A Perception Based Metric for Comparing Human Locomotion

M. Pražák, R. McDonnell, L. Kavan and C. O'Sullivan *

Graphics, Vision and Visualisation Group, Trinity College Dublin, Ireland

Abstract

Metrics measuring differences between skeletal animation frames (poses) form the core of a large number of modern computer animation methods. A metric that accurately characterizes human motion perception could provide great advantages for these methods, by allowing the systems to focus exclusively on perceptually important aspects of the motion. In this paper we present a metric for human locomotion comparison, derived directly from the results of a perceptual experiment.

Categories and Subject Descriptors (according to ACM CCS): I.3.7 [Computer Graphics]: Virtual Reality, Perception

1. Introduction

Computer animation of humanoid characters encompasses a very large number of methods with applications ranging from musculo-skeletal simulations for biological research to Virtual Reality (VR) and computer games. Based on this scale, we can distinguish between two main groups of algorithms: the physically-based and the data-driven, with some methods using a combination of both. In this paper we consider only the latter case, because it represents the application domain of our technique.

Underpinning most data-driven methods of human animation is a metric, which allows the similarity between two motion clips and/or short segments of input motions to be determined. At the lowest level, two frames describing two poses are compared, with an extension to short motion segments achievable by simply averaging the similarity values of corresponding frames of the input segments (thereby accounting for *changes* of the motion as well, achieving similar results as comparing derivatives of the pose changes [KGP02]). For longer sequences, a timewarping technique is required to align the clip segments in the time domain (see Section 2).

In this paper, we present a metric which computes the difference between two frames (poses) of human locomotion and is directly derived from the results of a perceptual experiment. The narrow focus of the metric on human locomotion only is necessary for two reasons. First, it might be possible that metrics for different motion types would need different parameters, requiring a classification to take place before the comparison. Secondly, as we want to base our metric directly on a perceptual experiment, for experiment design we prefer a relatively small selection of clips of one type.

The contribution presented in this paper can be summarized as follows:

- a general framework for forming frame comparison metrics based on perceptual experiments,
- a metric created using this framework, describing the difference between locomotions of different subjects,
- an evaluation of the most common analytical frame comparison metrics and their suitability for expressing perceptual metrics, and
- an evaluation of different metric functions (Euclidean, manhattan and sigmoid) with respect to perceptual results.

In the next Section we describe previous work related to our method. Then we continue with construction of the motion database (Section 3), the perceptual experiment and its result (Section 4), the construction of the metrics (Section 5) and finally we summarize our results in Section 6.

* [prazakm,Rachel.McDonnell,kavanl,Carol.OSullivan]@cs.tcd.ie

2. Background

The computer animation of humanoid characters is a large field with a great number of techniques available (for a general overview please refer to [VWVBE*09]). As described above, our focus is aimed towards data-driven methods, which use motion metrics to determine similarities between frames and/or animation clips.

Data-driven animation
The data-driven methods can be divided into graph-based techniques, parameterization methods (and their combinations) and motion compression algorithms (used for minimizing the storage space required by large databases of motions).

The first group, graph-based techniques (or *motion graphs*), create a new animation by concatenating motion sequences of source animation clips [KGP02, LCR*02]. The animation space is described as an oriented graph with nodes representing short clips of animation and edges their transition points (detected using a comparison metric). The resulting animation is created by traversing the graph structure based on the decision of a higher-level planning algorithm.

Parameterization methods adopt a motion blending technique to generate new animations based on interpolation between clips from a database of motions. A *parametric space* structure, describing the interpolation scheme, is built from the annotation of input sequences with vectors of parameters describing their properties [WH97, RCB98]. This group of methods utilize the metrics to determine the set of motions belonging to a single group and to align the motions in the time domain using timewarping techniques. In a strict sense, each parameter of the parametric space describes a specialized metric as well, but its aim is to describe one aspect of the whole animation clip rather than provide the means to compare two frames.

The combination of the two previous methods leads to a more general description of motions, allowing the use of blending to create motions of a particular motion type, while allowing to transition between different types using the motion graph approach [RCB98, SO06, HG07].

Motion compression methods usually involve a form of dimensionality reduction based data correlation, employing statistical models, time-domain compression (wavelets) or their combinations [FHP07]. The use of a similarity metric in this group is mainly in database search and indexing.

Our metric aims to provide a way to compare two frames/sequences in a manner consistent with human perception of motion, which should be applicable for all the methods described above.

Motion metrics
As described above, a large number of methods use a form of motion metric to compare whole animation clips, short sequences or single frames.

Metrics comparing whole animations are aimed at motion type classification and database search, and their results should therefore be mostly independent of animation length. Applicable methods include bodypart-based approaches (e.g., [MRC05, OFH08]) and averaging of similarities computed from normalized motion clips (in a similar manner as for short sequences, see below). As our metric is expressing the differences between two frames, it is applicable mainly to the group of methods that compute the average of similarities.

In order to compare animation sequences using metrics that compute single frames, the input animations need to be normalized in both the time domain and in global space (to achieve independence to environment layout). Normalization in the time domain is usually achieved using timewarping (e.g., [RCB98]), preferably with a uniform base function [KG03]. Another approach would involve statistical models (e.g., Hidden Markov Models) to achieve similar results [LCR*02]. In this paper we use timewarping (similar in spirit to Kovar et al. [KG03]) to create a blending curve in motion comparison space, allowing us to compare clips of motions using a frame-based metric.

Frame-based metrics often use information from a single frame (one character pose) per animation clip, but can also incorporate dynamic information (e.g., velocity, acceleration) that may require access to several neighboring frames. Kovar et al. [KGP02] used a linear metric based on positions of points on the character skin, using a window of frames to account for dynamic information. The spatial alignment is addressed using a closed-form optimization alignment step. A similar approach, with points placed in joint locations and joint orientation represented by two more points, was used by Arikan [Ari06].

Lee et. al. [LCR*02] used a hierarchical skeleton structure, computing frame difference by finding the weighted angular difference between each pair of corresponding joints. The spatial alignment is achieved by simply leaving the global character position and orientation in global space (i.e., transformation of the root of the hierarchy). The problem of setting the weights of the averaging function was further addressed by Wang and Bodenheimer [WB03]. They used an optimization algorithm to determine a set of weights that provided maximal variance between different input data, and evaluated their result by a perceptual experiment. Our method uses the opposite approach. By building the metric directly from perceptual data, we want our metric to perform the same way as a human participant, thereby corresponding to perceptual data not only in variation, but also in other statistical properties defining a metric.

Human motion and perception
Perceptual studies of human motion form a well-established topic in the field of neuroscientific research. A common stimulus is a *point-light walker*, describing the motion pattern of an actor as a set of moving points. Even though

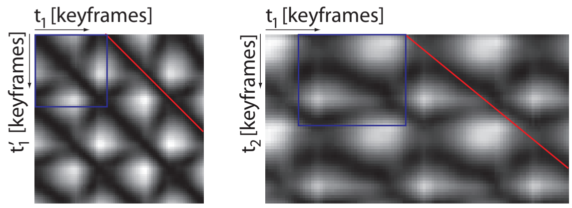

Figure 1: *Motion maps for the similarity metric between two animations with detected blending line (red) and one period of motion (blue). Left: the creation of a periodic clip formed by comparing the clip to itself (notice that there is no need for time synchronization, and the blending line has an angle of 45 °). Right: comparison between two different motion clips.*

this representation is extremely simplistic, the human visual system (HVS) can still, under certain conditions, interpret the trajectories as human motion, allowing motion perception to be explored independently of other visual information [Joh73].

The HVS performs well on these stimuli, recognizing not only a human walk [Joh73], but also often the walking subject [CK77]. Based on this evidence, Troje [Tro02] built a pointlight walker's parametric model based on a set of perceptual experiments, allowing perceptually-guided parameterizations to be created (such as "happy" or "nervous"). More recently, Giese and Poggio [GP03] performed a study uncovering the underlying principles of motion perception on pointlight walkers.

Perceptually based animation metrics became an active topic in computer animation research in recent years. Ren et al. [RPE*05] explored the possibility of defining a metric which would correspond to human perception of the naturalness of the motion. Closer to our topic, Wang and Bodenheimer [WB03] used an optimization technique to find the optimal weights for Lee's metric [LCR*02]. They also perceptually validated that their resulting function produced better transition points than the original weights proposed by Lee et al. Tang et al. [TKSL08] created a frame similarity metric based on concept of the joint relative distance.

Our approach differs mainly in the fact that we are trying to create a metric with perceptual evidence as its origin. As the metric is to be built from components based on commonly used metrics, we provide evaluation of each type and recommend the metric that produces the most accurate results.

3. Motion database

Our motion capture database used for this experiment contained walking motions of twenty one walkers (14 males and 7 females). The motions were captured using a 13 camera Vicon optical system, with 42 markers placed on body of the subject and a capturing framerate of 100 FPS (Frames Per Second). Each walker was instructed to walk up and down a straight line in our capture area until they felt comfortable and were walking at their natural pace. We then recorded a number of walks from each of them.

Captured data was then retargeted to a neutral wooden mannequin figure and converted into periodic clips by blending between two periods of the input motion. The period was detected by cross comparing the frames of the motion and detecting the minima of the resulting map. Because both periods of the motion are nearly similar, the resulting blending curve is a straight line (see Figure 1, left). By testing different metrics, we found that for this purpose all of them performed similarly, making the type of metric used unimportant.

4. Perceptual experiment and data analysis

Twenty-nine participants (18 male, 11 female) took part in our perceptual experiment. All participants were naive to the purpose of the experiment and had normal or corrected to normal vision. A matrix of 4 characters was displayed, with each character's motion different from the others and randomly chosen from the set of 21 (see Figure 2). Participants were asked to select two of the motions that they felt were most similar, by clicking a mouse. Sixty-three trials in total were shown to each participant, each providing 10 seconds for the selections. If they did not answer in the allotted time, an additional set of random motions was displayed on the screen. The 10 second time interval was chosen to avoid a long and detailed examination of the motion by participants, as we were more interested in the immediate reaction. The number of remaining trials was displayed on-screen be-

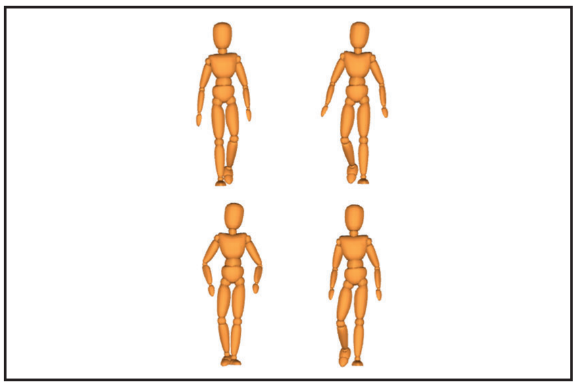

Figure 2: *Layout of the perceptual experiment. Users were asked to select two of the shown motions, which they felt were the most similar.*

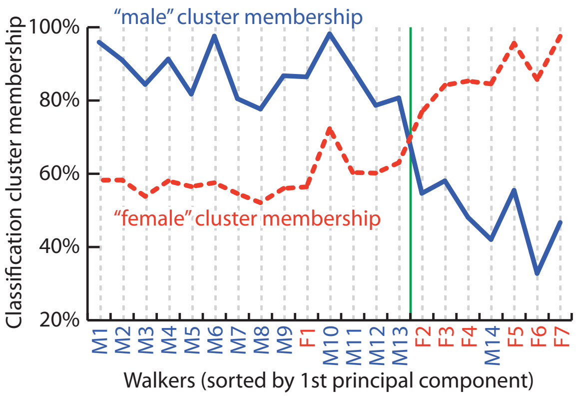

Figure 3: *Results of the classification performed on the data obtained in the perceptual experiment. The "male" (blue) and "female" (red) membership functions show the natural division of the motions by the gender of the actor. The order of actors on the horizontal axis is given by their projection onto first principal components of the source data (see Section 4). The green line shows the division line between "male" and "female" groups.*

tween each trial, making participants fixate on the center of the screen. For every trial, we recorded the four motions displayed and the chosen pair.

Motions were displayed on a neutral wooden mannequin figure, which was shown in previous studies to be gender-neutral (i.e., perceptual responses on the character's gender with applied animation correlate well with the gender of the subject originally performing the animation) [MJHN09]. As gender is one of the important factors in judging motion difference [KC77,Tro02], the gender neutrality of the character used in our study is very important.

The recorded results were then converted into a table, where the value in each cell represented the perceived difference between the two motions indexed by row and column number. The difference value was the ratio of the number of times the pair of animations was *not* perceived to be similar to the overall number of times they were shown together. The resulting value was therefore close to 1 if the motions were judged as very different, and 0 if they were very similar. The final table was symmetric, containing only values in the interval $[0,1]$ and with zeros on the diagonal, representing no difference when the motion is compared to itself. (Note: this table fulfills all the requirements of a metric with exception of triangle inequality, which was fulfilled for 98.6% of cases).

Further analysis of the results showed interesting properties of this table. To perform this analysis, we can interpret each row/column (because the matrix is symmetric, there is no difference between these two options) as a point in n-dimensional space (in our case $n = 21$). By performing Principal Component Analysis on this data and sorting the walks by their values projected on the first computed eigenvector, we achieve ordering that makes a clear distinction between gender. In the same manner we can apply a K-means clustering algorithm to the data with two clusters, receiving again a good distinction between the gender of the original motion captured walkers (see Figure 3). This shows an interesting property of the experiment results – the direction of the highest variance of the data is directly connected with the gender of walkers. Even though the importance of the gender in motion difference was already shown in previous work [KC77], the direct relation of gender with the direction of the highest variance is, to our knowledge, a new finding.

The distinction based on this analysis fails in two cases - female "F1" and male "M14" (Figure 3). The incorrect classification of male actor "M14" can be explained by his age, as this was the youngest male actor captured. The second misclassified walker, female "F1", was the same actor misclassified in a previous study [MJHN09]. As this is a phenomenon shown in a few previous studies as well (e.g., [KC77]), we can still assume that from perceptual point of view this classification is correct.

A graphical representation of the values in the source table, sorted in the way described above, is shown in Figure 4. The colours of the related motions appear darker, showing their lower difference and close relationship (with black on the diagonal meaning no difference), whilst light colours signify more distant motions.

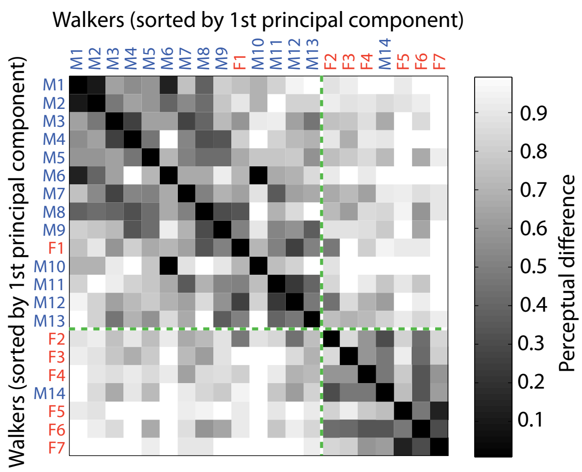

Figure 4: *Graphical representation of the data obtained in our perceptual experiment. A large value of the function signify large difference between two respective motions; the black diagonal means no difference when a motion is compared to itself. The green dashed line shows the threshold between the "male" and "female" clusters from Figure 3*

		Difference	Constant	Normalization Constant	Head	Clavicle	Upper Arm	Forearm	Hand	Thigh	Calf	Toe
Angles	euclid	23.335	0.508	659.272	0.000	0.000	0.006	0.000		0.669	0.325	
	manh	18.448	0.324	19.537	0.041	0.003	0.061	0.000		0.622	0.273	
	sigm	18.004	0.312	20.016	0.039	0.008	0.076	0.000		0.614	0.263	
Positions	euclid	17.117	0.290	206.509	0.062	0.000	0.551	0.181	0.000	0.000	0.050	0.156
	manh	13.423	0.069	9.909	0.020	0.000	0.362	0.429	0.000	0.000	0.000	0.189
	sigm	13.264	0.059	10.375	0.015	0.000	0.344	0.455	0.000	0.000	0.000	0.184
Velocities	euclid	25.779	0.431	1914.283	0.000	0.000	0.000	0.000	0.605	0.000	0.298	0.098
	manh	22.039	0.193	36.885	0.000	0.000	0.000	0.000	0.551	0.000	0.298	0.151
	sigm	21.993	0.189	36.532	0.000	0.000	0.000	0.000	0.564	0.000	0.306	0.157

Figure 5: *Table of results of our fitting algorithm, showing the best matches (difference column) and corresponding weight values for each metric type and shaping function.*

5. Motion metrics toolkit

With the result of the perceptual experiment (above), we try to define analytical metrics to reproduce the perceptual results as closely as possible, by changing the parameters of existing methods using an optimization technique described below. The resulting metric could then be used in any data-driven animation method, allowing human perception of the motion difference to be more closely mimicked.

Using the original metrics of Lee et al. [LCR*02] and Kovar et al. [KGP02] as a starting point, we define the distance functions 1 to 3 below, where c represents a scalar displacement constant, F is a measurement function (e.g., power of 2 would be the squared Euclidean distance), m is the number of joints and $q_{a,b}$ describes a rotation quaternion in frame a with index b, with similar notation for joint positions and velocities. Weights w_k for joints $k = 1 \dots m$ describe the estimated saliency of each joint, i.e., its relative perceived importance using the assigned measurement function.

$$d_{angles}(f_i, f_j) = c + \sum_{k=1}^{m} w_k F(q_{i,k} \cdot q_{j,k}) \quad (1)$$

$$d_{positions}(f_i, f_j) = c + \sum_{k=1}^{m} w_k F(\|p_{i,k} - p_{j,k}\|) \quad (2)$$

$$d_{velocities}(f_i, f_j) = c + \sum_{k=1}^{m} w_k F(\|v_{i,k} - v_{j,k}\|) \quad (3)$$

The final function then consists of a weighted average of components determined by the equations above. We can display each of these components in a separate motion map in a manner similar to that for whole clips of motion (e.g., Figure 1). Each component has unique properties, as can be seen in Figure 6, with each type of measured physical property having its own unique pattern. We are aiming to combine these patterns in a manner closely resembling the results of our perceptual experiment.

In order to achieve this, the metric itself was looked upon as an n-dimensional function, with every dimension describing one of the weight parameters w_k. To reduce the number of dimensions, we used the same weights for symmetrically corresponding bones, leading to a reduction of the parameter count to half of the original number and thus speeding up the fitting process. The difference function was described as L_2 (Euclidean) norm between the experimental data and the table created using the fitted component, with all dimensions cropped above zero (as negative weights do not have any physical meaning). This formulated the fitting problem as a non-negative least squares optimization. Because the parameter space is mostly smooth, a classical random start gradient descent algorithm leads to a successful solution for most starting points (tested on 1000 random points for each case). The results of the best fits are shown in Figure 5.

Four different options for distance function F were evaluated: power of two (squared Euclidean); absolute value (Manhattan); Sigmoid function ($d/(d+1)$ with d computed using Euclidean metric); and power of n. The fitting algorithm proved to be effective in the first 3 cases, as shown in Table 1. From these results we can conclude that it is not necessary to use all bones of the skeleton, as some have no effect on the result. The most obvious impact is on the velocities

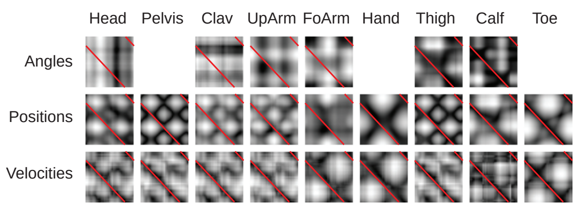

Figure 6: *Components of the metric toolkit described in terms of motionmaps (in a similar manner to Figure 1).*

term describing the dynamics of motion, which actually includes only hands, thighs and calves. On the other hand, for the angle and position distance functions, which describe the static posture properties of the skeleton, the most important feature of the arm is the elbow, as already shown by [Tro02]. The sigmoidal distance function matched human perception most closely, with the Manhattan metric a close second. Both were significantly better than the classical squared Euclidean metric. For the general power of n distance function, with n as one of the fitting parameters, the fitting algorithm did not prove to be efficient, although for $n \approx 0.3$ it produced even better results than the other tested functions.

6. Conclusions and Future Work

The development of a more general metric capable of emulating human perception of motion differences seems feasible from our results. In this light, our results can be seen as a first step towards developing a mathematically specified metric with the desired perceptual properties. We plan to further evaluate more different types of metric functions and measure the feasibility of different optimization methods for automatic specification of weights based on perceptual principles. We also plan to extend the approach to different types of motions.

References

[Ari06] Arikan O.: Compression of motion capture databases. *ACM Transactions on Graphics (TOG) 25* (2006), 890. 2

[CK77] Cutting J., Kozlowski L.: Recognizing friends by their walk: Gait perception without familiarity cues. *Bulletin of the Psychonomic Society* (1977). 3

[FHP07] Faloutsos C., Hodgins J. K., Pollard N. S.: Database techniques with motion capture. *International Conference on Computer Graphics and Interactive Techniques* (2007). 2

[GP03] Giese M. A., Poggio T.: Neural mechanisms for the recognition of biological movements. *Nature Reviews Neuroscience 2003* (2003). 3

[HG07] Heck R., Gleicher M.: Parametric motion graphs. *Proceedings of the 2007 symposium on Interactive 3D graphics and games* (2007). 2

[Joh73] Johansson G.: Visual perception of biological motion and a model for its analysis. *Perceiving events and objects* (1973). 3

[KC77] Kozlowski L., Cutting J.: Recognizing the sex of a walker from a dynamic point-light display. *Perception & Psychophysics* (1977). 4

[KG03] Kovar L., Gleicher M.: Flexible automatic motion blending with registration curves. *Symposium on Computer Animation* (2003). 2

[KGP02] Kovar L., Gleicher M., Pighin F.: Motion graphs. *ACM Transactions on Graphics (TOG) 21* (2002). 1, 2, 5

[LCR*02] Lee J., Chai J., Reitsma P. S. A., Hodgins J. K., Pollard N. S.: Interactive control of avatars animated with human motion data. *ACM Transactions on Graphics (TOG) 21* (2002). 2, 3, 5

[MJHN09] McDonnell R., Jörg S., Hodgins J., Newell F.: Evaluating the effect of motion and body shape on the perceived sex of virtual characters. *ACM Transactions on Applied Perception (TAP)* (2009). 4

[MRC05] Müller M., Röder T., Clausen M.: Efficient content-based retrieval of motion capture data. *ACM Transactions on Graphics (TOG) 24* (July 2005), 677. 2

[OFH08] Onuma K., Faloutsos C., Hodgins J. K.: FMDistance: A fast and effective distance function for motion capture data. *Short Papers Proceedings of EUROGRAPHICS* (2008). 2

[RCB98] Rose C., Cohen M. F., Bodenheimer B.: Verbs and Adverbs: Multidimensional Motion Interpolation. *IEEE Computer Graphics and Applications 18* (1998). 2

[RPE*05] Ren L., Patrick A., Efros A. A., Hodgins J. K., Rehg J. M.: A data-driven approach to quantifying natural human motion. *ACM Transactions on Graphics (TOG) 24* (2005). 3

[SO06] Shin H., Oh H.: Fat graphs: Constructing an interactive character with continuous controls. *Proceedings of the 2006 ACM SIGGRAPH/Eurographics symposium on Computer animation* (2006). 2

[TKSL08] Tang J., Komura T., Shum H., Leung H.: Emulating human perception of motion similarity. *Computer Animation and Virtual Worlds, 19* (2008). 3

[Tro02] Troje N.: Decomposing biological motion: A framework for analysis and synthesis of human gait. *Journal of Vision* (2002). 3, 4, 6

[VWVBE*09] Van Welbergen H., Van Basten B., Egges A., Ruttkay Z., Overmars M.: Real Time Animation of Virtual Humans: A Trade-off Between Naturalness and Control. In *Eurographics* (2009), p. 45Ű72. 2

[WB03] Wang J., Bodenheimer B.: An evaluation of a cost metric for selecting transitions between motion segments. *Symposium on Computer Animation* (2003). 2, 3

[WH97] Wiley D., Hahn J.: Interpolation synthesis of articulated figure motion. *IEEE Computer Graphics and Applications* (1997). 2