputs were recognized. Even so, the second error rate was weighted by a parameter k, stated to be application dependent.

Now, that the states were obtained, all that was needed to finalize the state machine that modeled a gesture was to train an SVM to learn the transitions between states. Next, the recognition path was built as a series of states from input sample data.

Finally, in the experimentation phase, two hand gestures (a simple one and a complex one) were recorded ten times each, using Wii Remote controllers. The feature vectors consisted of 3D accelerations from each hand. Best recognition for both gestures resulted for a number of four states in the corresponding state machines. For the simple gesture the recognition rate was 98%, and for the complex gesture the recognition rate was 80%.

4 | Behavior in Crowds

Crowd simulation has been a prodigious subject that has generated numerous quality simulation techniques over the past years, pushing the boundaries of real time performance, the number of simulated agents and diversities of scenarios. Complementary to this work, a recent interest in automatically validating the quality of these simulations has emerged.

4.1 A Steering Benchmark

SteerBench was presented in [Singh et al., 2008] (under the name of "Watch Out!"), improved in [Singh et al., 2009] and refined in [Kapadia et al., 2011a]. The framework consisted of two major components. The first one was a benchmark suite of numerous steering scenarios, designed to cover a comprehensive range of real life situations, classified in five categories (simple scenarios, one-on-one interactions, agent-agent interactions including obstacles, group interactions and large scale scenarios). The second was comprised of a set of evaluation metrics designed to be customizable and independent from the steering algorithm, as well as of a method to compare the results between two steering algorithms.

The principal metrics chosen were the following: the number of unique collision events, time efficiency - which showed how fast an agent achieves its goal, and effort efficiency - measuring the optimality of total kinetic energy use that an agent required towards its goal. Other numerous detailed metrics concerned collision, turning, distance, speed and acceleration.

[Kapadia et al., 2011b] described a thorough analysis of the scenario space. Considering that the complete scenario space consisted of all possible scenarios that could be obtained by combining user defined parameters (namely environment size, obstacle discretization, number of agents and target speed of agents), the most trivial possibilities were filtered out by imposing constraints. For example, one of those constraints was making all the agents interact with a reference agent. This method reveals the most representative subspace of scenarios that were considered challenging for the tested steering algorithms.

The analysis also proposed new metrics to measure the performance of steering algorithms in that subspace, in terms of coverage and quality. Coverage gave the ratio of successfully handled scenarios by a steering algorithm with respect to a particular metric. The average quality was the average value of the respective metric over all the sampled scenarios. The authors showed that the representative subspace of scenarios was covered by a steering algorithm in a finite number of test case samples.

As the primary metrics put forward in [Singh et al., 2009] were found to be unintuitively weighted in the final score, [Kapadia et al., 2011a] proposed another measure based on the Principle of Least Efforts (also used by [Guy et al., 2010]). Furthermore, by exploring the results of [Kapadia et al., 2011b] as described above, three of the most challenging initial scenarios were maintained and other eight were further added.

[Kapadia et al., 2009] built an interactive framework that provided a predefined set of rules which could detect abnormal behaviors (considered actions of interest, as their presence or absence determined the quality of a simulation) like steering in a circle, deviating from the target, or unnatural oscillations. Users could combine different rules to build up a complex behavior, such as pick-pocketing.

As the authors have shown a consistent focus on developing this benchmark framework over the last few years, which has already been used to test new methods [Karamouzas et al., 2009], it is most likely to progress into a standard for steering algorithms evaluation.

4.2 Data Driven Evaluation

[Lerner et al., 2009] searched similar state characters in a simulated and in a real life crowd and compared their actions in the respective situations. The example set was selected from video input where the trajectories were tracked manually and the entries were checked for redundancy.

The state was formulated as the density of agents surrounding a particular character, while the action was given by the trajectory across a two second time window centered at the corresponding state (a "short-term" decision). The similarity function between the simulation state-action pair and the example set state-action pair consisted of a normalized combination of the distance between states and the similarity between actions. States were compared by calculating the difference of densities in the surrounding regions, and actions were evaluated by distances between trajectories. The obtained score raised the probability of an individual behavior to be either natural or curious. It could not be determined for sure that a behavior fell in one of these two categories as the authors themselves did not claim that their example set was exhaustive.

[Banerjee and Kraemer, 2011] recreated a real environment and populated it with virtual agents, then compared the behavior in the simulated with the real scenarios by checking the match between agents' distribution in the virtual region and the corresponding one in reality.

[Guy et al., 2010] developed a crowd simulation method based on the Principle of Least Efforts (PLE), namely on minimizing the total metabolic energy used up when walking on a path. Knowing the instantaneous power during walking as $P = e_s + e_w|v|^2$, where e_s was the cost of being alive and e_w captured the biomechanical efficiency of locomotion (constants per agent), they modeled the biomechanical energy as $E = m \int \left(e_s + e_w|v|^2 \right) dt$.

A thorough validation of this technique was performed against other popular simulation techniques, running on a number of known scenarios. The validation included an analytical comparison - by calculating the biomechanical energy of two agents exchanging position and then comparing it to an optimal value -, a numerical comparison between the total biomechanical energies spent by agents in complex scenarios where the optimal value was unknown, a quantitative comparison of their agents' responses against empirical observations from crowd studies and lastly resorted to visual inspection to evaluate the naturalness next to emergent behaviors.

[Xing et al., 2012] tackled the issue of real world data acquisition by employing Human Computation (basically getting a task done by volunteers without them realizing it) and developed an evacuation game to study the decisions real humans take when trying to escape from a room full of people. They found that when choosing from two routes of different sizes towards the same goal, the users will prefer the clearer route, once the shorter route is occupied by ten or more agents. The influence of the exit size on choosing the way out was also tested and the experiments revealed that even though the larger one is prefer, when more agents prefer the smaller one, the user will be pushed towards that one as well.

[Musse et al., 2012] build a 4D histogram from the vector $(x(t), y(t))$, where $(x(t), y(t), \dot{x}(t), \dot{y}(t))^T$ was a discrete time trajectory and the velocity vector was represented in polar coordinates. To deal with memory constraints, all the sizes were mapped to a different number of discretized bins $H\left(x^d, y^d, \theta^d, s^d\right)$. Up to this point, the histogram provided information over the global flow of the crowd. To investigate the subflows, the trajectories were clustered depending on their displacement vectors made up by their start and end points. This was a useful feature in case of environments that anticipated main flows.

Crowds comparison was performed considering different aspects. The similarity between global flows was given by a distance metric based on the Bhattacharyya coefficient applied directly to the two normalized

4D histograms. Relative spatial occupancy was given by considering the position information from the histograms, resulting in 2D histograms that contained the relative spatial occupancy. Agents' density could be easily deduced from these histograms knowing the total number of agents for each crowd. In a similar manner, the orientation distribution histogram could be determined from the 4D histogram and could be used to determine whether the two crowds have similar main orientations or not.

The algorithm was tested both on simulated crowds as well as on real life data. In both cases, the flows were simple and the compared crowds moved within the same environment. For the simulated experiments, variations were introduced in the comparison aspects previously mentioned. The visual inspection validated the experiment results showing that similar crowds obtained the expected similarity score, with respect to all the analyzed aspects of global flow, global orientation, spatial occupancy and speeds distribution.

The main limitation of this method was that the comparison aspects were evaluated on a time frame, so changes within that time frame could not be detected. However, this method provided a reliable solution towards comparing two crowds, which was a step towards assessing a simulation's naturalness when compared with real life data.

[Guy et al.] developed a measure of similarity between ground-truth data and simulated crowds, aimed at eliminating some of the problems of the previous works, such as the inability of density-based metrics [Lerner et al., 2009] to cope with sparse scenarios. Simulators were formulated as functions that produced an estimation of the crowd state at the next unit in time. In this context, an entropy metric was introduced to measure the size of the prediction error for a given simulator. The entropy metric was demonstrated to provide rankable results, namely unique scores for different simulators, to be discriminative, general, consistent in results across similar datasets and robust to noise existent in the data.

The metric was tested by running different simulation models (a rule-based steering approach, a social-forces model and a predictive planning approach) on a common set of scenarios. Visual results showed that indeed the simulators with the lowest entropy metric performed best in the given scenarios. A perceptual study revealed a strong correlation between the entropy metric and similarity between the real crowd and the simulated one as perceived by human subjects.

5 | Style

Several elements compose the human motion: the action, the cadence and the motion signature [Vasilescu, 2001a]. The human visual system can easily recognize the motion signature of a particular person, as a means of evolutionary adaptation. Isolating the motion signature can reveal details about physical attributes (such as gender, age, body structure), emotional state, or the individual's own style of moving. Computer animation and computer vision researchers have worked on various fashions of separating and parameterizing these attributes in order to recognize new motions as belonging to a known individual/set of attributes or to recognize the action performed by a known person.

5.1 Gender

In order to discriminate gender from walking motion, [Troje, 2002a] regarded the walking sequence as a time series of postures p, and first applied principal component analysis separately to the postures of each walker, with the purpose of capturing redundancy. They obtained the following representation $p = p_0 + \sum_i c_i p_i$, where p_0 was the average postures and p_i were the first four principal components accounting for more than 98% of the overall variance.

Observing that the temporal behavior of the first four components could be modeled with pure sine functions, each walk was fully described by the average posture, the first four eigenpostures, the fundamental frequency and the phases of the second, third and fourth principal components with respect to the first component.

PCA was computed again on this space (of dimensionality 229), finding that a linear discriminant function (classifier) could correctly separate male from female walkers in the existing dataset using between 4 and 14 components. Further, the authors investigated the capability of the classifier to generalize to unknown motions and eliminated one of the motions in the data set in turn and obtained the smallest classification error for a space dimensionality of 4. This showed that Fourier decomposition of walking data was a nearly optimal representation in terms of covering variance with the smallest number of components and was further used in [Troje, 2002b].

More, structural information was separated from dynamic information and different classifiers were trained based on the two. The separation was done considering that the average posture p_0 encoded structural information, while the eigenpostures contained dynamic information. The authors found that using only dynamic information for classification gave the better performance.

[Troje, 2002b] focused on obtaining a system that was accurate enough to extract both biologically and psychologically (emotional) attributes. The difference in this approach was that, to linearize the motion data, first the postures were decomposed as second order Fourier expansions (leading to a representation of a walking motion by $p_{(j,0)}$ called the average posture, and $p_{(j,1)}, p_{(j,2)}, p_{(j,3)}, p_{(j,4)}$, called characteristic postures and ω_j, the fundamental frequency). Second the space dimensionality was reduced, by applying PCA which lead to the following representation of a walker: $w_j = v_0 + \sum k_{ij} v_i$, where v_0 was the average walker and v_i were called the Eigenwalkers. The dimensionality was reduced from 226 to 15 (accounting for 80% of the overall variance).

To determine gender, a linear classifier was found as the best solution of the overdetermined linear system $cK = r$, where r_j was 1 for a male walker and -1 for a female one, and K is the matrix of k_{ij} coefficients from the previous equation. The vector c therefore generalized the gender attribute. Adding or subtracting this vector changed the appearance of a character from male to female.

Next, [Troje, 2008] used a similar method to [Troje, 2002b], with the slight difference that the data was matched in the frequency domain, after computing the Fourier transform. The results in [Troje, 2002a] concerning structural and dynamic data were reconfirmed, by observing that removing the structural information did not affect classification, while misclassification rose when dynamic information was removed from the data.

5.2   Individuality

[Vasilescu, 2001a] and [Vasilescu, 2001b] decomposed the motion in the action performed and motion signature (or style). Three different actions were performed in the styles of multiple subjects, and the data was processed in order to parameterize distinct styles, to recognize specific individuals and to synthesize new motion in the style of a particular person.

Motion capture was used to record three different actions (walking, ascending and descending stairs) from several subjects. For each person, the actions were averaged and represented as joint angles over time in dataset matrix D. The dataset matrix was decomposed as follows $D = \left(Z^{VT} P^T\right) A^T$ [1] , where Z was called the core matrix and contained basis motions independent of people and of actions, P was the people matrix which contained the invariance across actions for each person, A was the action matrix encoding different actions invariant across people and $S = \left(Z^{VT} P^T\right)^{VT}$ contained person-specific signatures. The unknown factors Z, P, and A were solved by applying the 2-mode vector analysis algorithm from n-mode component analysis [Kapteyn et al., 1986] in numerical statistics.

This decomposition facilitated the calculation of a new signature for a person for whom only some of the actions were known. As main result of this study, each motion was generated in all distinctive styles and was successfully compared to ground-truth mocap data.

Using the same mathematical basis (n-mode component analysis) for the approach, [Vasilescu, 2002] represented the motion capture data differently, namely in higher-order arrays or tensors instead of matrices. To decompose the tensors according to the same principle, the author employed a formalism called higher-order singular value decomposition.

This work answers more specifically the assessment problem that we investigate in this review through a recognition method, aimed at identifying both the person who performs a known action, as well as the action performed by a known person. Recognition was made possible by an existing mapping of the motions either into the space of people parameters or the space of action parameters. A projection of the motion onto one of these spaces was computed, and then a nearest neighbor recognition algorithm retrieved either the best matching signature or action.

5.3   Emotion

Using the method described in [Vasilescu, 2002], [Kobayashi and Ohya, 2006] set out to identify gait patterns corresponding to different emotions. The main differences were the following: first, instead of computing a signature matrix, they computed an emotion matrix, and second, instead of capturing several subjects performing three distinct actions, they recorded four professional actors (male and female) performing a gait under a specific state of mind (angry, disgust, fear, joy, sad, surprise).

---

[1]VT is a mathematical operator comprising of a transposition and a stacking of the matrix columns to obtain a column vector.

Taking advantage of the periodicity of the gait cycle, a wavelet analysis was performed to extract specific motion features for each emotion. The authors focused on the angles between upper and front arm, between upper arm and upper body and between lower and upper leg. The findings showed that all emotions displayed either some sort of periodicity or temporal peaks (e.g. surprise depicts a decaying temporal large peak in the angle between upper arm and upper body as a motion feature). In the light of these findings, it looks like the wavelet analysis could be used to identify motion features in new motions in order to assess whether they belong to a certain state of mind.

[Troje, 2008] and [Troje, 2002b] also looked at emotion classification in motion and used the same mechanism as for gender (see 5.1). Instead of the binary values used for female and male in the linear classifier calculation, averaged ratings from human observers were inputted in the vector r. These rating were obtained in an experiment where six observers rated walking sequences as being within the "nervous" and "relaxed" or the "happy" and "sad" range. The need to involve human observers in this experiment comes naturally as emotional states are perceptual notions, unlike gender.

5.4 Other Attributes

[Sigal et al., 2010] and [Livne et al., 2012] focused on learning models for different attributes (gender, emotional state, as well as weight and age) from a combination of partially labeled video and motion capture data. They also explored the biological cues that determine humans to rate a virtual character motion with respect to the previously mentioned attributes. The mathematical method for processing mocap data and experimental methods for the attributes rating by human observers were similar to the work of [Troje, 2008]. The rest of the works covered attributes inference by human observers from video-based 3D trackers output, which although an interesting study on motion style, does not directly concern our review.

6 | Kinematics

In this chapter we will review a number of papers that use kinematics for three different purposes: to calculate transition probabilities or costs in motion graphs (6.1), to analyze joint angles and velocities during gait (6.2), or to assess motion estimation and interactive control techniques (6.3).

6.1 Motion Graphs Transitions

Motion capture databases are used to aid animators in their task of creating new, realistic motions of human virtual characters. Motion graphs represent the ways in which recorded motions can be combined. Assessment of whether it is possible to combine two motions is given by a measure of resemblance, called a distance metric [van Basten and Egges, 2009]. We will discuss distance metrics that involve kinematics in the following.

[Lee et al., 2002] designed three interfaces through which a user could interactively control a virtual character, by choosing from a number of options every few seconds of where the avatar to go, or how to behave; by sketching on the terrain the path for the character to follow; or by performing in front of a video camera a motion for the character to reproduce. To enable this, a database of recorded motions was organized in such way as to comprise the transition possibilities from one motion to another, then clustered to optimize searching appropriately to each interface described above.

The first step was made possible by modeling the data as a first-order Markov process, where the transitions between two states depended only on that respective current state. Transitions were expressed as probabilities, given by an exponential function depending on the distance between two frames, and a σ term that controlled the mapping between the distance and the corresponding probability. The distance between two frames was computed as $D_{ij} = d\left(p_i, p_j\right) + \nu d\left(v_i, v_j\right)$, where $d\left(p_i, p_j\right)$ represented the weighted differences of joint angles, $d\left(v_i, v_j\right)$ represented the weighted differences of joint velocities and ν, a weighting term.

For a similar purpose - that of interactively controlling a virtual character -, [Arikan and Forsyth, 2002] also used the differences between joint positions and joint velocities to calculate the transition costs from one frame to another. In addition, they also used the difference between the torso velocities and accelerations. All were expressed in the torso coordinate frame.

[Wang and Bodenheimer, 2003] assessed the cost metric proposed by [Lee et al., 2002]. In the original cost metric, the distances between velocities were calculated as Euclidean distances, while for the positions, the following formula was employed: $d\left(p_i, p_j\right) = \left\|p_{(i,0)} - p_{(j,0)}\right\|^2 + \sum_{k=1}^{m} w_k \|log\left(q_{(j,k)}{}^{-1} q_{(i,k)}\right)\|^2$, where the first term represented the squared norm of the difference in global translational positions, and the second term represented the weighted sum of squared geodesic norms of the orientations of joint k in frames i and j in quaternion space. The original set of weights were one for shoulders, elbows, hips, knees, pelvis and spine, the other were zeros. [Wang and Bodenheimer, 2003] found the optimal set of weights, with the aid of an animation expert who examined good and bad transitions. For the ν term, they found that its influence was rather small and varied the term from 0 to 100 and found no significant differences, so they fixed its value at one.

A cross-validation study revealed that the optimal weights were robust and provided a good choice for transitions between a wide variety of motions. A user study confirmed that the optimal weights showed better and more natural results than the original weights.

Next, [van Basten and Egges, 2009] also investigated the efficiency of the cost metric proposed by [Lee et al., 2002] in comparison with two other metrics, proposed by [Kovar et al., 2002] (also used in [Zhao and Safonova, 2009]), based on point clouds and by [Egges et al., 2004] and by [Forbes and Fiume, 2005] based on principal components. Assessing [Lee et al., 2002]'s metric was done using the weights found by [Wang and Bodenheimer, 2003], as mentioned above.

Three aspects were assessed: foot skating, path deviation and on-line running time. The joint-angle metric performed best in terms of path deviation, as well as in running time.

6.2   Gait Analysis

Joint kinematics measurement shows potential in aiding clinical evaluation and therapeutic treatment comparisons [Favre et al., 2008]. For this reason, several techniques were developed to measure joint motion, especially for the lower limbs.

[O'Donovan et al., 2007] used angular rate and magnetic (AARM) sensors to measure the joint angles from the orientation of one segment relative to another, and applied their method to the ankle joint. Evaluation of their approach was performed through an experiment which involved two subjects who performed 13 leg exercises. Comparison was made with a 3D motion analysis system based on markers. The root mean square error was calculated for the angles measured through the authors' approach and the angles measured by the motion analysis system, as $RMSE = \sqrt{1/T \sum_{k=1}^{T} \left(\phi\left(k\right)^S - \phi\left(k\right)^E \right)^2}$, where S meant angles measured with the AARM sensors and E meant angles measured with the motion analysis system. The considered angles were: flexion, internal/external rotation and in/eversion. Results showed a strong correlation between the two measurement methods.

[Favre et al., 2008] measured the knee joint angle using two inertial measurement units (IMUs), attached to the thigh and shank. They estimated the orientations of the two IMUs, and then aligned the two reference frames, to get the knee angle. Validation of the method was made through an experiment where two subjects performed two hip abduction/adduction movements and two level ground walks of 30 meters. For comparison, a magnetic tracking device was used to simultaneously record the experimental data. A thorough assessment of the experimental data included five parameters: firmness of the knee during hip abduction/adduction movement, errors of alignment, repeatability of the transformation angle between the reference frames, repeatability of the standing posture and errors of the system for 3D knee angle measurement.

Knee firmness was estimated by measuring the variation of knee angle during movement. The differential orientation between the thigh frame and the shank frame were measured by the magnetic tracking system and the standard deviation of the total angle corresponding to the differential orientation was calculated. The firmness hypothesis was validated by the small variations of the standard deviation.

The alignment error was represented by the differential orientation between the shank inertial and magnetic frames. Results showed that the two reference frames of the IMUs could be aligned accurately with an error of less than 5% in the horizontal plane.

The errors of the system for 3D knee angle measurement were assessed using the joint coordinate system recommended by the International Society of Biomechanics. The inertial and magnetic measured angles were compared by calculating an offset error, due to misalignment of the thigh and shank fixed reference frames, a dynamic error during walking trials, and the correlation coefficient (CC) between the two systems.

The offset error was found to be small for flexion/extension and abduction/adduction and high for internal/external rotation, as a consequence of the horizontal alignment error. The dynamic error was small

for all angles, and the CC was high for flexion/extension, for internal/external rotation and acceptable for abduction/adduction.

[Favre et al., 2010] proposed a combined system including a stationary motion capture system for calibration and wearable sensors for measuring lower body joint angles and segmental angular velocities. The motivation behind this system was the advanced calibration procedures for stationary systems and the possibility to measure gait over long distances. The same equipments were used as in [Favre et al., 2008], with the magnetic-based motion capture device as the stationary system and the IMUs as the wearable system.

Reliability of this method was shown by calculating the anatomical landmarks dispersion across calibrations and the effects of this dispersion on anatomical frames and kinematics. To assess kinematics intra and inter-trial repeatability, the following factors were employed: the offset $\mu_j = 1/100\| \sum_{i=1}^{100} X\left(i\right) - \sum_{i=1}^{100} x_j\left(i\right) \|$, calculated between the original X and the corrupted x mean cycle kinematics (corrupted kinematics were obtained by applying the difference of orientation between the mean anatomical frame and the anatomical frames obtained for each of the subjects repetitions of the experiment to the anatomical frame of a healthy subject), the coefficient of multiple correlation (CMC) for the similarity between the relative corrupted patterns, the dispersion σ_i of the corrupted kinematics, and δ as the dispersion of characteristic features among the corrupted cycles.

[Ferrari et al., 2010] assessed "Outwalk", a protocol to measure thorax-pelvis and lower-limb kinematics during gait, with the aid of Xsens, as an inertial and magnetic measurement system (IMMS). The system's ability to measure joint kinematics was compared with Vicon, via a clinical gait analysis protocol consistent with International Society of Biomechanics, called CAST.

During experimentation, both systems measured the same gait cycles synchronously. Three tests were performed: Outwalk and CAST applied to Vicon data, determining accuracy of Xsens and Vicon data processed with CAST and the differences between the kinematics of Outwalk applied to Xsens data and CAST applied to Vicon data. Five parameters were assessed for each test. Namely, for the third test, $OX(t)$ and $CV(t)$ represented the waveforms for each measured joint angle, given by the Outwalk-Xsens and CAST-Vicon data and protocols.

The computed parameters were: an offset $off = mean\left(OX\left(t\right)\right) - mean\left(CV\left(t\right)\right)$, r - the Pearson's correlation coefficient, the difference between the range of motion $\Delta ROM = ROM\left(OX\left(t\right)\right) - ROM\left(CV\left(t\right)\right)$, and two coefficient of multiple correlation (CMC) calculated before and after zeroing the offset off for each pair $OX(t)$, $CV(t)$.

Results showed that removing the offset improved the CMC values significantly. Also, the results for the other parameters showed a clear correspondence between the Outwalk kinematics with Xsens data and the CAST kinematics with Vicon data.

[Djurić-Jovičić et al., 2011] presented a method based on digital filtering to estimate leg joints angles using accelerometer arrays attached to body segments. The angles were obtained by subtracting the absolute angles of the neighboring leg segments.

To assess the accuracy of the algorithm, an experiment was performed where several subject walked at their natural pace and on a treadmill at various speeds. Goniometers were used as the reference system. The difference between the angles obtained by the proposed method and the angles obtained from the goniometers was calculated. Then the Pearson's correlation coefficient and the root square mean error were calculated between the experimental angles and the reference angles. Results showed that the method was reliable for measuring angles for clinical applications.

6.3   Motion Estimation

We will refer in this section to a number of papers that studied motion estimation and interactive control of virtual humans from different sources like video cameras or a simplified set of sensors.

[Chai and Hodgins, 2005] investigated an approach to performance animation aided by video cameras and a small number of markers as well as by a prerecorded motions database. Users were able to control virtual characters in real time, performing different behaviors. The performance of this method was correlated to the low-dimensional human motion representation developed in this scope. The method was therefore compared with other dimensionality reduction techniques by comparing the reconstruction error, calculated as the L_2 distance between the original motion and the reconstructed motion. In an end-to-end evaluation, the motion of a user was recorded using a full marker system and compared to the reconstructed motion via the small set of markers and video cameras. A reconstruction error of 2.54 degrees per joint angle was obtained.

[Sminchisescu et al., 2005] developed a probabilistically motivated tracking algorithm, based on a Bayesian Mixture of Experts Model. A human motion capture database was used in the learning algorithm. Motions were reconstructed from video camera input data.The approach was evaluated by computing the root mean square error per joint angle in degrees. A similar measure was used by [O'Donovan et al., 2007], discussed in the previous section 6.2.

Furthermore in this area, [Liu et al., 2011] studied an approach to control virtual characters with the aid of motion data captured with inertial sensors. They employed statistical motion modeling in which the closest poses from the database to the current pose were used as training data to learn a dynamic model mapping previous poses to the current pose. Among others, they used leave-one-out evaluation to validate their results. In this phase, human animations were synthesized and the errors were measured by degrees per joint angle per frame.

[Sigal and Black, 2006] debated that using joint angle distances as error measures (as we've seen so far in this section) depended on the parameterization of the human body, therefore it could not be used to compare different methods applied on characters with different degrees of freedom or different parameterizations of the joint angles. To overcome this difficulty, they provided a dataset to be used for comparisons and proposed an error measure with wide applicability. This measure was based on a sparse set of virtual markers that corresponded to joint and limb endpoints locations. Mathematically, the proposed measure was formulated as $D\left(X, \hat{X}, \hat{\Delta}\right) = \sum_{m=1}^{M} \frac{\hat{\delta}_m \|x_m - \hat{x}_m\|}{\sum_{i=1}^{M} \hat{\delta}_i}$, where M was the number of markers, X the body state expressed according to the marker positions m, and $\hat{\delta}_m$ was 1 if the proposed algorithm could recover the respective marker and 0 otherwise. This way, the binary selection variable $\hat{\Delta} = \{\hat{\delta}_1, \hat{\delta}_2, \ldots, \hat{\delta}_M\}$ could ensure that algorithms that used different representations could be compared. The 3D error was calculated in millimeters.

7 | Physics

As perceived realism and the physical realism of a motion are interconnected [Geijtenbeek et al., 2010], it is needless to say that evaluating if an animated character obeys the laws of physics gives a clear indicator of whether the technique that generated the motion is able to produce visually realistic and physically feasible results.

7.1 Studies on Interpolation, Concatenation and Adaptation

[Safonova and Hodgins, 2005] studied linear interpolated human motions and identified a number of properties that needed to be ensured in order to obtain a physically correct resulting motion. The analysis was conducted on the three separate phases of a motion: the flight, the contact and the transition between the two.

It was shown that interpolating the center of mass trajectories instead of the root positions would eliminate the non-linearity in the center of mass trajectory during the flight phase, leaving the components that are not influenced by gravity constant. In this phase, the angular momentum would remain constant should the motions lack visible rotations, or rotate approximately around the same axis.

Imposing that the feet should not slide on contact with the environment, the analysis showed that interpolating the feet positions, the body center of mass positions and the non-redundant degrees of freedom (root position, all joint angles except legs and two "knee circle" parameters) would preserve continuity of the motion in this situation.

Next, the analysis showed that if two motions were statically balanced, the resulting interpolated motion would be the same. Ground contact was imposed to require a limited coefficient of friction, and it was shown that if the interpolated motions had a ground reaction force within the friction cone, the resulting motion would also satisfy this property.

Last, in the transition phase, it was shown that for motions that lacked rotation during flight, rotated approximately around the same axis, or took place in the vertical plane, the velocity of the center of mass would be continuous.

This analysis can be regarded as an inverse procedure than that of the assessment that we are discussing in this review, as it reveals a procedure that ensures physical correctness. Subsequently, we will point out several works that evaluate the physical correctness of motions resulted either from interpolation, retargeting or concatenation.

[Pronost and Dumont, 2006] focused on retargeted and interpolated motions and investigated the physical validity of an adapted motion by comparing the resulting forces and torques with biomechanical literature and by comparing ground reaction forces with experimental data from force plates. Using a virtual character with morphology identical to a real one and analyzing the plots of different measures (ground reaction forces, vertical, lateral and fore-aft forces and torques at each joint) showed that these measures behave as expected from known biomechanical properties. Different real locomotions were analyzed during a full gait cycle and several observations were made that confirmed known properties such as the double hump of the vertical force, the negative lateral force during the stance phase, or the backward

direction of the fore-aft force during the first half of the support and the opposite orientation during the second half.

Further, [Pronost and Dumont, 2007] used the forces and torques that drove the motion to synthesize new physically valid motions. The authors investigated the relationship between force and torques normalizations and motion style of character morphology. For this purpose, a set of motions performed by two characters were compared based on the kinematical difference between the motions (by analyzing the lateral, horizontal and vertical components of the root node) and the differences between the normalized ground reaction forces. This comparison was a statistical evaluation using the root means square error (of the first order), the average value of Euclidian distances, the average value of correlation coefficient and the average first time derivative.

[Multon et al., 2007] studied the physical correctness when retargeting acrobatic aerial motions to characters with different topologies. They corrected the center of mass position and ensured that the angular momentum was constant during the aerial phase. The latter was done through two different strategies: by adapting the angular velocity of the root while keeping the pose unchanged and by adjusting the rotation of the body segments (first arms, then legs) to keep the total angular velocity unchanged. The X, Y and Z components of the angular momentum during the aerial phase of a corrected and an original motion capture sequence scaled on a different character were plotted and compared. The graph revealed that indeed the angular momentum was constant for the corrected motion.

[Shum et al., 2009] explored the angular momenta of two motions to obtain a physically correct concatenated motion. Similarly to [Multon et al., 2007], in order to validate the results, the angular momentum trajectories during two consecutive forwards flips were plotted as executed by a real performer and by a virtual character driven by their method. A close look at this simple graph revealed that the results were similar to the real performance.

Similarly, to validate their method for real-time motion adaptation, [Hoyet et al., 2010] plotted the trajectory of the center of pressure and compared a pushing motion with an adapted one where 200 N force were added. The graph showed that the center of pressure for the adapted motion remained close to the center of pressure of the original one.

Therefore, we can conclude from the last four cited articles that comparing trajectories, either by observing similarities in graphs or using statistical measures is a useful and visual inspection correlated indicator of the physical validity of a motion. It would be interesting to see which other measures beside the root node position, angular momentum and center of pressure could provide reliable assessment material.

7.2 Muscular Models and Injury Assessment

[Geijtenbeek et al., 2010] took a step further and looked at a musculoskeletal model of a virtual character. They designed two quality measures. The dynamic error measure accounted for the amount of external force and moment required for the motion to satisfy the Newton-Euler laws of motion. These laws state that changes in linear and angular momentum have to correspond to external forces due to gravity and environment interaction.

The second measure, called the muscle error measure was defined as the total amount of excess muscle force (on top of the maximum capacity) required for a character to perform an animation, normalized by the total maximum force of all muscles in the model. This showed whether the effort that a muscle put in a motion was realistic or not.

The two measures were combined in a final score as the average across all frames in an animation. Experiments showed that changing the weight of the virtual body resulted in an increased muscle error for more effort-intensive motions. The muscle error as well as the dynamics error were found to be correlated with slowing down or speeding up the animation for dynamic motions, like walking or jumping, showing that these modifications cannot be done without the respective adjustments in the motion.

[Geijtenbeek et al., 2011] also tackled the issue of injury in an animated character. Research on injuries from motor-vehicle accidents underlay a set of individual measures for head, neck, chest, pelvis, arms, legs, ankles and feet. Except for the head, the individual measure represented the normalized maximum of a physical property averaged over a time window. This physical property was taken from literature and was given by, for example, the maximum acceleration for the pelvis during side impact for the pelvis injury measure, or the total magnitude of constraint forces applied by the wrist joint to hand and lower arm for the arms measure.

The final score was an average of all the independent measures. A user study showed a significant correlation between the measure as defined earlier and the injury level perceived by observers, thus validating the fact that the measure was reliable in assessing the injury inflicted on a virtual character.

8 | Soft Tissue Simulation

Soft tissue simulation is a widely researched area due to multiple applications in computer graphics, biomechanics and medicine. Main interests include modeling organs, muscles and skin deformations. As this research provides numerous applications in the domains mentioned earlier, the validation of the researched models is a main concern. The most popular means of evaluating experimental results are: comparison with magnetic resonance imaging (MRI) or ultrasound data (mainly discussed in section 8.1), comparison with dense marker data (discussed in section 8.2) and in vivo experiments. The last type of validation experiments is common for organ simulation for surgical purposes. In this case, it is of paramount importance to have a thorough validation of the model. However, in vivo experiments are mostly carried out on animals. In this review, we would like to take a look at evaluation possibilities in the case of human virtual character muscles and skin modeling.

In the following, we will discuss muscle modeling and skin deformation modeling techniques. In this area, there are three main methodologies developed so far: geometrically-based, physically-based and data-driven approaches [Lee et al., 2010]. We will see how each of these matches the evaluation techniques mentioned above.

8.1   Muscles

[Arnold et al., 2000] developed a method to construct models of musculoskeletal geometry from MR images. They used around 250 images to build the models for three lower extremity cadaveric specimens.

For evaluation, the tendon excursion method was used to determine the hip flexion-extension and knee flexion moment arms. This method implied measuring the length changes of three muscles (the medial hamstrings comprised of the semimembranosus and semitendinosus muscles and the psoas muscle, chosen because of the interventions performed on them to treat movement abnormalities provoked by cerebral palsy) during flexion. The moment arms were calculated as the partial derivative of the muscle-tendon lengths with respect to joint angle. All the data was collected through an experiment performed on cadaveric specimens. This data was compared with the moment arms predicted by the implemented model. The average error was computed as the average of the absolute difference between the moment arms, in millimeters and as percentage of the experimental moment arms.

The average errors were within 10% of the experimental moment arms. Further, the errors in the length changes of the muscles that corresponded to the movement arm errors, were calculated and compared to variations in the peak muscle-tendon lengths from 18 unimpaired subject. The maximum errors found were less than one standard deviation of the peak experimental lengths.

[Lemos et al., 2005] developed a non-linear dynamic finite element model to solve a continuum model for general muscle fiber architecture. Theoretical results provided by the model were compared to ultrasound medical imaging experimental results. The structural changes and force production in the tibialis anterior muscles during contraction were investigated by measuring relaxed and activated fascicle lengths (in millimeters), angle of pennation (in degrees) and external forces, for different levels (percentages of the maximum torque) of maximum voluntary contraction. The theoretical results were found in agreement with experimental data.

[Blemker and Delp, 2005] used MR images of a live subject to create 3D finite-element models of four muscles that crossed the hip. To evaluate their model, the muscle moment arms ma^{fiber} were calculated according to the principal of virtual work, $ma^{fiber} = \partial l^{fiber}/\partial\theta$, where the function $l^{fiber}(\theta)$ was obtained by fitting a fourth-order polynomial to the observed fiber lengths and joint angles. The moment arms were compared with the previously discussed results of [Arnold et al., 2000], among others. Next, the changes in shape of the 3D muscle models were compared to those obtained from MR images. Each point of the surfaces segmented from MR images was projected on the surface of the 3D muscle model, resulting in distance errors for each point. The average and root mean square (RMS) errors were calculated across all points.

[Südhoff et al., 2009] studied a method to reconstruct 3D knee muscles models from MRI images. The method was an adaptation of the deformation of a parametric specific object (DPSO) approach, which required a low number of axial MRI images (slices) to reconstruct the geometry of the knee. The contours of 12 muscles from 15 young subjects (both asymptomatic and suffering of ligament rupture) were outlined manually using specific software.

A reference model was built from the manual identification of 300 continuous images, for two random asymptomatic subjects. The volume error was investigated as the difference between the reference and the model built using the DPSO method. They found that for five to seven slices, the error was inferior to 5%.

Also, similarly to [Blemker and Delp, 2005], the error in shape was calculated by projecting each point on the experimental model surface onto the reference model surface and taking the distance between point and projection (called point-surface error). The RMS was less than 5 mm for seven slices or more.

To test the reproducibility of muscle reconstruction with the DPSO method, two operators reconstructed the muscles from ten asymptomatic subjects, using the optimum number of slices for an under 5% error determined by the DPSO method. The work load was reduced from 12 hours to one hour per subject. The relative volume reproducibility was calculated as the absolute difference in volume between the two reconstructions, divided by the first one. The mean difference and standard deviation of the relative volume were examined, as well as the interclass correlation coefficient. Shape reproducibility was calculated as the RMS of the point-surface error. Variable results were found, however the point-surface errors were placed in a confidence interval for all muscles.

[Oberhofer et al., 2009] implemented the Host Mesh Fitting (HMF) technique to predict muscle deformation of a subject-specific musculoskeletal model during walking. MR scans of the lower body from a female subject were used for this purpose. Similarly to [Südhoff et al., 2009] and [Blemker and Delp, 2005], the validation was aided by examining the shape changes, as the RMS of the differences between the points of the MR image data and their projection on the surface obtained by the HMF method. This was done for five muscles, during a motion from a 15° knee angle to a 45° knee angle. The best RMS error was obtained for the tibialis anterior, at 1.8 mm.

[Vasavada et al., 2008] developed a method to determine wrapping surface parameters for muscle paths that best approximated the centroid paths of muscles, and applied it on the neck musculature. MRI data was collected for 18 neck muscles from a single male subject in seven different postures. This data was used to determine muscle geometry. Centroid path data from a neutral posture was used to characterize a wrapping surface at each vertebra.

For evaluation, the wrapping parameters from the neutral posture were rotated with the corresponding vertebral body to correspond to the other different postures. The average distance between the centroid path and the modeled path was calculated as the sum of distances between the centroid path to the modeled path at each MRI slice, normalized by muscle length. This represented the error metric. It was also calculated for straight and centroid paths, to reveal the necessity of the wrapping surface. It was decided that an error metric for a straight line that was less than 10% of its distance from the centroid path to the vertebral center made the wrapping surface unnecessary for the respective muscle. Thirteen out of the eighteen modeled muscles benefited from a wrapping surface. Also, it was found that, for the semispinalis capitis muscles, all postures benefited in a considerable manner from the the the use of wrapping surfaces.

Most recently, [Suderman and Vasavada, 2012] modeled the curved muscle paths in the cervical spine accounting for soft tissue deformation determined by posture changes. For this, they used moving muscle points (MMP), which could move with respect to the body segment to which the point was linked, thus allowing for muscle paths to bend according to surrounding deformable soft tissue. MRI scans were collected for 15 muscle pairs, from two male subjects, in five flexion-extension postures. From this data, muscle paths were modeled using MMP.

In the evaluation phase, the results from [Vasavada et al., 2008] were used for comparison. The error metric represented the average distance between modeled muscle path and the smoothed centroid path. Also, in this case the curved paths were compared to the straight path, and a percentage of improvement was calculated for each muscle as the difference between the error metric for the straight path and the error metric for the curved path, divided by the first. A statistical analysis was run using the repeated measures one-way ANOVA. Results showed that paths produced by MMP closely followed the centroid paths for all postures, and showed better results than other two tested methods - fixed via point method and straight line paths.

8.2 Skin

[Park and Hodgins, 2006] presented a data-driven approach to capture the fine details of the human body surface during movement, by using around 350 markers placed on muscular parts of the body, to obtain both the motion of the skeleton and that of the skin. They extracted the rigid body motion and computed local deformations of a subject-specific polygonal model using the marker set information. Evaluation was purely visual based and found better results than two other methods: rigid deformation and quadratic deformation without resolving residuals. This work represented the base for [Park and Hodgins, 2008] who used the previous data as ground truth in an automatic evaluation procedure.

[Park and Hodgins, 2008] simplified the work previously required to record the motion capture data. They used previous data to statistically derive static and dynamic deformation models. Static deformation were formulated as pose functions, while dynamic deformations were obtained by fitting dynamic equations to pre-recorded data to model the effects of muscles moving the joints and of muscles and fat inertia. These models were applied to skeletal body motion captured using a number of 40-50 markers, thus obtaining a virtual character that depicted skin deformations during movement.

Learning static and dynamic deformation was aided by two different databases. The static one contained the slower versions of the motions from the dynamic one. For evaluation, a dynamic model was built on top of the static database, and then the motion was reconstructed. Reconstruction errors were measured as comparisons between original markers positions and simulated marker positions (in millimeters per frame). The dynamic deformation over time was evaluated as the comparison between the original PCA component of dynamic deformation and the simulated one. They found that the results were well matched on an overall level to the original motion. Synthesized results were similarly compared to ground truth data.

[Bickel et al., 2008] focused on facial details and developed a technique that combined computational and data-driven approaches to transfer fine-scale details to novel facial animations. They decomposed facial geometry into large-scale motion - using a linear shell deformation model obtained through a sparse set of markers or handle points defined by the user - and fine-scale details learned from a set of example poses. The learning was done through a novel pose-space deformation technique. Then, the technique computed the fine-scale details for new facial expressions.

For evaluation, the performance of an actor captured with high-resolution surface details was divided in two halves, the first used as training data and the second as ground-truth data. Comparison of the implemented method with ground-truth data was done based on the L^2 error (in millimeters).

9 | Conclusions

We have seen in this review that virtual character assessment is starting to take shape. We have addressed a number of papers that focused on evaluating motion naturalness, perceptual plausibility, comparing behavior in crowds or put forward unified frameworks to compare different methods (for steering and motion tracking). Although a difficult task, due to different models and representations, automatic assessment is taking steps forward.

While user studies will probably still remain a reliable validation method for a long time, the tendency now is to use them as complementary to an automatic sort of validation, which highliths the correlation between what humans perceive to be natural or correct and what mathematical and physical laws determine to be so. However, there is still one strong advantage of user studies, namely that they can draw attention to the algorithms' weaknesses.

We have also seen a number of papers that do not concern virtual character animation directly, for example the ones discussed in 6.2. Although these papers referred to measuring kinematical parameters and assessing the precision of their findings, the evaluation metrics (such as the root mean square error for joint angles) could be (and already have been) used in applications such as controlling a virtual character in real time with the aid of a suit of sensors. In this case the fidelity of the characters' motion corresponding to the user's was assessed.

Even though it is not always possible to propose a straightforward metric to assess motion, the works which offer guidelines based on perceptual studies proved to have real applicability for subsequent research. We can therefore distinguish that virtual character assessment is currently approached in two ways: either by proposing evaluation metrics and procedures against other methods or ground truth data or by observing thresholds and patterns to bring up guidelines for developers who want to achieve a specific tradeoff between naturalness and computational requirements.

In the closing of this review, we offer a very brief summary of metrics and guidelines covering most of the chapters we discussed, in the following table.

<div align="center">TABLE 9.1: Literature Review Summary</div>

Chapter	Reference	Short Description
Motion Naturalness	[Ren et al., 2005]	naturalness measure based on ensemble statistical models for natural motion
		$s_i = \frac{\log P(D\|\theta_i)}{T}$
		$s = \min_i \left(\frac{(s_i - \mu_i)}{\sigma_i} \right)$
	[McDonnell et al., 2005]	1.164 : 1 pixel to texel ratio at which impostors become distinguishable
Behavior in Crowds	[Lerner et al., 2009]	density based measure for crowd comparison
	[Guy et al.]	entropy metric for crowd comparison
	[Musse et al., 2012]	used 4d histograms to compare global flows, spatial occupancy, agents' density and orientation distribution
Style	[Vasilescu, 2002]	mapped motions into the space of people parameters or the space of action parameters to recognize the person performing or the action performed
Kinematics	[Lee et al., 2002]	transition probability between two frames as exponential function of the distance D_{ij}
		$D_{ij} = d(p_i, p_j) + \nu d(v_i, v_j)$
	[O'Donovan et al., 2007]	compared angle joint measurements to ground-truth data using RMSE
		$RMSE = \sqrt{1/T \sum_{k=1}^{T} \left(\phi(k)^S - \phi(k)^E \right)^2}$
	[Chai and Hodgins, 2005]	compared performance animation with ground-truth data using the L_2 error in degrees per joint angle
	[Sminchisescu et al., 2005]	compared video based reconstructed motion with ground-truth data using the RMSE in degrees per joint angle
	[Sigal and Black, 2006]	comparison measure for motion tracking algorithms based on sparse set of virtual markers, in millimeters
		$D\left(X, \hat{X}, \hat{\Delta}\right) = \sum_{m=1}^{M} \frac{\hat{\delta}_m \|x_m - \hat{x}_m\|}{\sum_{i=1}^{M} \hat{\delta}_i}$
Physics	[Multon et al., 2007]	compared the angular momentum trajectories to physically validate retargeted motion
	[Shum et al., 2009]	compared the angular momentum trajectories to physically validate concatenated motions
	[Geijtenbeek et al., 2010]	dynamic error measure and muscle error measure for musculoskeletal models
Soft Tissue Simulation	[Arnold et al., 2000]	muscle reconstruction error equal to the average of the absolute difference between the moment arms, in millimeters
	[Blemker and Delp, 2005]	RMSE across point of muscle surfaces from MRI and their projections on simulated muscle model
	[Park and Hodgins, 2008]	skin reconstruction errors as comparisons between original markers positions and simulated marker positions (in millimeters per frame)
	[Bickel et al., 2008]	compared facial details based on L_2 error in millimiters

Bibliography

L. Ren, A. Patrick, A.A. Efros, J.K. Hodgins, and J.M. Rehg. A data-driven approach to quantifying natural human motion. In *ACM Transactions on Graphics (TOG)*, volume 24, pages 1090–1097. ACM, 2005.

M. Vicovaro, L. Hoyet, L. Burigana, and C. O'Sullivan. Evaluating the plausibility of edited throwing animations. In *Eurographics/ACM SIGGRAPH Symposium on Computer Animation*, pages 175–182. The Eurographics Association, 2012.

L. Hoyet, R. McDonnell, and C. O'Sullivan. Push it real: Perceiving causality in virtual interactions. *ACM Transactions on Graphics (TOG)*, 31(4):90, 2012.

T. N. Cornsweet. The staircase-method in psychophysics. *American Journal of Psychology*, 75:485–491, 1962.

F. Tecchia, C. Loscos, and Y. Chrysanthou. Visualizing crowds in real-time. In *Computer Graphics Forum*, volume 21, pages 753–765. Wiley Online Library, 2003.

Rafael Rodriguez, Eva Cerezo, Sandra Baldassarri, and Francisco J. Seron. Technical section: New approaches to culling and lod methods for scenes with multiple virtual actors. *Comput. Graph.*, 34(6): 729–741, December 2010. ISSN 0097-8493.

J. Hamill, R. McDonnell, S. Dobbyn, and C. O'Sullivan. Perceptual evaluation of impostor representations for virtual humans and buildings. In *Computer Graphics Forum*, volume 24, pages 623–633. Wiley Online Library, 2005.

Rachel McDonnell, Carol O'Sullivan, and Simon Dobbyn. Lod human representations: A comparative study. 2005.

Simon Dobbyn, John Hamill, Keith O'Conor, and Carol O'Sullivan. Geopostors: a real-time geometry / impostor crowd rendering system. In *Proceedings of the 2005 symposium on Interactive 3D graphics and games*, I3D '05, pages 95–102, New York, NY, USA, 2005. ACM. ISBN 1-59593-013-2.

Francisco Ramos, Oscar Ripolles, and Miguel Chover. Continuous level of detail for large scale rendering of 3d animated polygonal models. In *Proceedings of the 7th international conference on Articulated Motion and Deformable Objects*, AMDO'12, pages 194–203, Berlin, Heidelberg, 2012. Springer-Verlag. ISBN 978-3-642-31566-4.

K.A. Yuksel, A. Yucebilgin, S. Balcisoy, and A. Ercil. Real-time feature-based image morphing for memory-efficient impostor rendering and animation on gpu. *The Visual Computer*, pages 1–10, 2012.

Dennis Majoe, Lars Widmer, and Jürg Gutknecht. Enhanced motion interaction for multimedia applications. In *Proceedings of the 7th International Conference on Advances in Mobile Computing and Multimedia*, MoMM '09, pages 13–19, New York, NY, USA, 2009a. ACM.

Ronan Billon, Alexis Nédélec, and Jacques Tisseau. Gesture recognition in flow based on pca analysis using multiagent system. In *Proceedings of the 2008 International Conference on Advances in Computer Entertainment Technology*, ACE '08, pages 139–146, New York, NY, USA, 2008. ACM.

Marcel Martin, Jonathan Maycock, Florian Paul Schmidt, and Oliver Kramer. Recognition of manual actions using vector quantization and dynamic time warping. In *Proceedings of the 5th international conference on Hybrid Artificial Intelligence Systems - Volume Part I*, HAIS'10, pages 221–228, Berlin, Heidelberg, 2010. Springer-Verlag.

Masaki Oshita and Takefumi Matsunaga. Automatic learning of gesture recognition model using som and svm. In *Proceedings of the 6th international conference on Advances in visual computing - Volume Part I*, ISVC'10, pages 751–759, Berlin, Heidelberg, 2010. Springer-Verlag.

Przemysław Głomb, Michał Romaszewski, Arkadiusz Sochan, and Sebastian Opozda. Unsupervised parameter selection for gesture recognition with vector quantization and hidden markov models. In *Proceedings of the 13th IFIP TC 13 international conference on Human-computer interaction - Volume Part IV*, INTERACT'11, pages 170–177, Berlin, Heidelberg, 2011. Springer-Verlag.

Dennis Majoe, Lars Widmer, Philip Tschiemer, and Jürg Gutknecht. Tai chi motion recognition using wearable sensors and hidden markov model method, 2009b.

Kai Kunze, Michael Barry, Ernst A. Heinz, Paul Lukowicz, Dennis Majoe, and Jürg Gutknecht. Towards recognizing tai chi - an initial experiment using wearable sensors. In *Heinz1, Paul Lukowicz, Dennis Majoe, Jurg Gutknecht; Institute for Computer Systems and Networks UMIT*, 2006.

Jian Xiang, Jian-guang Weng, Yue-ting Zhuang, and Fei Wu. Ensemble learning hmm for motion recognition and retrieval by isomap dimension reduction. *Journal of Zhejiang University - Science A*, 7: 2063–2072, 2006.

VI Levenshtein. Binary codes capable of correcting deletions, insertions and reversals. In *Soviet Physics Doklady*, volume 10, page 707, 1966.

T. Kailath. The divergence and bhattacharyya distance measures in signal selection. *Communication Technology, IEEE Transactions on*, 15(1):52–60, 1967.

L.E. Baum, T. Petrie, G. Soules, and N. Weiss. A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains. *The annals of mathematical statistics*, pages 164–171, 1970.

Meinard Müller. *Information Retrieval for Music and Motion*, chapter 4. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2007.

Donald J. Berndt and James Clifford. Advances in knowledge discovery and data mining. chapter Finding patterns in time series: a dynamic programming approach, pages 229–248. American Association for Artificial Intelligence, Menlo Park, CA, USA, 1996.

Ronan Billon, Alexis Nédélec, and Jacques Tisseau. Recognition of gesture sequences in real-time flow, context of virtual theater. In *Proceedings of the 8th international conference on Gesture in Embodied Communication and Human-Computer Interaction*, GW'09, pages 98–109, Berlin, Heidelberg, 2010. Springer-Verlag.

Shawn Singh, Mishali Naik, Mubbasir Kapadia, Petros Faloutsos, and Glenn Reinman. Motion in games. chapter Watch Out! A Framework for Evaluating Steering Behaviors, pages 200–209. Springer-Verlag, Berlin, Heidelberg, 2008.

Shawn Singh, Mubbasir Kapadia, Petros Faloutsos, and Glenn Reinman. Steerbench: a benchmark suite for evaluating steering behaviors. *Computer Animation and Virtual Worlds*, 20(5-6):533–548, 2009.

Mubbasir Kapadia, Matthew Wang, Glenn Reinman, and Petros Faloutsos. Improved benchmarking for steering algorithms. In *Motion in Games*, pages 266–277, 2011a.

Mubbasir Kapadia, Matt Wang, Shawn Singh, Glenn Reinman, and Petros Faloutsos. Scenario space: characterizing coverage, quality, and failure of steering algorithms. In *Proceedings of the 2011 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, SCA '11, pages 53–62, New York, NY, USA, 2011b. ACM.

Stephen J. Guy, Jatin Chhugani, Sean Curtis, Pradeep Dubey, Ming Lin, and Dinesh Manocha. Pledestrians: a least-effort approach to crowd simulation. In *Proceedings of the 2010 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, SCA '10, pages 119–128, Aire-la-Ville, Switzerland, Switzerland, 2010. Eurographics Association.

Mubbasir Kapadia, Shawn Singh, Brian Allen, Glenn Reinman, and Petros Faloutsos. Steerbug: an interactive framework for specifying and detecting steering behaviors. In *Proceedings of the 2009 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, SCA '09, pages 209–216, New York, NY, USA, 2009. ACM.

Ioannis Karamouzas, Peter Heil, Pascal Beek, and Mark H. Overmars. A predictive collision avoidance model for pedestrian simulation. In *Proceedings of the 2nd International Workshop on Motion in Games*, MIG '09, pages 41–52, Berlin, Heidelberg, 2009. Springer-Verlag.

Alon Lerner, Yiorgos Chrysanthou, Ariel Shamir, and Daniel Cohen-Or. Data driven evaluation of crowds. In *Proceedings of the 2nd International Workshop on Motion in Games*, MIG '09, pages 75–83, Berlin, Heidelberg, 2009. Springer-Verlag.

Bikramjit Banerjee and Landon Kraemer. Evaluation and comparison of multi-agent based crowd simulation systems. In *Agents for Games and Simulations II*, volume 6525 of *Lecture Notes in Computer Science*, pages 53–66. Springer Berlin / Heidelberg, 2011.

Pengfei Xing, Michael Lees, Hu Nan, and T. Vaisagh Viswanthatn. Validation of agent-based simulation through human computation: an example of crowd simulation. In *Proceedings of the 12th international conference on Multi-Agent-Based Simulation*, MABS'11, pages 90–102, Berlin, Heidelberg, 2012. Springer-Verlag.

Soraia R. Musse, Vinicius J. Cassol, and Cláudio R. Jung. Towards a quantitative approach for comparing crowds. *Computer Animation and Virtual Worlds*, 23(1):49–57, 2012.

S.J. Guy, J. van den Berg, W. Liu, R. Lau, M.C. Lin, and D. Manocha. A statistical similarity measure for aggregate crowd dynamics.

M. Alex O. Vasilescu. An algorithm for extracting human motion signatures. In *IEEE Conference on Computer Vision and Pattern Recognition*, Kauai, HI, 2001a.

N. F. Troje. Decomposing biological motion: A framework for analysis and synthesis of human gait patterns. In *Journal of Vision*, volume 2, pages 371–387, 2002a.

Nikolaus Troje. The little difference: Fourier based synthesis of genderspecific biological motion. In Rolf P. Würtz and Markus Lappe, editors, *Dynamic Perception*, pages 115–120, Berlin, 2002b. AKA Press.

Nikolaus F. Troje. 12. Retrieving Information from Human Movement Patterns. *Understanding Events*, pages 308–335, March 2008.

M. Alex O. Vasilescu. Human motion signatures for character animation. August 2001b.

A. Kapteyn, H. Neudecker, and T. Wansbeek. An approach ton-mode components analysis. *Psychometrika*, 51(2):269–275, 1986.

M. Alex O. Vasilescu. Human motion signatures: Analysis, synthesis, recognition. In *Proceedings of the 16 th International Conference on Pattern Recognition (ICPR '02)*, volume 3 of *ICPR '02*, 2002.

Yuichi Kobayashi and Jun Ohya. Em-in-m: Analyze and synthesize emotion in motion. In *Advances in Machine Vision, Image Processing, and Pattern Analysis*, volume 4153 of *Lecture Notes in Computer Science*, pages 135–143. Springer Berlin Heidelberg, 2006.

Leonid Sigal, David J. Fleet, Nikolaus F. Troje, and Micha Livne. Human attributes from 3d pose tracking. In *Proceedings of the 11th European conference on computer vision conference on Computer vision: Part III*, ECCV'10, pages 243–257, Berlin, Heidelberg, 2010. Springer-Verlag.

Micha Livne, Leonid Sigal, Nikolaus F. Troje, and David J. Fleet. Human attributes from 3d pose tracking. *Computer Vision and Image Understanding (CVIU)*, 116:648–660, 2012.

B. J. H. van Basten and A. Egges. Evaluating distance metrics for animation blending. In *Proceedings of the 4th International Conference on Foundations of Digital Games*, FDG '09, pages 199–206, New York, NY, USA, 2009. ACM.

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 Trans. Graph.*, 21(3):491–500, July 2002. ISSN 0730-0301.

Okan Arikan and D. A. Forsyth. Interactive motion generation from examples. *ACM Trans. Graph.*, 21 (3):483–490, July 2002. ISSN 0730-0301.

Jing Wang and Bobby Bodenheimer. An evaluation of a cost metric for selecting transitions between motion segments. In *Proceedings of the 2003 ACM SIGGRAPH/Eurographics symposium on Computer animation*, SCA '03, pages 232–238, Aire-la-Ville, Switzerland, Switzerland, 2003. Eurographics Association.

Lucas Kovar, Michael Gleicher, and Frédéric Pighin. Motion graphs. In *Proceedings of the 29th annual conference on Computer graphics and interactive techniques*, SIGGRAPH '02, pages 473–482, New York, NY, USA, 2002. ACM.

Liming Zhao and Alla Safonova. Achieving good connectivity in motion graphs. *Graph. Models*, 71(4): 139–152, July 2009.

Arjan Egges, Tom Molet, and Nadia Magnenat-Thalmann. Personalised real-time idle motion synthesis. In *Proceedings of the Computer Graphics and Applications, 12th Pacific Conference*, PG '04, pages 121–130, Washington, DC, USA, 2004. IEEE Computer Society.

K. Forbes and E. Fiume. An efficient search algorithm for motion data using weighted pca. In *Proceedings of the 2005 ACM SIGGRAPH/Eurographics symposium on Computer animation*, SCA '05, pages 67–76, New York, NY, USA, 2005. ACM.

J. Favre, B.M. Jolles, R. Aissaoui, and K. Aminian. Ambulatory measurement of 3d knee joint angle. *Journal of Biomechanics*, 41(5):1029 – 1035, 2008.

Karol J. O'Donovan, Roman Kamnik, Derek T. O'Keeffe, and Gerard M. Lyons. An inertial and magnetic sensor based technique for joint angle measurement. *Journal of Biomechanics*, 40(12):2604 – 2611, 2007.

J. Favre, X. Crevoisier, B.M. Jolles, and K. Aminian. Evaluation of a mixed approach combining stationary and wearable systems to monitor gait over long distance. *Journal of Biomechanics*, 43(11): 2196 – 2202, 2010.

Alberto Ferrari, Andrea Giovanni Cutti, Pietro Garofalo, Michele Raggi, Monique Heijboer, Angelo Cappello, and Angelo Davalli. First in vivo assessment of "outwalk": a novel protocol for clinical gait analysis based on inertial and magnetic sensors. *Med. Biol. Engineering and Computing*, 48(1):1–15, 2010.

Milica D. Djurić-Jovičić, Nenad S. Jovičić, and Dejan B. Popović. Kinematics of gait: New method for angle estimation based on accelerometers. *Sensors*, 11(11):10571–10585, 2011.

J. Chai and J.K. Hodgins. Performance animation from low-dimensional control signals. In *ACM Transactions on Graphics (TOG)*, volume 24, pages 686–696. ACM, 2005.

C. Sminchisescu, A. Kanaujia, Z. Li, and D. Metaxas. Discriminative density propagation for 3d human motion estimation. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 390–397. IEEE, 2005.

Huajun Liu, Xiaolin Wei, Jinxiang Chai, Inwoo Ha, and Taehyun Rhee. Realtime human motion control with a small number of inertial sensors. In *Symposium on Interactive 3D Graphics and Games*, I3D '11, pages 133–140, New York, NY, USA, 2011. ACM.

L. Sigal and M.J. Black. Humaneva: Synchronized video and motion capture dataset for evaluation of articulated human motion. *Brown Univertsity TR*, 120, 2006.

Thomas Geijtenbeek, Antonie J. Van Den Bogert, Ben J. H. Van Basten, and Arjan Egges. Evaluating the physical realism of character animations using musculoskeletal models. In *Proceedings of the Third international conference on Motion in games*, MIG'10, pages 11–22, Berlin, Heidelberg, 2010. Springer-Verlag.

Alla Safonova and Jessica K. Hodgins. Analyzing the physical correctness of interpolated human motion. In *Proceedings of the 2005 ACM SIGGRAPH/Eurographics symposium on Computer animation*, SCA '05, pages 171–180, New York, NY, USA, 2005. ACM.

Nicolas Pronost and Georges Dumont. Validating retargeted and interpolated locomotions by dynamics-based analysis. In *Proceedings of the 4th international conference on Computer graphics and interactive techniques in Australasia and Southeast Asia*, GRAPHITE '06, pages 65–74, New York, NY, USA, 2006. ACM.

Nicolas Pronost and Georges Dumont. Dynamics-based analysis and synthesis of human locomotion. *Vis. Comput.*, 23(7):513–522, May 2007.

F. Multon, L. Hoyet, T. Komura, and R. Kulpa. Dynamic motion adaptation for 3d acrobatic humanoids. In *Proceedings of IEEE Humanoids'07*, Pittsburgh, PA, November 2007.

Hubert P. H. Shum, Taku Komura, and Pranjul Yadav. Angular momentum guided motion concatenation. *Comput. Animat. Virtual Worlds*, 20:385–394, June 2009. ISSN 1546-4261.

Ludovic Hoyet, Franck Multon, Taku Komura, and Anatole Lecuyer. Perception based real-time dynamic adaptation of human motions. In *Proceedings of the Third international conference on Motion in games*, MIG'10, pages 266–277, Berlin, Heidelberg, 2010. Springer-Verlag.

Thomas Geijtenbeek, Diana Vasilescu, and Arjan Egges. Injury assessment for physics-based characters. In *Proceedings of the 4th international conference on Motion in Games*, MIG'11, pages 74–85, Berlin, Heidelberg, 2011. Springer-Verlag.

Dongwoon Lee, Michael Glueck, Azam Khan, Eugene Fiume, and Ken Jackson. A survey of modeling and simulation of skeletal muscle. 2010.

Allison S. Arnold, Silvia Salinas, Deanna J. Asakawa, and Scott L. Delp. Accuracy of muscle moment arms estimated from mri-based musculoskeletal models of the lower extremity. *Computer Aided Surgery*, 5 (2):108–119, 2000.

Robson R. Lemos, Jon Rokne, Gladimir V. G. Baranoski, Yasuo Kawakami, and Toshiyuki Kurihara. Modeling and simulating the deformation of human skeletal muscle based on anatomy and physiology. *Computer Animation and Virtual Worlds*, 16(3-4):319–330, 2005.

Silvia Blemker and Scott Delp. Three-dimensional representation of complex muscle architectures and geometries. *Annals of Biomedical Engineering*, 33:661–673, 2005. ISSN 0090-6964.

I. Südhoff, Jacques A. de Guise, A. Nordez, Erwan Jolivet, D. Bonneau, V. Khoury, and Wafa Skalli. 3d-patient-specific geometry of the muscles involved in knee motion from selected mri images. *Med. Biol. Engineering and Computing*, 47(6):579–587, 2009.

Katja Oberhofer, Kumar Mithraratne, Ngaire S. Stott, and Iain A. Anderson. Anatomically-based musculoskeletal modeling: prediction and validation of muscle deformation during walking. *Vis. Comput.*, 25(9):843–851, July 2009.

Anita N Vasavada, Richard A Lasher, Travis E Meyer, and David C Lin. Defining and evaluating wrapping surfaces for mri-derived spinal muscle paths. *Journal of Biomechanics*, 41(7):1450–7, 2008.

B.L. Suderman and A.N. Vasavada. Moving muscle points provide accurate curved muscle paths in a model of the cervical spine. *J Biomech*, 45(2):400–4, 2012.

Sang Il Park and Jessica K. Hodgins. Capturing and animating skin deformation in human motion. *ACM Trans. Graph.*, 25(3):881–889, July 2006. ISSN 0730-0301.

Sang Il Park and Jessica K. Hodgins. Data-driven modeling of skin and muscle deformation. *ACM Trans. Graph.*, 27(3):96:1–96:6, August 2008. ISSN 0730-0301.

Bernd Bickel, Manuel Lang, Mario Botsch, Miguel A. Otaduy, and Markus Gross. Pose-space animation and transfer of facial details. In *Proceedings of the 2008 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, SCA '08, pages 57–66, Aire-la-Ville, Switzerland, Switzerland, 2008. Eurographics Association.