Received: 21 April 2023 | Accepted: 1 May 2023

DOI: 10.1002/cav.2157

SPECIAL ISSUE PAPER

WILEY

Correlation as a measure for alignment and similarity of human motions

Mathew Randall | **Carlo Harvey** | **Ian Williams**

DMTLab, Birmingham City University, Birmingham, UK

Correspondence
Mathew Randall, School of Computing and Digital Technology, Birmingham City University, Birmingham B4 7XG, UK.
Email: mathew.randall@bcu.ac.uk

Abstract

The ability to measure similarity and alignment of motions is a key tool in motion retrieval and motion editing. Similarity metrics based on distance functions are often utilized when measuring similarity of human motions, however, metrics based on correlation can also potentially useful for measuring similarity and alignment. This paper evaluates the use of correlation as a method of measuring the alignment and similarity of human motion and compares them against more established distance-based metrics. Three correlation methods and five methods of parameterising rotation are evaluated. The results show that parameterization based on displacement vectors and Kendall Tau rank correlation are optimal for measuring the alignment between two motions. If measuring similarity of motions, however, an approach based on distance metrics for angular or positional distance should be used.

KEYWORDS

human motion, motion similarity, motion alignment

1 | INTRODUCTION

Many techniques and approaches to re-using and combining motion captured data are dependent on reliable and accurate methods of time warping a motion, to temporally align it with another motion, using techniques such a Dynamic Time Warping (DTW). In the field of motion synthesis, the temporal alignment of motions allows styles and emotions to be transferred between captured performances,[1] and motions to be accurately blended together.[2] While in the field of interaction, time warping is used in motion training[3] and movement analysis.[4]

To support these applications, an accurate metric is required to measure the temporal alignment of two motions, both to identify motions in a database with the best alignment, and to evaluate the performance of time warping algorithms. Existing metrics typically use approaches based on Euclidean distance[2,5] to search motion databases for similar motions which contain matching movements, rather than aligned motions, which temporally match each other. Meanwhile, correlation has been used to evaluate time warping algorithms,[6] using it to measure the alignment between the resulting warped input motion, and the target motion being aligned to.

This paper evaluates correlation as a method for measuring both alignment and similarity, comparing it against more established distance based similarity metrics. A number of different correlation based similarity metrics are implemented and tested, using alternative approaches to parameterize joint angles and measure correlation.

To support this study, statistical and cluster analysis techniques are used to determine which distance or correlation based similarity metric is optimal for measuring: (i) how accurately the temporal features of a pair a motions are aligned, and (ii) how similar a pair of motions are to each other.

2 | RELATED WORK

Measuring how accurately the temporal features of two motions align is a non-trivial problem. Motion data is multidimensional in nature, for example a simple joint rig of 18 joints, each encoded with 3 parameters, would use 54 dimensions to define a character's pose in a given frame. Any measurement of alignment of two motions is therefore based on an aggregated best fit of multiple time series representing multiple joints, however it is important to focus measurements on joints which are salient to human motion. Lee et al.[7] suggest focusing similarity metrics on joints which have more impact on the general pose of a motion, such as: shoulders; elbows; hips; and knees, and ignore less significant end effector joints such hands and feet. Optimal joint weightings determined using regression, agreed with this selection of joints.[8]

Metrics for measuring similarity should also take into account both the posture and dynamics of human motion, many researchers measure similarity using distance functions based on angular, positional and velocity differences between the two motion's respective joints. Quaternions are used to measure the geodesic distance between joint angles,[9] while Euclidean approaches are used to measure differences in position and velocity, either in global space[10] or relative to the hip joint.[5] An alternative, hip joint invariant approach, plot point clouds of the position of each joint in each frame within a window of time, then measure the difference between the two set of points.[2]

Rather than use a distance based metric, Etemad and Arya, used a metric based on linear correlation, utilizing Pearson's Correlation Coefficient (PCC), to measure how well two motions align.[6] They also discuss the shortcomings of distance based metrics, pointing out how they are more affected by signal noise or any offset between the motion signals being compared, than a correlation based approach. As well as linear correlation this paper considers the use of the methods Spearman's and Kendall's rank correlation coefficients,[11] which are more suited to working with non-parametric data. Although rank based correlation methods have not previously been used to measure the similarity or alignment of two motions, they have proven effective in recognizing human motions from movement sensors.[12]

Rather than evaluating the alignment of two motions based on the similarity of the motions performed, Folgado et al.[13] proposes a metric based on the similarity of the alignment paths used to time warp the motions, allowing a given alignment path to measure against a gold standard. However, this approach relies on an optimal alignment already being established to compare against.

The performance of a given similarity metric can be evaluated by measuring its ability to distinguish between aligned and non-aligned motions or between similar and dissimilar motions. Chan et al.[3] used a statistical approach to determine which of the three distance measures: joint position, velocity and angle are best for measuring the similarity of dance motions. For each measure, the distribution plots of similar and dissimilar motions were visually compared, to identify which measure produced the least overlap, and therefore discriminated best between similar and dissimilar plots. This paper proposes using a non-parametric t-test to accurately measure the overlap. Valik et al.[14] utilizes a number of search algorithm evaluation tools, such as Mean Average Precision (MAP), to evaluate the ability of different similarity metrics to correctly retrieve similar motions from a motion database.

This paper proposes a number of similarity metrics based on different methods of correlation and angle parametrization. The performance of these metrics will be evaluated and compared with established distance base similarity metrics, assessing their ability to measure both the similarity and alignment of two motions.

3 | MEASURING SIMILARITY

3.1 | Preparing motion data

Regardless of the method used to parameterize joint angles, each parameter can be treated as a time series of n data points, where n is the number of frames as defined by $f = 1, \dots, n \in \mathbb{N}$. The similarity of two motion sequences $(a_1, a_2, \dots, a_n)$ and $(b_1, b_2, ..., b_n)$ can be measured using either, a cost based approach measuring the distance between each respective data point or correlation approach measuring the overall correlation between the two time series.

To facilitate these similarity measurements both motion sequences must have the same number of frames (n), which typically requires the length of one the motions to be adjusted. This can be achieved by using Uniform Time Warping (UTW), to uniformly stretch or squash a motion, or Dynamic Time Warping (DTW) to monotonically align the frames of an input motion to best match the frames of a target motion.

Both correlation and distance based approaches, can only reliably compare motions applied to the same joint system or skeletal model. Differences in the joint systems will affect the joint angles used to encode character poses. This means that an identical pose on two different joint systems can be defined using different joint angles.

3.2 | Distance based similarity metrics

Distance based metrics evaluate the similarity of two motions by extracting an identical set of parameters from the joints of both motions, then measuring the amount of difference between the respective parameters in a cost function. In this paper four distance based similarity metrics are evaluated based on position, angle, velocity and point cloud parameters. Although not considered in this study, distance based similarity metrics that combine position, velocity and accelerating have been proposed.[5]

A common feature to use in distance base metrics is difference between the orientations of corresponding joints within the two motions. The distance between the angular rotation of two joints can be determined by representing the rotations as quaternions and taking the absolute value of their inner product, $|q_a \cdot q_b|$, giving the geodesic distance between them. The geodesic distance is the length of a curve representing the shortest path between two angles when plotted onto the surface of a sphere. The angular distance between motion sequences a and b, can be determined using Equation (1), where both motions contain n frames, and the similarity is based on set of joints m.

$$c_\theta = \sum_{f=1}^{n}\sum_{j=1}^{m}\frac{2}{\pi}\arccos|q_{ja}^{(f)} \cdot q_{jb}^{(f)}|. \tag{1}$$

When measuring the alignment or similarity of two motions using position or velocity, it is often more desirable to parameterize positions in local space, relative to the hips, than in global space. This eliminates the need to align motions before measuring the similarity. Moreover, most applications that require aligned or similar motions, such as: motion blending; motion recognition and translating motion styles, do not have an explicit need for the hip joint to be considered. To determine the position of a joint with respect to the hip, the transform of the hip joint needs to be removed or reversed, by multiplying the inverted transform matrix of the hip joint (H^{-1}) by the global position of the joint j, (G_j). The similarity between motions a and b, based on the positions of their joints relative to their root, can be determined using Equation (2), where $\vec{t}(M)$ is a vector representing the translation element of a given transform matrix M and $|\vec{v_j}|$ is magnitude of vector, $\vec{v_j}$, which represents the positional difference in the local space of joint, j, between motions a and b at a given frame, f.

$$c_p = \sum_{f=1}^{n}\sum_{j=1}^{m}|\vec{v_j}|,$$
$$\text{where } \vec{v_j} = \vec{t}(H_a^{-1}G_{ja}) - \vec{t}(H_b^{-1}G_{jb}). \tag{2}$$

A velocity distance metric can be based on either positional or angular velocity. Equation (3) demonstrates how the positional distance metric in Equation (2), can be adapted to create a metric measuring the difference in positional velocity, where $\vec{\Delta}_j$ is vector representing the velocity of joint j in local space.

$$c_{pv} = \sum_{f=1}^{n}\sum_{j=1}^{m}|\vec{\Delta}_{ja} - \vec{\Delta}_{jb}|,$$
$$\text{where } \vec{\Delta}_j = \vec{t}(H_f^{-1}G_{jf}) - \vec{t}(H_{f-1}^{-1}G_{j(f-1)}). \tag{3}$$

Point clouds of joint positions at a given frame and its neighboring frames, represent both the pose (position) and movement (velocity and acceleration) of a motion over a small window of time.

The distance between two sets of point clouds can be determined using Equation (4), giving a measure of similarity, where w is the number of neighboring frames to be included each side of the current frame, α_p is the weighting for the point in the point cloud and p_a and p_b are points in the point cloud created from motion a and b respectively. Transform (T)[2] is applied to the point cloud of one motion, optimizing the fit (sum of the squared distances) between the two sets of point clouds.

$$c_{pc} = \sum_{f=1}^{n} \sum_{p=1}^{m(2w+1)} \alpha_p |p_a - T_{\theta,x,z} p_b|. \tag{4}$$

3.3 | Correlation based similarity metrics

A weakness of the distance based approach to measuring similarity, is the way in which it can be overly affected by any difference in the offset or amplitude of two signals in comparison to correlation based approaches. Figure 1 shows a normalized signal of a joint parameter with offset, multiplication, gamma and noise applied. Each of these transforms represent ways in which a repeated action or motion can vary between performance captures. A performer spatially missing a marked position would result in an offset, while a more exaggerated performance would result in some joint rotations being extended or amplified. Gamma represents the way in which a movement starts and stops, akin to the ease-in and ease-out of key-frame animation, this could be considered a pseudo representation of acceleration, which is also affected by the exaggeration of a performance. Just like any digital recording process, a motion captured performance will also contain a certain level of noise, this noise will vary between each recording.

Despite the original (blue) and transformed (red) signals being very similar or identical in form, the distance based approach considers the two signals to be different, whereas the Pearson, Spearman and Kendall Tau correlation based approaches consider the signals to be very similar. While the distance based approach measures the similarity of each data point, correlation approaches consider the similarity between the overall forms of the two signals.

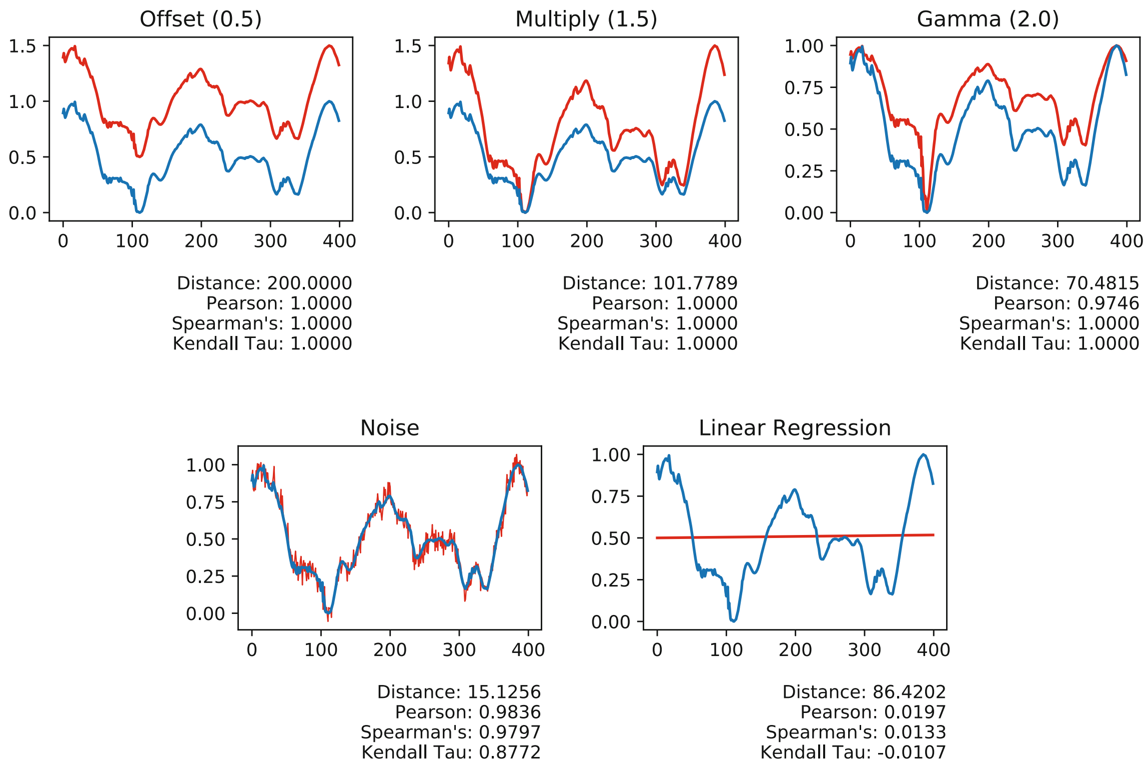


FIGURE 1 A comparison of the effects of different signal transforms on distance and correlation based measurements of similarity. Despite the original signal (blue) and transformed signal (red) being similar in form, the distance based metric considers them to be dissimilar.

The distance based metric considers the straight linear regression line to be more similar to the original signal than the offset or amplified versions of the signal, despite the line being a completely different shape. This is in contrast to the correlation methods which all gave the offset and multiplied versions of the signal a score of one, which is the same score two identical signals would achieve. Within this study, similarity metrics utilizing three different commonly used correlation methods are compared. One linear correlation method, PCC, and two rank based methods Spearman's and Kendall Tau.

Unlike PCC, neither of the rank based correlation coefficients are effected by a gamma transform, as shown in Figure 1. Although a gamma transform results in a non-linear relationship between the original and transformed signal values, the rankings of these values are unaffected by the transform and are therefore identical for both signals.

The Kendall Tau correlation coefficient is particularly sensitive to changes in direction (i.e. the derivative of a signal changing from positive to negative or vice versa), occurring in one signal and not the other, as this produces discordant pairs in the ranked values. This sensitivity can see in the noise plot within Figure 1.

Unlike the distance based metric, all of the correlation methods discussed here provide normalized results, where $r \in \mathbb{R} : -1.0 \leq r \leq 1.0$, regardless of how large the data series values are. This is useful when aggregating the results of many similarity scores, as is often the case when working with motion data, and avoids the need to normalize data across an entire data-set.

3.4 | Parametrising joint angles

Motion data is encoded using joint lengths and joint rotations, rather than joint positions. While joint lengths remain constant through-out the motion, often only being specified at the beginning of a motion data file, the joint rotations for each joint are specified for every frame of a motion sequence. Apart from the root joint, joint rotations are specified with respect to a parent joint in their own local space, rather than all being specified within a single global space.

This study utilizes five different approaches to parameterising joint angles for use with correlation based similarity metrics. Three established approaches: Euler, Quaternions and rotational matrices, and two less commonly used approaches: displacement vectors and logarithmic maps. The later two methods both express three dimensional angles as $\mathbb{R}3$ vectors in a form that avoids gimbal lock.

Displacement vectors represent an orientation as unit vector $V(x, y, z)$. A displacement vector, v' for a given joint, can be derived by multiplying the joint's rotation matrix R by unit vector $v = [0, 1, 0]$, as follows $v' = vR$. This representation loses the *roll* dimension of the angle but has previously been used with PCC to measure the correlation[6] between motions.

A number of different approaches have been explored using logarithmic maps to convert quaternions to an $\mathbb{R}3$ vector representation, that avoids gimbal lock, and still specifies the axis and magnitude of the rotation.[15] These angular representations are considered particularly useful for machine learning or deep learning applications. Logarithmic maps exploit the fact that when a logarithm is applied to a quaternion, the real component w becomes zero, resulting in the definition $log\ q \equiv \begin{bmatrix} 0 & x & y & z \end{bmatrix}$. Equation (5) uses a natural logarithm as a logarithmic map to convert quaternion q into a linear vector, where $q\mathfrak{R}$ and $q\mathfrak{I}$ are the respective real and imaginary vector components of q.[15]

$$ln\ q = \frac{1}{sinc(arccos(q_{\mathfrak{R}}))} q_{\mathfrak{I}}. \tag{5}$$

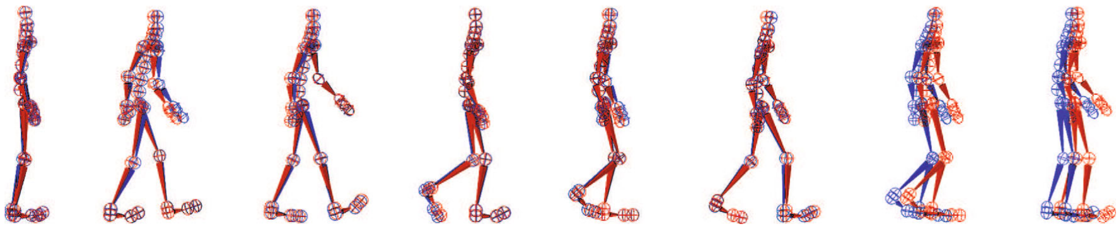
Any given angle can be represented by two quaternions, as $q = -q$. This double covering or antipodal symmetry can be visualized as two hemispheres, where all three dimensional angles can be specified within each hemisphere. Before converting a set of quaternions into a logarithmic map, a hemispherization process needs to be performed to ensure that all the quaternions for a given joint in both motion a and b, are specified within the same local hemispherical space.[15]

4 | METHODOLOGY

4.1 | Sourcing the data-sets

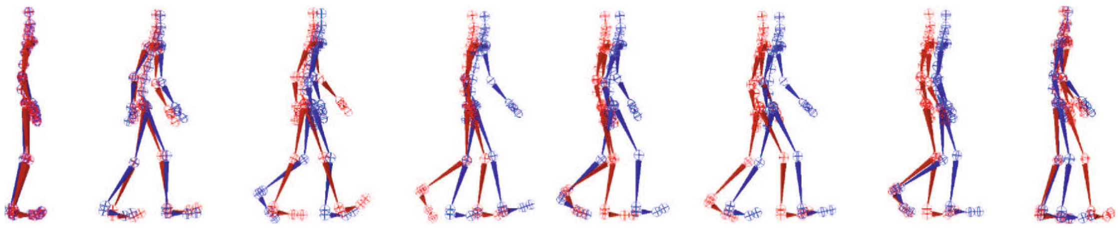
To facilitate this study a number of data-sets, consisting of captured human motions, were required. Each data-set consists of pairs of motions each containing two motions that were either aligned, similar or dissimilar as described in Figure 2. In total four data-sets of motion pairs were required as follows:

(a)



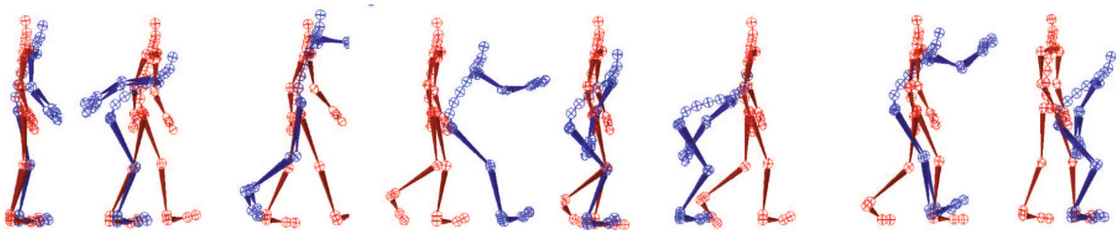
Aligned motion pair: two recordings of the same motion with one motion aligned to the other. This motion pair consists of blue walking motion that has been aligned to a red walking motion.

(b)



Similar motion pair: two recordings of the same motion that have not been aligned. This motion pair consists of two walking motions.

(c)



Dissimilar motion pair: two recordings of different motions. This motion pair consists of jumping motion (blue) and walking motion (red).

FIGURE 2 Examples of the three different types of motion pairs used in this study.

- To evaluate the performance of similarity metrics at measuring the alignment of two motions, two data-sets, one consisting of, similar aligned pairs of motions, and the other, similar unaligned pairs of motions, were required. These data-sets are referred to as *Aligned* and *Non-Aligned* respectively.
- To evaluate the performance of similarity metrics at measuring the similarity of two motions, two data-sets, one consisting of similar motion pairs and the other dissimilar motion pairs, were required. These data-sets are referred to as *Similar* and *Dissimilar* respectively.

A set of 63 motions were captured, consisting of 21 different movements each captured three times. The movements consist of common actions such walking, jumping, sitting and picking up objects. Each movement was carefully choreographed to achieve a high level of similarity between each of the three recordings. Care was taken to match the number of steps, make sure the motions started and ended on the same foot and markers were set out on the floor for the actor to hit with each step. The motions were all captured during a single capture session using the same actor.

The motions were captured using a Vicon motion capture system consisting of eight 2.2 megapixel Vero 2.2 cameras recording at 120 frames per second. The system was configured in 7×7 m volume with a level floor. None of the motion captured contained dropped frames and no manipulation or clean-up of the data was performed. This data-set of motions has been made freely available to the public.[16]

4.2 | Preparing and generating data-sets

The captured motions were organized into the required four data-sets as follows:

- *Aligned data-set*: consists of sets of two *Aligned* motions, in which one motion has been aligned to another similar motion using DTW. Each unique two way combination of two similar motions was used, with each set of three recordings of the same motion generating six sets, producing a data-set of 126 pairs of motions.
- *Non-aligned data-set*: consisting of *Non-Aligned* pairs of motions. This data-set is made up of each unique combination of two similar motions, plus the aligned motions from the *Aligned* data-set above, each combined with the similar motion that was neither used as the input or target motion for the alignment. Each set of three recordings of the same motion generated 9 sets, producing a total data-set of 189 pairs of motions. The motivation for including aligned motions was increase the size of the data-set.
- *Similar data-set*: consisting of a mixture of *Aligned* and *Similar* pairs of motions. For each set of three recorded motions, each unique combination of the three recorded motions and six aligned motions from the *Aligned* data-set (9 in total) was used. For each set of three recordings of the same motion, this generated 36 sets, producing a total data-set of 756 pairs of motions. As with the data-set above, the motivation for including aligned motions was to increase the size of the data-set.
- *Dissimilar data-set*: consisting of pairs of recordings of two different *Dissimilar* movements. Two recordings of each movement were combined with two recordings from the other 20 movements, to create a data set of 840 pairs of motions. Only two of the three recordings of each movement were used to avoid an overly large data-set that might bias results.

4.3 | Implementation of DTW

The motions within Aligned data-set were prepared using DTW to align the temporal features of an input motion to match that of a target motion. The distance between the joint poses in the frames input and target motions were measured based on rotational distance, determined using Equation (1) populating a cost matrix, C. The costs in matrix C were accumulated to create matrix T, starting at $T_{0,0}$ and ending at $T_{m-1,n-1}$, using Equation (6), where m and n are the number of frames in the input and target motions respectively.

An alignment path was then plotted backwards through the accumulated cost matrix T from $T_{m-1,n-1}$ to $T_{0,0}$, enforcing a boundary condition. Starting at $T_{m-1,n-1}$, Equation (7) is used to plot an alignment path through the accumulate cast matrix, T, where p' and p represent the coordinates in matrix T of the next and most recent points on the alignment path respectively. No local continuity or global constraints were applied to constraint the path that could be plotted.

$$T_{i,j} = C_{i,j} + min\{T_{i-1,j}, T_{i,j-1}, T_{i-j,k-j}\}, \tag{6}$$

$$p'(i,j) = arg\ min\{T_{p_{i-1},p_{j-1}}, T_{p_{i-1},p_j}, T_{p_i,p_{j-1}}\}. \tag{7}$$

4.4 | Joint weightings

With joint rotations being parameterized in local space within this study, joints at the end of the kinematic chain are of less significance. Therefore, the time-warping and similarity measurements within this study were performed using a subset of equally weighted joints as shown in Figure 3. The subset consisted of both the left and right: shoulders, elbows, hips and knees.

4.5 | Measuring motion similarity

The similarity of the motion pairs in every data-set were measured using 19 different similarity metrics (15 correlation and 4 distance base metrics). The PCC, Spearman's and Kendall Tau correlation methods were each implemented with

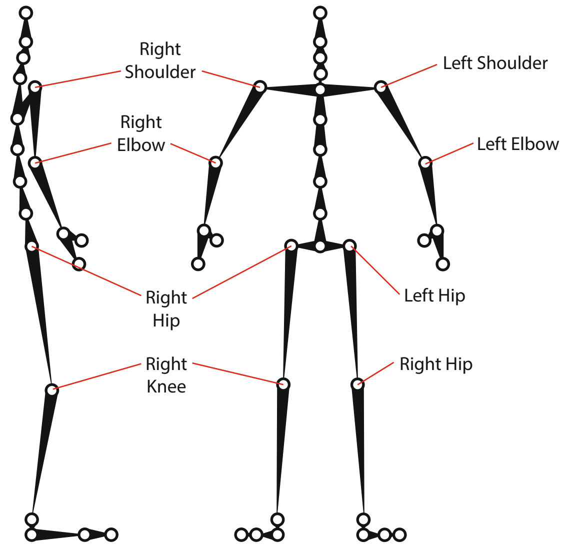


FIGURE 3 Configuration of joints used in study.

five different joint angle parametrization methods (Euler, quaternion, matrix, displacement vector and logarithmic map) as described in Section 3.4, giving a total of 15 different correlation based similarity metrics. Four different distance based similarity metrics were also implemented based on: joint angles, joint position, joint velocity and point clouds as described in Section 3.2. The point clouds were generated using a window size of seven, encompassing three frames immediately before and after the frame being sampled. As the focus of this study is to compare different measurements of similarity, compound approaches which combine multiple different similarity measurements were avoided.

To avoid unutilized axes, resulting from joints with limited degrees of freedom affecting the results of correlation based similarity metrics, any axis containing only low amplitude noise below 1×10^{-5} was given a correlation score of 1. In order to weight each joint equally, the distance based similarity metrics based on angle, position and velocity, require the similarity scores (i.e., distance) for each joint to be normalized across the data-set.

4.6 | Evaluating similarity metrics

The performance of each similarity metric was evaluated using two different performance tests, Overlap and MAP, which measure the ability of each similarity metric to distinguish between aligned and non-aligned motions or between similar and dissimilar motions.

4.6.1 | Overlap test

A Mann–Whitney U test[17] was used to measure how much the similarity scores of two data-sets overlapped each other. The Mann–Whitney U was used as the similarity scores within the sample groups failed the Shapiro-Wilk, D'Agostino Skew and D'Agostino Kurtosis tests for normal distribution, therefore a non-parametric test was required.

The test measures the probability of a randomly selected score from the sample group with the highest mean being less than a randomly selected score in other sample group using Equation (8). p is the probability of a random pair of motions from one data-set a with n_a motion pairs, scoring a lower similarity score than a random pair of motions from data-set b with n_b motion pairs, where U is the test statistic produced by the Mann-Whitney U test. If the distribution of two sample groups perfectly match, the probability of a score from one sample group being higher than the score of another sample group is 50%, hence p has a maximum possible value of 0.5, therefore $p \in \mathbb{R} : 0 \leq p \leq 0.5$.

$$p = \frac{U}{n_a n_b}. \tag{8}$$

The similarity metrics whose scores have the lowest probability p of overlapping, are considered to differentiate better between the two data-sets and should be considered more optimal.

4.6.2 | Mean average precision test

Mean average precision (MAP) tests are commonly used to evaluate search algorithms. Just like the overlap test, the MAP test measures how well a metric can distinguish between the motion pairs from two different data-sets, but using a contrasting approach. Rather than measuring overlap between two sets of scores, the MAP test combines the two sets of scores into single sorted list, then considers how many of the scores with the closest k values to each score are from the same data-set.

Given a query score q, precision P_k is the fraction of scores within the k nearest scores, which are from the same data-set as q. This is determined using $P_k = \frac{m \cup k}{k}$, where $m \cup k$ is the number scores from the same data-set as q within the nearest k neighboring scores.

To determine the average precision AP for a given query score q, the P_k for every value of k is considered in the range $\{1 \cdots n\}$ using Equation (9), where n is the total number of scores in the data-set that q belongs to, and rel_k is an indicator function which is equal to 1 if the kth nearest neighboring score is from the same data-set as q, or 0 otherwise.

$$AP = \frac{\sum_{k=1}^{n} P_k rel_k}{n}. \tag{9}$$

The mean average precision (MAP) is the average AP for every score within the combined list of scores. This determined using Equation (10), where Q is to the total number of scores in the combined lists.

$$MAP = \frac{\sum_{q=1}^{Q} AP_q}{Q}. \tag{10}$$

The similarity metrics whose scores achieve the highest results in the MAP tests, show that they are able to differentiate better between the two data-sets and should be considered more optimal.

5 | RESULTS

The results of the Overlap and MAP performance tests for each similarity metric can seen in Table 1. The two tests were used to measure the performance of 19 different similarity metrics when executing two different tasks, measuring either the alignment or similarity of two motions. The results of each performance test has been color coded to show the best (green) and worst (red) performing metrics. For the overlap tests the lower the result, the smaller overlap, indicating a better a performance. For the MAP test the higher the result the better performance.

The correlation method used in each correlation based metric is identified using the following symbols: ρ_p Pearson, ρ_s Spearman's, ρ_{kt} Kendall Tau. The results for performance tests on the correlation based metrics were averaged across each of the five joint parameterization techniques and three correlation methods used. This allowed the performance impact of the different approaches to joint parameterization and correlation to be evaluated independently of one another.

5.1 | Measuring motion alignment

The *Aligned* and *Non-Aligned* data-sets were used to determine the ability of each similarity metric to accurately measure the temporal alignment of two motions. The metrics which performed best on these data-sets are ideal for applications which require motions to be identified with temporally aligned features, such as finding motions to blend with existing motion.

Overall the correlation based metrics performed significantly better than distance based metrics at distinguishing between aligned and non-aligned pairs of motions. This suggests that correlation based metrics are a better choice when measuring alignment.

TABLE 1 The results of the performance tests for each similarity metric. The ability of metrics to measure alignment was evaluated by comparing the similarity scores of motion pairs from the *Aligned* and *Non-Aligned* datasets. Each metrics ability to measure similarity was evaluated by comparing the similarity scores of motion pairs from the *Similar* and *Dissimilar* datsets. A continuous linear color grade going from green (best) to red (worst), was applied to each column of results, to make it easier to identify the best and worst preforming metric in each test. For correlation based metrics the correlation method used is identified using ρ_p Pearson, ρ_s Spearman's, ρ_{kt} Kendall Tau. Rows showing mean scores are identified by μ.

	Performance as an Alignment Measure using Aligned and Non-aligned Dataset		Performance as an Similarity Measure using Similar and Dissimilar Dataset	
	Overlap	MAP	Overlap	MAP
Eular ρ_p	0.111	0.481	0.073	0.747
Eular ρ_s	0.073	0.557	0.073	0.750
Eular ρ_{kt}	0.062	0.586	0.072	0.743
Eular μ	0.082	0.541	0.073	0.747
Quaternion ρ_p	0.115	0.478	0.074	0.745
Quaternion ρ_s	0.067	0.566	0.072	0.748
Quaternion ρ_{kt}	0.057	0.599	0.071	0.740
Quaternion μ	0.080	0.548	0.072	0.744
Matrix ρ_p	0.108	0.487	0.074	0.755
Matrix ρ_s	0.070	0.561	0.072	0.759
Matrix ρ_{kt}	0.059	0.592	0.071	0.750
Matrix μ	0.079	0.547	0.072	0.755
Displacement Vector ρ_p	0.090	0.511	0.083	0.737
Displacement Vector ρ_s	0.065	0.565	0.081	0.741
Displacement Vector ρ_{kt}	0.052	0.603	0.080	0.732
Displacement Vector μ	0.069	0.560	0.081	0.737
Logarithmic Map ρ_p	0.137	0.447	0.073	0.748
Logarithmic Map ρ_s	0.078	0.540	0.072	0.754
Logarithmic Map ρ_{kt}	0.063	0.581	0.071	0.746
Logarithmic Map μ	0.093	0.523	0.072	0.749
Peason $(\rho_p)\mu$	0.112	0.481	0.075	0.746
Spearman's $(\rho_s)\mu$	0.071	0.558	0.074	0.750
Kendall Tau $(\rho_{kt})\mu$	0.059	0.592	0.073	0.742
Angular Distance	0.116	0.514	0.016	0.775
Positional Distance	0.197	0.412	0.014	0.769
Positional Velocity	0.246	0.395	0.341	0.288
Point Cloud	0.196	0.399	0.039	0.670
σ (without outliers)	0.026	0.067	0.001	0.009

Kendall Tau was the best performing correlation method in both performance tests, with both rank correlation methods consistently performing better than Pearson's linear correlation in all tests. The overlap and MAP tests showed displacement vectors to be the best method of parameterising joints. The optimal metric for measuring alignment is a correlation based metric in which joint rotations are parameterized using displacement vectors and the correlation is measured using Kendall Tau.

The averages of correlation based metrics in both tests show a smaller deviation between metrics that used different approaches to parameterising joint angles, than between those that used different methods of correlation. This suggests that the choice of correlation method is a more important factor to consider than the choice of joint angle parameterization.

The angular distance metric also performed well in both tests at distinguishing between aligned and non-aligned motion pairs. This method would be a good choice where correlation based metrics are hard to implement, such as in real-time applications.

Both the overlap and MAP tests were better able to differentiate between the performance of different correlation based metrics at measuring motion alignment than similarity. However, this is potentially indicative of the unsuitability of the correlation metrics at measuring similarity.

5.2 | Measuring motion similarity

The *Similar* and *Dissimilar* data-sets were used to determine the ability of each similarity metric to accurately measure the similarity of two motions. This identifies optimal metrics to use for tasks such as identifying or classifying motions.

In general the distance based metrics performed better than the correlation based metrics when measuring motion similarity. In particular, metrics based on angular and positional distance performed the best in both the overlap and MAP tests, with a clear performance gap between these and other metrics on the overlap test. The high performance of angular and positional distance based metrics in the MAP and overlap tests, shows that these metrics are ideal for performing discrete or Boolean decisions such as identifying a motion as the same or not the same.

Given the marginal difference in performance between angular and positional based distance metrics, the optimal choice between these two metrics will be dependent on the use case. The angular distance metrics is a better choice for comparing the overall joint poses within the motions, as it is less affected by any potential differences in joint lengths between the two motions. However, if the priority is to measure similarities in the positions of end effectors, such as hands and feet, then a metric based on joint position should still be used.

The results of the performance tests from distance based metrics corroborate the findings of Chan et al.[3] which also found that distance metrics based on joint angle and position performed best at discriminating between similar and dissimilar motions and that a distance metric based on joint velocity performed poorly. For applications where joint velocity is important, such as working with dynamic or ballistic motions, a point cloud metric should be considered, rather than a joint velocity metric.

Across both performance tests there is significantly more variation between the results of the distance based metrics than the correlation based metrics. This is to be expected as each distance based metric is based on different motion features such as velocity, rotation or position, whereas all the correlation based metrics are all based on joint rotation.

5.3 | Computational performance

The computation performance of each similarity metric implemented in this study was evaluated using a small subset of motions, to determine the average time take by each metric to evaluate the similarity of a frame of motion. The computational efficiency of a similarity metric is an important consideration in real-time applications or when searching through large data-sets of motions.

The results of this test can be seen in Tables 2 and 3, for the correlation and distance base similarity metrics respectively. Both tables show the average time to process each frame of motion in microseconds (μs). Although in line with

TABLE 2 The average time taken by each correlation based similarity metric to evaluate the similarity of a frame of motion in microseconds (μs).

	Correlation method		
Rotation Parameter	**Pearson**	**Spearman's**	**Kendall Tau**
Eular	36	50	39
Quaternion	84	93	96
Matrix	50	86	69
Displacement vector	720	740	716
Logarithmic map	1463	1559	1490

TABLE 3 The average time taken by each distance based similarity metric to evaluate the similarity of a frame of motion in microseconds (μs).

Distance metric	Time
Angular distance	114
Positional distance	1016
Postional velocity	1017
Point cloud	1769

expectations, these results should be treated as indicative and not conclusive, as the code used to implement the metrics has not been fully optimized. Each similarity metric was ran using a single threaded implementation on an Intel i7 processor running at approximately 3.85 GHz.

Correlation metrics based on Eular angles ran the fastest, this is was expected as joint rotations are stored as Eular angles, therefore no interpretation or conversion of the angle was required to support this. Correlation metrics based on displacement vectors and the distance metric based on local joint positions, utilize transform matrices to convert or interpret the joint information, resulting in them running more slowly. Correlation metrics based on logarithmic maps where also quite computationally inefficient, due to the need make sure that every quaternion is expressed on the same hemisphere, which involves finding the mean quaternion rotation for each joint. As expected the most computationally inefficient metric is based on point clouds, due larger number of data points being compared and need to determine an optimal transformation for one of the point clouds.

6 | CONCLUSIONS

This paper compared a variety of similarity metrics based on both distance and correlation. Different methods of representing angular data and measuring correlation were evaluated, including novel approaches such as the use of displacement vectors, logarithmic maps and rank correlation. The tests revealed that correlation based metrics are better for measuring the alignment, while distance based metrics are better for measuring similarity. The results also showed that the alignment and similarity of two motions should not be considered the same, with different similarity metrics performing best in each use case. Applications concerned with motion alignment, such as motion blending and motion graphing, should therefore use correlation based metrics, while applications concerned with motion similarity, such as recognizing or classifying motions should use distance based metrics.

The results showed a correlation metric based on displacement parameterization and Kendall Tau rank correlation to be the best method for measuring the alignment of two motions. They also showed that distance based metrics based on angular or positional distance should be used to measure the similarity of two motions. Depending on the application, angular distance should be used to compare the overall pose of a motion and positional distance should be used to compare the position of end effectors.

The optimal approaches to measuring alignment identified in this paper, will be particularly useful in identifying candidate motions in motion synthesis and accurately measuring the performance of time warping algorithms. However, it is important to keep in mind that no matter how accurately a similarity metric measures alignment, it will not consider important factors such how much a motion is distorted or the physical plausibility of the output motion, other metrics could be used to do this.[6,18]

Although the optimal metrics for measuring the similarity and alignment of two motions have been shown to be different, they can be used together. A robust approach to identifying a candidate motion to blend with another motion, could be to first use a distance metric based on joint position, to identify a short list of candidate motions. Then a correlation metric would be used to identify which of the short listed motions has the best alignment. This is similar to approaches implemented by other researchers.[10]

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in BCU Single Person Actions Dataset at https://github.com/matRandall/Mocap_SinglePersonActions.

ORCID

Mathew Randall https://orcid.org/0009-0000-2869-1008
Carlo Harvey https://orcid.org/0000-0002-4809-1592
Ian Williams https://orcid.org/0000-0002-0651-0963

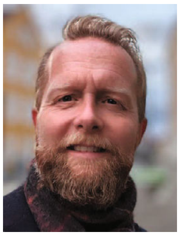
REFERENCES

1. Xia S, Wang C, Chai J, Hodgins J. Realtime style transfer for unlabeled heterogeneous human motion. ACM TOG. 2015;34(4):1–10.
2. Kovar L, Gleicher M, Pighin F. Motion graphs. ACM SIGGRAPH 2008. 2002. p. 473–82.
3. Chan JC, Leung H, Tang JK, Komura T. A virtual reality dance training system using motion capture technology. IEEE Trans Learn Technol. 2010;4(2):187–95.
4. Tormene P, Giorgino T, Quaglini S, Stefanelli M. Matching incomplete time series with dynamic time warping: an algorithm and an application to post-stroke rehabilitation. AI Med. 2009;45(1):11–34.
5. Arikan O, Forsyth DA. Interactive motion generation from examples. ACM Tog. 2002;21(3):483–90.
6. Etemad SA, Arya A. Correlation-optimized time warping for motion. Visual Comput. 2015;31(12):1569–86.
7. Lee J, Chai J, Reitsma PS, Hodgins JK, Pollard NS. Interactive control of avatars animated with human motion data. ACM CGIT 2002. p. 491–500.
8. Wang J, Bodenheimer B. An evaluation of a cost metric for selecting transitions between motion segments. ACM SCA. 2003; p. 232–8.
9. Krzeszowski T, Switonski A, Kwolek B, Josinski H, Wojciechowski K. DTW-based gait recognition from recovered 3-D joint angles and inter-ankle distance. ICCVG 2014; Springer; 2014. p. 356–63.
10. Yang H, Guan T. Motion similarity analysis and evaluation of motion capture data, Master's thesis. Edmonton, Canada: University of Alberta; 2004.
11. Puth MT, Neuhäuser M, Ruxton GD. Effective use of Spearman's and Kendall's correlation coefficients for association between two measured traits. Anim Behav. 2015;102:77–84.
12. Li Z, Xie X, Zhou X, Guo J, Bie R. A generic framework for human motion recognition based on smartphones. IIKI 2015; IEEE. p. 299–302.
13. Folgado D, Barandas M, Matias R, Martins R, Carvalho M, Gamboa H. Time alignment measurement for time series. Pattern Recognit. 2018;81:268–79.
14. Valcik J, Sedmidubsky J, Zezula P. Assessing similarity models for human-motion retrieval applications. ICCVG. 2016;27(5): 484–500.
15. Johnson MP. Exploiting quaternions to support expressive interactive character motion. PhD thesis. Cambridge, MA, USA: Massachusetts Institute of Technology; 2003.
16. Randall M. Motion capture dataset of single person actions. Available at https://github.com/matRandall/Mocap_SinglePersonActions. 2022.
17. Haslwanter T. An introduction to statistics with python. *With applications in the life sciences*. Switzerland: Springer International Publishing; 2016.
18. Reitsma PS, Pollard NS. Perceptual metrics for character animation: sensitivity to errors in ballistic motion. ACM SIGGRAPH 2003 Papers; 2003. p. 537–42.

AUTHOR BIOGRAPHIES

Mathew Randall is a Senior Lecturer at Birmingham City University in the UK. He is a member of the Graphics and Vision Research group, within DMTLabs, where he is currently completing his Ph.D. and specializes in motion capture, virtual production and the analysis and manipulation of motion capture data. He is a senior fellow of the Higher Education Academy, lecturing in visual effects and computer graphics.

Carlo Harvey is an Associate Professor in Computer Games Technology at Birmingham City University, in the School of Computing and Digital Technology. He teaches C++ for Games, Computer Graphics and Artificial Intelligence / Machine Learning. He is the Director of Future Games and Graphics overseeing the research arm for Games and Graphics within the DMTLab at BCU.



Ian Williams is Professor of visual computing and Head of the Digital Media Technology Lab (DMT Lab) at Birmingham City University. He gained his PhD from Manchester Metropolitan University in 2008 in low level feature analysis and Ai for multiple-scale biomedical image analysis. His current research interests center on the visual and interactive computing domains, with a key emphasis on creating novel methods, theories, and paradigms for interacting with, and improving on, the Quality of Experience for users of Augmented Reality (AR) and Virtual Reality (VR).

How to cite this article: Randall M, Harvey C, Williams I. Correlation as a measure for alignment and similarity of human motions. Comput Anim Virtual Worlds. 2023;34(3):e2157. https://doi.org/10.1002/cav.2157