graphics • vision • visualisation

# Towards a Perceptual Metric for Comparing Human Motion

**Martin Pražák, Rachel McDonnell, Ladislav Kavan, Carol O'Sullivan**
Graphics, Vision and Visualisation group, Trinity College Dublin, Ireland

## Introduction

Most of the commonly used approaches for editing human motion, such as motion graphs and motion blending, use some form of distance metric in order to compare character poses in keyframes. These metrics utilize a combination of three traditional methods - joint angular differences, distances between points on an object and velocities of specified bodyparts.

The presented method attempts to find a metric and its parameters (not limited to the usual Euclidean metric), which would match a dataset formed by a direct perceptual experiment as closely as possible. Previous methods used peception for evaluation alone, but we use perception as the basis of our metric.

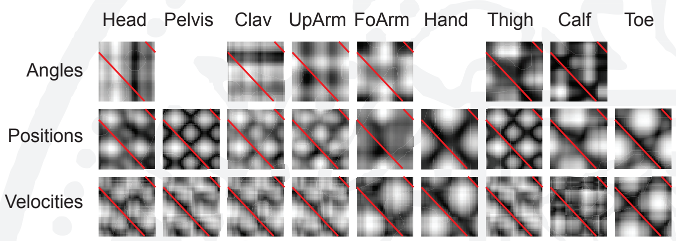
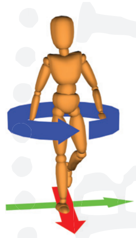



Figure 1: *Maps of body part components, red line indicates the best detected transition*

## Motion Metrics



In order to compare two different motion clips a pose alignment needs to be performed as the very first step. Following the method of [KGP02], we change only the X and Z character positions and their orientations along the Y axis (yaw) of one of the clips.

Using the original three metrics as a starting point, we define the distance functions 1 to 3 below, where *F* is a measurement function (e.g., power of 2 would be 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.

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

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

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

## Motion database and data preprocessing

The first step of motion processing is to produce one normalized period of the walk in order to make sure that the processed motion clips results are directly comparable. This can be achieved by comparing the animation with itself, forming a form of autocorelation function, and detecting its local minima.

In order to compare two different walks, they must be first aligned with the coordinate system as described above, and then aligned in the time domain by evaluating each of their 2D comparison functions and selecting the best blending curve between them.

## Perceptual experiment and data analysis

A perceptual experiment was conducted in order to gather information on human perception of motion similarity. A matrix of four different motions from a set of 21 was displayed and participants were asked to select two of the motions they felt were the most similar by clicking the mouse.

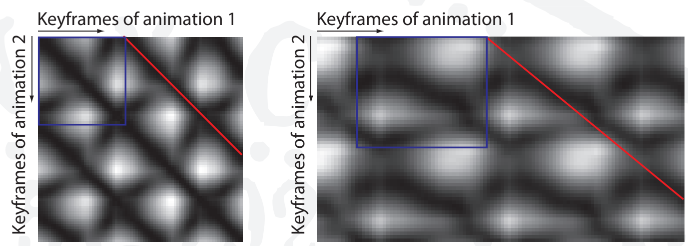



Figure 2: *Left - detection of one period of motion, right - blending curve and comparison window demonstration*

Data gathered from this experiment formed a table describing a "distance" between each pair of animation. Because this table fulfils the requirements for a distance function usable in a metric, it was then used as the target for fitting algorithm.

The fitting algorithm was a simple n-dimensional random-start gradient-descent algorithm. At least 1000 fitting trials were computed for every case in order to achieve the best possible parameters settings for a particular combination of distance function and metric type.

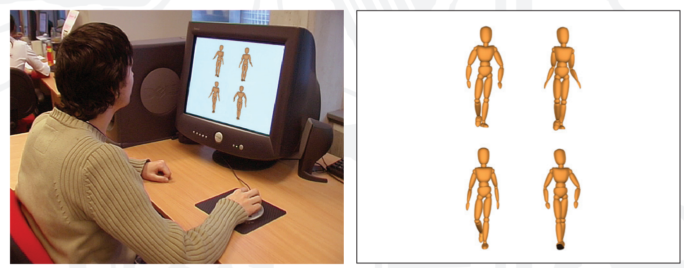


Figure 3: *The perceptual experiment layout*

## Results and future work

Four different options for the distance function *F* were evaluated: power of two (Euclidean distance), absolute value (Manhattan), Sigmoid function and a power of *n*. The fitting algorithm proved to be effective in the first 3 cases. ***From these preliminary results we can conclude that it is not necessary to use all bones of the skeleton, as some have no correspondence to the perception of motion similarity.***

As regards the distance function, the ***sigmoidal metric matched human perception most closely***, with the Manhattan metric a close second, while both signifficantly better than the usually considered Euclidean metric. For a more general power of *n* distance function, with *n* as one of the fitting parameters, the fitting algorithm did not prove to be sufficiently efficient. Although, interesting to note is that for $n \approx 0.3$ it produced an even closer fit than the other tested functions.

This is still a work in progress, so if you have any feedback, please do tell us! :)

## References

[KG03] Kovar L., Gleicher M.: Flexible automatic motion blending with registration curves. In *Proc. SCA '03* (2003), pp. 214-224.

[KGP02] Kovar L., Gleicher M., Pighin F.: Motion graphs. *ACM Trans. Graph. 21*, 3 (2002), 473-482.

[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 Trans. Graph. 21*, 3 (2002), 451-500.

[Tro02] Troje N. F.: Decompositing biological motion: A framework for analysis and synthesis of human gait patterns. *Journal of Vision 2*, 5 (2002), 371-387.

[WB08] Wang J., Bodenheimer B.: Synthesis and evaluation of motion transitions. *ACM Trans, Graph. 27*, 1 (2008)