

Investigating Similarity Measures for Locomotor Trajectories Based on the Human Perception of Differences in Motions

Annemarie Turnwald, Sebastian Eger, and Dirk Wollherr

Abstract—Providing robots with the ability to move human-like is one of the recent challenges for researchers who work on motion planning in human populated environments. Human-like motions help a human interaction partner to intuitively grasp the intention of the robot. However, the problem of validating the degree of human-likeness of a robot motion is rarely addressed, especially for the forward motion during navigation. One approach is using similarity measures to compare the robot trajectories directly with human ones. For this reason, this paper investigates different methods from the time series analysis that can be applied to measure the similarity between trajectories: the average Euclidean distance, the Dynamic Time Warping distance, and the Longest Common Subsequence. We aim to identify the measure that performs the same way as a human who rates the similarity. Thus, the evaluation of the methods is based on a questionnaire that examines the human perception of differences between walking motions. It is concluded that the human similarity perception is reproduced best by using the Dynamic Time Warping and comparing the derivatives of the path and velocity profiles instead of the absolute values.

I. INTRODUCTION

A prerequisite for seamless human-robot interaction is that a human can easily interpret the intentions of its robotic partner. The interpretation is assumed to be facilitated if a robot moves *human-like* [1]. However, evaluating if motions are actually human-like or to which extent is still an open issue. So far, it has mostly been assessed qualitatively: researchers on human crowd simulations validated if their methods reproduce characteristic behaviors like line forming or flocking [2], [3]; others checked for the adherence to social standards [4] like keeping a comfortable distance, or carried out a Turing test [5]; the most common approach is to learn cost functions from human data and to compare generated paths visibly by plotting them next to human ones [6], [7], [8]. To our knowledge, a quantitative method is still missing.

We aim to identify a *measure for the similarity of trajectories* such that artificial trajectories of a motion planner can be compared to human ones. Of our particular interest is rating the human-likeness of *forward walking motions during navigation*, meaning the traveled trajectory while walking. Two exemplary walking trajectories are shown in Fig. 1. One difficulty with these trajectories is that they vary in their path as well as their velocity profile. A suitable measure has to account for changes in both. This leads to a second challenge: determining which differences weight more.

This work was supported in part within the ERC Advanced Grant SHRINE Agreement No. 267877 (www.shrine-project.eu).

The authors are with the Chair of Automatic Control Engineering (LSR), Technische Universität München, 80333 Munich, Germany, {a.turnwald, dw}@tum.de, seger@lsr.ei.tum.de.

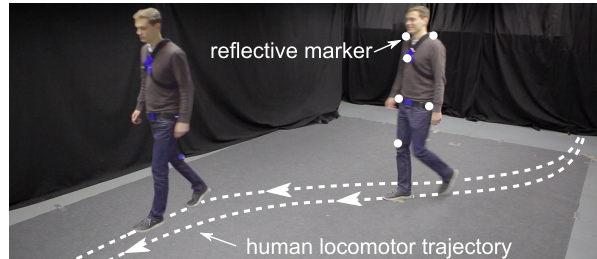


Fig. 1. Video setup of the questionnaire. Participants had to rate the difference between forward walking motions. The marker distribution for the tracking system is shown.

In order to approach the first problem we interpret the path and velocity profile of a human motion as two distinct time series. Common methods from the time series analysis are applied to compare the profiles separately. Specifically, the *average Euclidean distance*, the *Dynamic Time Warping*, and the *Longest Common Subsequence* are evaluated. Importantly, the local derivatives of the profiles are also considered such that emphasis is put on differences between the shape.

The second problem is addressed by basing the weighting on the human perception of similarity between motions. A video questionnaire is set up to assess if and to what extent humans perceive different motions during walking as being different. For example, a human may perceive different path shapes as striking while overlooking velocity changes. The results are used to adjust the weighting between the path and velocity comparisons and to evaluate the similarity measures.

The following Sec. II surveys the related works. Sec. III presents the questionnaire about the human perception of similarity, followed by descriptions of the considered measures and their adoptions to an application for human trajectories in Sec. IV. Their performance is evaluated in Sec. V.

II. RELATED WORK

The research on computer animation that creates *full-body* motions of humans is highly related. In [9] and [10] different measures to divide generated motions in two classes – natural and unnatural – are investigated. The authors let the measures compete with human assessment to evaluate the performance. Reitsma *et al.* [11] suggested to involve the human perception of motions from the outset in the development of a similarity measure. They measured the sensibility of humans to changes in horizontal and vertical velocity. The findings are used as guidelines for rating the human-likeness of animated body motions. Following Reitsma’s idea, Tang *et al.* [12] asked in a questionnaire if motions are similar and presented a measure based on its results and on machine learning.

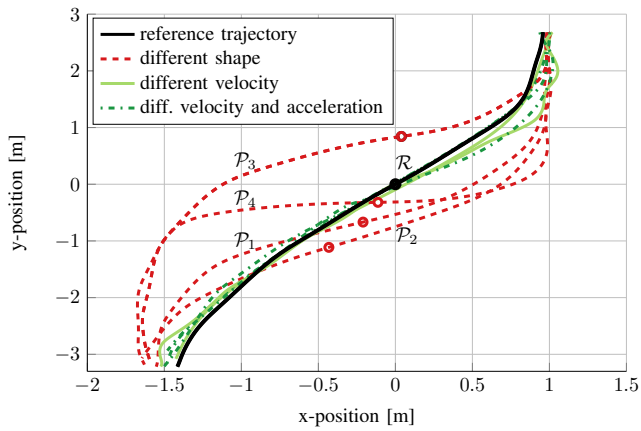


Fig. 2. Path profile of the reference trajectory and the nine investigated motions. Inflection points are drawn in as circles.

However, the method mentioned so far focused primarily on the binary decision if a motion appears natural or not. As one of few, Pražák *et al.* [13] were interested in the perceived similarity of several motions. Therefore, humans were asked to select motions that seem to be most similar. Based on the findings a metric was developed. An analogical approach was shown by Krüger *et al.* [14].

Note that all mentioned works deal with full body motions such as jumping, dancing or gesturing. These are complex and require that the motions of all joints and limbs are analyzed. The ordinary forward walking has received little attention, although this is the essential motion in order to compare the human-likeness of the navigation of robotic platforms. We differ from the related work by concentrating on forward walking motions. Moreover, we specifically ask for the level of perceived difference between motions.

III. HUMAN MOTION PERCEPTION

A questionnaire assesses which kinds of changes during a walking motion are noticed pre-eminently by humans. Its setup and results are presented in the following.

A. Data Processing and Experimental Setup

1) *Data recording and processing:* Motions were recorded with a video camera and an optical motion capture system from Qualisys. For the latter six reflective markers were put on a person as in Fig. 1 and their 3D positions over time were recorded at 204 Hz. The mean position of all markers was calculated at each time step and smoothed with a Butterworth filter (4th order, 0.01 cutoff frequency¹) to remove the torso oscillations.

In this work, a *walking motion* will be defined by a single *trajectory*. Different motions come about by changing the path shape or the velocity profile of a *reference trajectory*, denoted as \mathcal{R} . Its path and velocity profiles are shown as solid black lines in Fig. 2 and 3. It is characterized by its inflection point at the origin, the slope $m \approx 1.5$ through the inflection point, and a mean velocity of 1.1 m/s. In order to define new trajectories, either the path or velocity profile of the reference were altered. Note that the trajectories are real

¹Note that alternatively we presented a spline based method in [15].

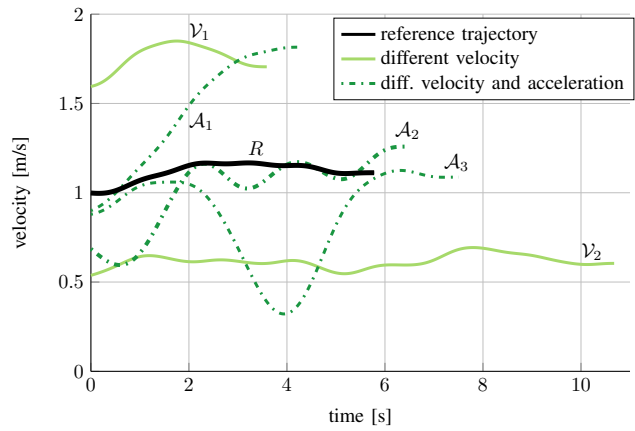


Fig. 3. Velocity profile of the reference trajectory and the five investigated motions with altered velocity profiles.

human motions. Thus, path and velocity profile never match exactly. All in all, nine trajectories were recorded:

- Different path profile, similar velocity profile:
 - \mathcal{P}_1 : small difference in the path, inflection point shifted south-westwards, slope $m \approx 0.6$.
 - \mathcal{P}_2 : medium difference in the path, inflection point shifted south-westwards, slope $m \approx 0.8$.
 - \mathcal{P}_3 : medium difference in the path, inflection point shifted northwards, slope $m \approx 0.4$.
 - \mathcal{P}_4 : large difference in the path, inflection point shifted south-westwards, point $m \approx 0.0$.
- Different velocity profile, similar path profile:
 - \mathcal{V}_1 : larger mean velocity of 1.7 m/s.
 - \mathcal{V}_2 : smaller mean velocity of 0.6 m/s.
 - \mathcal{A}_1 : high acceleration while walking.
 - \mathcal{A}_2 : one stop at inflection point.
 - \mathcal{A}_3 : changing acceleration and deceleration.

2) *Experimental Setup and Procedure:* An online questionnaire² was set up wherein the participants had to state for each motion how different it is in comparison to the reference trajectory. For this purpose, the participants were each shown two videos: the first one always showed a person walking the reference trajectory; the second one showed the same person walking one of the other recorded motions. The videos could be restarted if desired. The participants were asked to rate the perceived difference between the walking motions on a scale between 1 (small difference) and 11 (big difference). The sequence of shown motions was randomized.

B. Results

Overall, 77 participants rated all nine motions. The resulting boxplots of the ratings are shown in Fig. 4. Although the boxplots neglect that some persons tend to rate high while others prefer low values, they already indicate that trajectory \mathcal{A}_1 – the behavior where the person accelerates – is perceived as more similar to the reference trajectory than the other motions. In contrast, trajectory \mathcal{P}_4 – the large deviation from the path – is perceived as highly different. Moreover, the distribution may have unequal variances.

²Software SoSci Survey, <https://www.soscisurvey.de/>.

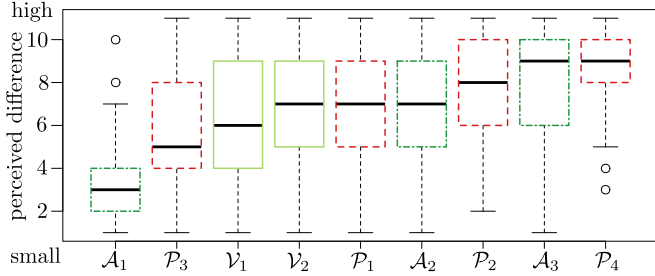


Fig. 4. Level of perceived difference between the reference trajectory and the rated motion behavior shown with boxplots.

Thus, the non-parametric Friedman-Test is chosen to check if any of the motions is rated consistently higher or lower than the others. It takes within subject data into account and is suitable if the distribution is unknown. The resulting p-value is $\ll 0.001$ with a 5% significance level, hence, at least one group differs significantly from another group.

In order to decide which motions are perceived different, a post hoc analysis was performed. Fig. 5 shows the p-values of the group comparisons and the boxplots of the differences of the ratings. Significant differences are marked grey. As assumed, the median rating value for the acceleration trajectory \mathcal{A}_1 is significantly smaller than all other trajectories, see the eight grey boxplots in Fig. 5 on the left. The post hoc test also confirmed that the median of the rates for trajectory \mathcal{P}_4 is higher than the rest. The only exception is trajectory \mathcal{A}_3 with changing acceleration and deceleration. This trajectory itself is also perceived as very different to the reference. Apart from that, the difference of trajectory \mathcal{P}_3 is perceived as smaller than the trajectories \mathcal{P}_2 and \mathcal{A}_2 .

To sum up, the participants rated the unusual motion behaviors like the huge path deviation with high curvature \mathcal{P}_4 , the stopping motion \mathcal{A}_2 , and the changing accelerations \mathcal{A}_3 as highly different to the reference trajectory. The constant acceleration motion \mathcal{A}_1 was only marginally noticed. Interestingly, the ratings for the small and medium path deviations ($\mathcal{P}_1, \mathcal{P}_2, \mathcal{P}_3$) are similar to the one of the velocity deviations ($\mathcal{V}_1, \mathcal{V}_2$). Hence, a proper similarity measure for human trajectories has to account equally strong for changes in both path and velocity. Importantly, also sudden acceleration changes need to be detected.

IV. SIMILARITY MEASURES FOR HUMAN LOCOMOTOR TRAJECTORIES

A trajectory recorded with a motion capture system can be interpreted as a time series. One of the most popular methods to compare the similarity between two time series are to compute the average Euclidean distance, the Dynamic Time Warping or the Longest Common Subsequence. Therefore, these three approaches are evaluated for their suitability to rate the difference between trajectories similar as humans would do. First, the approaches are presented generally for the application of comparing multivariate time series. After that we elaborate how to use these methods such that they account for the specifics of human trajectories.

A. Similarity Measures for Time Series

A time series is denoted as \mathbf{A} (or \mathbf{B} , respectively) and consists of a series of N -dimensional observations $\mathbf{a}[t]$: $\mathbf{A} = (\mathbf{a}[1], \dots, \mathbf{a}[t], \dots, \mathbf{a}[T_a])$, with $a_n[t]$ being the n -th dimension of observation $\mathbf{a}[t]$ at time step t . T_a is the number of observations. A time series is called *univariate* if $N = 1$, otherwise it is *multivariate*.

1) *Average Euclidean Distance*: A simple way to compare time series is to compute the *average Euclidean distance* between the observations. It is defined by:

$$D_{\text{EUC}}(\mathbf{A}, \mathbf{B}) = \frac{1}{T_a} \sum_{i=1}^{T_a} d(\mathbf{a}[i], \mathbf{b}[i]), \quad \text{with} \quad (1)$$

$$d(\mathbf{a}[i], \mathbf{b}[j]) = \|\mathbf{a}[i] - \mathbf{b}[j]\|_2. \quad (2)$$

All trajectories are sampled to have equal number of elements that are equidistant because this measure requires the time series to have the same amount of elements.

2) *Dynamic Time Warping*: The *Dynamic Time Warping* (DTW) method [16] can compare time series of different length (T_a and T_b). It uses a one-to-many comparison to find an optimal match for each element with certain restrictions. A distance matrix \mathbf{D} is computed first:

$$\mathbf{D}(\mathbf{A}, \mathbf{B}) = \begin{pmatrix} d(\mathbf{a}[1], \mathbf{b}[1]) & \dots & d(\mathbf{a}[1], \mathbf{b}[T_b]) \\ \vdots & \ddots & \vdots \\ d(\mathbf{a}[T_a], \mathbf{b}[1]) & \dots & d(\mathbf{a}[T_a], \mathbf{b}[T_b]) \end{pmatrix}, \quad (3)$$

with $\mathbf{D}_{i,j}$ being the cell in the i -th row and the j -th column. Note that the original DTW is for univariate time series only. An extended DTW is used to cope with multivariate time series (e.g., trajectories) by using the Euclidean distance over all dimensions of the time series as proposed in [17], [18].

Secondly, a warping path $W = \{w_1, \dots, w_k, \dots, w_K\}$ through the matrix \mathbf{D} is sought. It starts at $w_1 = \mathbf{D}_{1,1}$ and ends at $w_k = \mathbf{D}_{T_a, T_b}$. At the same time, it has to minimize the normalized warping sum:

$$D_{\text{DTW}}(\mathbf{A}, \mathbf{B}) = \frac{1}{K} \min \left\{ \sqrt{\sum_{k=1}^K w_k} \right\}, \quad (4)$$

with w_k being an element $\mathbf{D}_{i,j}$ and K being a normalizing factor to compensate warping paths of different length. Additionally, the warping path has to fulfill the continuity and monotonicity constraints (compare [19]).

Dynamic programming is used to find the warping path: based on the distance matrix \mathbf{D} , a cumulative matrix $\tilde{\mathbf{D}}$ is computed with the elements

$$\tilde{\mathbf{D}}_{i,j}(\mathbf{A}, \mathbf{B}) = \mathbf{D}_{i,j} + \min\{\tilde{\mathbf{D}}_{i-1,j-1}, \tilde{\mathbf{D}}_{i-1,j}, \tilde{\mathbf{D}}_{i,j-1}\}. \quad (5)$$

The last cell $\tilde{\mathbf{D}}_{T_a, T_b}$ corresponds to the minimum warping sum in Eq. (4). Its normalized value is the Dynamic Time Warping distance D_{DTW} .

One problem is that this method admits that a large number of consecutive elements in one time series are matched with a single element in the other series. This is prevented by forcing the warping path to stay within a region around the

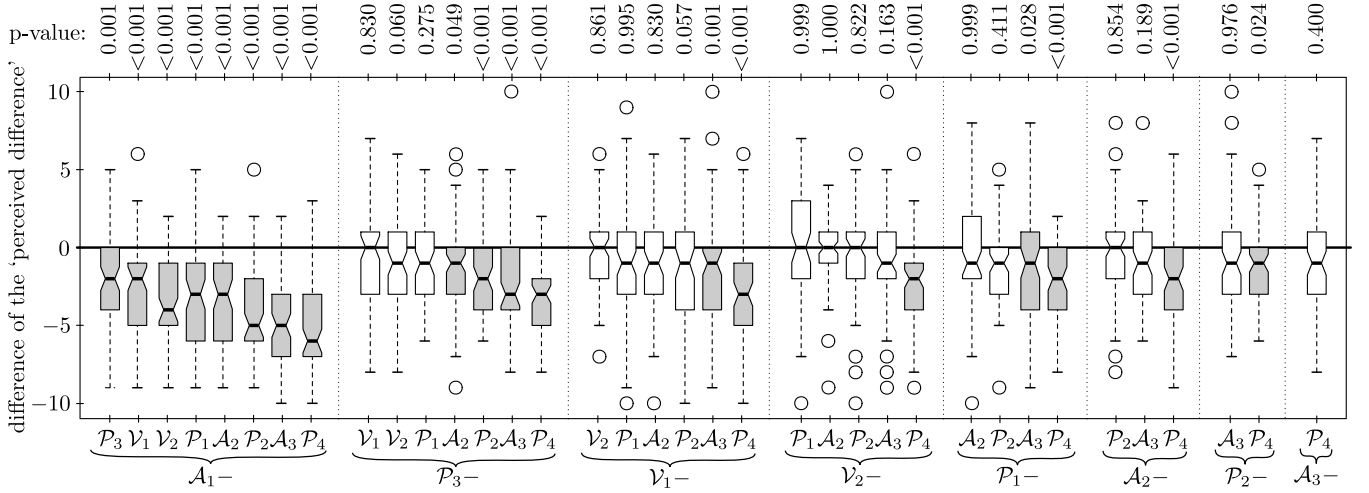


Fig. 5. Level of perceived difference between the reference trajectory and the rated motion behavior.

diagonal of matrix $\tilde{\mathbf{D}}$ [19]. Elements can only be matched if they are within δ temporal units, $\tilde{\mathbf{D}}_{i,j} = \infty$ if $|i - j| \geq \delta$.

3) *Longest Common Subsequences*: Euclidean distance and DTW punish dissimilar parts because each element of the time series needs to be matched. Contrarily to that, identifying the *Longest Common Subsequence* (LCSS) [20] focuses on the parts that are similar: it counts the number of consecutive matches, whereas two elements match if they are within δ temporal and ϵ spacial units of each other.

The LCSS constructs a matching matrix \mathbf{M} with dynamic programming similar to $\tilde{\mathbf{D}}$ in (5). The cells are defined by:

$$\mathbf{M}_{i,j}(\mathbf{A}, \mathbf{B}) = \begin{cases} 0 & \mathbf{A} \text{ or } \mathbf{B} \text{ empty,} \\ 1 + \mathbf{M}_{i-1,j-1} & \mathbf{D}_{i,j} < \epsilon \text{ and} \\ & |i - j| < \delta, \\ \max\{\mathbf{M}_{i-1,j}, \mathbf{M}_{i,j-1}\} & \text{otherwise.} \end{cases} \quad (6)$$

Since the LCSS counts the matches, we define

$$D_{\text{LCSS}}(\mathbf{A}, \mathbf{B}) = 1 - \frac{\mathbf{M}_{T_a, T_b}(\mathbf{A}, \mathbf{B})}{\min\{T_a, T_b\}} \quad (7)$$

to be the LCSS distance between two time series.

B. Adapting Similarity Measures for Human Trajectories

The presented measures are mostly used to compare signals that contain the same information, meaning they have the same unit like frequency or meter. However, for human trajectories the following adaption is necessary because the questionnaire revealed that one has to account for changes in path and velocity, which have different units.

1) Splitting Motions into Position and Velocity Profiles:

We propose to interpret the path and velocity profile of a human motion as two distinct time series. For example, the reference motion \mathcal{R} can be split into

$$\mathcal{R} = \{\mathbf{A}_{\mathcal{R}}^{\text{pos}}, \mathbf{A}_{\mathcal{R}}^{\text{vel}}\}. \quad (8)$$

$\mathbf{A}_{\mathcal{R}}^{\text{pos}}$ marks the multivariate time series that contains the positions of a person over time. It consists of the two dimensional

observations $\mathbf{a}_{\mathcal{R}}[t] = (a_1[t], a_2[t])^T$ that represent the x and y -position.

The time series $\mathbf{A}_{\mathcal{R}}^{\text{vel}}$ contains the forward velocity over time. It is obtained by calculating the covered distance over time based on the position observations. Note that $\mathbf{A}_{\mathcal{R}}^{\text{vel}}$ is univariate since only the forward velocity is considered. Additionally, the velocity is smoothed with a Butterworth filter (2th order, 0.004 cutoff frequency) to remove the step oscillations. This is not strictly necessary. However, our desired application is the performance evaluation of a robot motion planner by comparing its trajectories to human ones. Robotic platforms are mostly wheeled, hence, go without the typical accelerations during step motions.

In the following, \mathbf{A}^{pos} and \mathbf{A}^{vel} are defined to be different *types* of time series. Let types to differ in the information they contain: e.g., either path or velocity information.

2) *Considering the Derivative*: Two further types are considered: the local derivative of the position and velocity, respectively. Thus, not only the raw path and velocity profiles are compared but also their ‘shapes’. Inspired by Keogh and Pazzani [21], the derivative of the time series is taken in order to account for differences in the rising and falling trends of a curve.

Let us denote $D_t \mathbf{A}$ to be the time series that consists of the (approximate) derivative of \mathbf{A} . For simplicity the following estimate of the derivative is used [21]:

$$D_t[\mathbf{a}] = \frac{(\mathbf{a}[t] - \mathbf{a}[t-1]) + ((\mathbf{a}[t+1] - \mathbf{a}[t-1])/2)}{2}, \quad (9)$$

with $1 < t < T_a$.

The derivative of the time series containing positions is denoted as $D_t \mathbf{A}^{\text{pos}}$. Note that it is different from the forward velocity because it is taken for each dimension. $D_t \mathbf{A}^{\text{vel}}$ corresponds to the derivative of the velocity.

In the following, it is differed between four types of time series. They are named as ‘Pos’, ‘dPos’, ‘Vel’ and ‘dVel’.

3) *Combining Position and Velocity Comparison*: The similarity of two motions can be calculated by comparing one of these four types each. Which type is suited best for human

TABLE I

PERFORMANCE RATING PR OF THE THREE CONSIDERED SIMILARITY MEASURES DEPENDING ON THE TYPE OF THE COMPARED TIME SERIES.

$PR \in [0, 1]$, WITH 1 BEING THE BEST POSSIBLE PERFORMANCE.

Type of time series	D_{EUCL} PR	D_{DTW} PR	D_{LCSS} PR
Pos	0.68	0.63	0.37
dPos	0.74	0.68	0.63
Vel	0.53	0.79	0.47
dVel	0.53	0.63	0.58

motion will be examined in the next section. Additionally, it is investigated if it is advisable to sum up the results of the single comparisons. Four combinations are considered:

- Pos+Vel: $D(\mathbf{A}^{\text{pos}}, \mathbf{B}^{\text{pos}}) + \lambda D(\mathbf{A}^{\text{vel}}, \mathbf{B}^{\text{vel}})$
- Pos+dVel: $D(\mathbf{A}^{\text{pos}}, \mathbf{B}^{\text{pos}}) + \lambda D(D_t \mathbf{A}^{\text{vel}}, D_t \mathbf{B}^{\text{vel}})$
- dPos+Vel: $D(D_t \mathbf{A}^{\text{pos}}, D_t \mathbf{B}^{\text{pos}}) + \lambda D(\mathbf{A}^{\text{vel}}, \mathbf{B}^{\text{vel}})$
- dPos+dVel: $D(D_t \mathbf{A}^{\text{pos}}, D_t \mathbf{B}^{\text{pos}}) + \lambda D(D_t \mathbf{A}^{\text{vel}}, D_t \mathbf{B}^{\text{vel}})$

with λ being a constant for weighting the influence of path and velocity differences.

In order to have an equal weighting for $\lambda = 1$, the dimensions of the time series are normalized before the summation. Particularly, each dimension of the time series \mathbf{A} and \mathbf{B} together is normalized separately to zero mean and unit variance [17].

The suggested adaptations are grounded on the assumption that humans account for changes in path, velocity and shape. The next section evaluates for each measure which types have to be compared and which λ has to be used in order to mimic human motion perception as closely as possible.

V. EVALUATION OF THE SIMILARITY MEASURES

This section evaluates how suitable the presented similarity measures are for comparing human trajectories.

A. Evaluation Approach

We aim to identify the measure that reproduces the human assessment of the difference between trajectories best. Therefore, the distances between the nine motions introduced in Sec. III and the reference is calculated with the three similarity measures presented in Sec. IV.

Each of the nine motions was recorded ten times since comparing the reference to only one specific representative of a motion lacks generalizability. For each trajectory the distance is calculated and the mean of all ten distances is considered for the further evaluation.

1) *Rating the Performance:* The performance rating of each similarity measure is based on a ‘bonus point system’. The questionnaire identified the criteria a measure has to fulfill in order to reproduce the human similarity assessment. According to the post hoc analysis the similarity ratings were significantly different in 19 cases. The grey boxplots in Fig. 5 show which comparisons are significant. They also reveal which motion was perceived as less different to the reference. Thus, a similarity measure has to fulfill at least these 19

TABLE II

PERFORMANCE RATING PR OF THE THREE CONSIDERED SIMILARITY MEASURES DEPENDING ON THE TYPE OF THE COMPARED TIME SERIES AND WEIGHTING λ .

Type of time series	D_{EUCL}		D_{DTW}		D_{LCSS}	
	PR	$\lambda \in$	PR	$\lambda \in$	PR	$\lambda \in$
Pos+Vel	0.68	[0.0]	0.79	[0.3, 3.0]	0.47	[1.1, 2.2]
Pos+dVel	0.68	[0.0, 3.0]	0.63	[0.0]	0.53	[1.3, 3.0]
dPos+Vel	0.74	[0.0, 0.3]	0.74	[0.1, 3.0]	0.63	[0.0, 0.3]
dPos+dVel	0.84	[1.9, 2.0]	0.84	[0.4, 1.4]	0.74	[0.6, 1.4]

criteria in order to match the human perception. For example, the distance $D(\mathcal{R}, \mathcal{A}_1)$ has to be smaller than $D(\mathcal{R}, \mathcal{P}_1)$ because the boxplot is below zero (compare leftmost boxplot in Fig. 5). $D(\mathcal{R}, \mathcal{A}_1)$ also needs to be smaller than $D(\mathcal{R}, \mathcal{V}_1)$, $D(\mathcal{R}, \mathcal{V}_2)$, and so on.

Whenever a similarity measure got the ratio right (smaller of bigger, respectively) it was credited a bonus point. The sum of collected bonus points was divided by 19, leading to a performance rating $PR \in [0, 1]$, with 1 being the best possible performance.

2) *Determining the Values for δ , ϵ and λ :* Using LCSS required to fix the values of δ and ϵ . Their values clearly depend on the type of data and the application. However, Vlachos et al. [22] state that choosing δ to be more than 20 – 30% of the trajectories length did not yield substantive improvements for most examined datasets. Thus, δ is set to be 20% of the reference trajectory length. This same δ is used to fix the size of the warping window of the DTW.

The value of ϵ is application specific. Best results were achieved by setting ϵ to a quarter of the average Euclidean distance in Eq. (1).

Several values for the weighting factor λ were examined by incrementally increasing λ from 0 to 3 with a step size of 0.1. The values yielding the best results are discussed in the next section.

B. Results

First, the performance ratings PR are presented for the case that the path and velocity profiles of the motions as well as their derivatives are compared separately. Tab. I summarizes the results. The best performance rating of $PR = 0.79$ was achieved by applying the DTW on the velocity profiles. The Euclidean distance is with its best PR of 0.74 only slightly lower. In contrast to the DTW, this result was obtained by comparing the derivative of the position, hence, the shape of the trajectories. The same applies for the LCSS that was the worst performing measure with a best PR of 0.63.

The performance rate could be further enhanced in case that the separately calculated distances in Tab. I were combined as shown in Sec. IV-B.3. Tab. II lists the performance. Only the best PR values are shown depending on the corresponding interval of the weighting parameter λ . Clearly, comparing and combining the derivative of both the path as well as the velocity profile (dPos+dVel) was most successful:

all measures performed best with these types of time series inputs. The Euclidean distance and DTW are on par with $PR = 0.84$. The results of the DTW seem to be more robust since its range for λ is obviously larger than for the Euclidean distance. Again the performance of LCSS is slightly worse with $PR = 0.74$. The reason may be that the LCSS with its binary decision for each element – match or no match instead of a distance value – is unable to specify the amount of difference over a certain threshold.

Although Euclidean distance and DTW perform well, non of the measures satisfied all criteria. We examined which criteria were violated in which cases. If the derivative of the velocity was disregarded, the two motions with changing accelerations \mathcal{A}_2 (one stop) and \mathcal{A}_3 (slow and fast) were assigned very low difference values when compared to the reference. This is in great contrast to the human rating which assigns these motions high differences, probably because they appear as rather unusual. Only the DTW detected the differences without relying on the derivative of the velocity, but still performs best by using the combination of dPos and dVel. However, using dVel is a trade-off because it assigns the remaining (constant) acceleration trajectory \mathcal{A}_1 additional distance although human rarely noticed the constant acceleration. Apart from that all similarity measures rated the difference of \mathcal{P}_3 as too high compared to the human perception. The ratios got slightly better in case the derivative of the position was used.

As final remark we note that the good results with using the derivatives suggest by implication that humans rather pay attention to differences in the shape than in the positions.

VI. CONCLUSIONS

This paper evaluated three similarity measures for times series with regard to their eligibility for the comparison of human locomotor trajectories during walking. The evaluation was based on how far the measures agreed with the human perception of similarity between motions. On that account, a video based questionnaire revealed that humans perceive differences in position and velocity similarly strong. However, a huge deviation from the path or unnatural acceleration during the motion were rated as significantly different. These results could be reproduced best by applying Dynamic Time Warping on the derivative of the position and velocity profiles of the motions that are to be compared. The performance of the Euclidean distance was almost as good. However, none of the similarity measures reflect the human assessment exactly.

Further studies are needed to gain more insights about the human similarity assessment of walking motions since our questionnaire yields merely a rough valuation. Based on this, it would be advisable to learn a suitable weighting between position and velocity differences. Future work will also take further, existing similarity measures into account. Particularly, Dynamic Motion Primitives [23] hold promises since they yield a generic framework to represent motions. Motion segments are compared by regarding the weighting of the basis functions.

REFERENCES

- [1] T. Kruse, A. Pandey, R. Alami, and A. Kirsch, “Human-aware robot navigation: A survey,” *Robotics and Autonomous Systems*, vol. 61, no. 12, pp. 1726–1743, 2013.
- [2] S. Hoogendoorn and P. Bovy, “Simulation of pedestrian flows by optimal control and differential games,” *Optimal Control Applications and Methods*, vol. 24, no. 3, pp. 153–172, 2003.
- [3] N. Pelechano, J. Allbeck, and N. Badler, “Controlling individual agents in high-density crowd simulation,” in *ACM SIGGRAPH/Eurographics symposium on Computer animation*, 2007, pp. 99–108.
- [4] B. Kim and J. Pineau, “Human-like navigation: Socially adaptive path planning in dynamic environments,” in *RSS Workshop on Inverse Optimal Control & Robot Learning from Demonstrations*, 2013.
- [5] H. Kretzschmar, M. Kuderer, and W. Burgard, “Learning to predict trajectories of cooperatively navigating agents,” in *IEEE Int. Conf. on Robotics and Automation*, 2014, pp. 4015–4020.
- [6] M. Luber, L. Spinello, J. Silva, and K. O. Arras, “Socially-aware robot navigation: A learning approach,” in *IEEE Int. Conf. on Intelligent Robots and Systems*, 2012, pp. 902–907.
- [7] M. Kuderer, H. Kretzschmar, C. Sprunk, and W. Burgard, “Feature-based prediction of trajectories for socially compliant navigation,” in *Proc. of Robotics: Science and Systems*, 2012.
- [8] D. Vasquez, B. Okal, and K. O. Arras, “Inverse reinforcement learning algorithms and features for robot navigation in crowds: An experimental comparison,” in *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, 2014, pp. 1341–1346.
- [9] L. Ren, A. Patrick, A. A. Efros, J. K. Hodgins, and J. M. Rehg, “A data-driven approach to quantifying natural human motion,” *ACM Transactions on Graphics*, vol. 24, no. 3, pp. 1090–1097, 2005.
- [10] B. van Basten and A. Egges, “Evaluating distance metrics for animation blending,” in *ACM Int. Conf. on Foundations of Digital Games*, 2009, pp. 199–206.
- [11] P. Reitsma and N. S. Pollard, “Perceptual metrics for character animation: Sensitivity to errors in ballistic motion,” *ACM Transactions on Graphics*, vol. 22, no. 3, pp. 537–542, 2003.
- [12] J. Tang, H. Leung, T. Komura, and H. P. Shum, “Emulating human perception of motion similarity,” *Computer Animation and Virtual Worlds*, vol. 19, no. 3-4, pp. 211–221, 2008.
- [13] M. Pražák, R. McDonnell, L. Kavan, and C. O’Sullivan, “A perception based metric for comparing human locomotion,” in *Irish Workshop on Computer Graphics*, 2009, pp. 75–80.
- [14] B. Krüger, J. Baumann, M. Abdallah, and A. Weber, “A study on perceptual similarity of human motions,” in *Workshop on Virtual Reality Interaction and Physical Simulation*, 2011, pp. 65–72.
- [15] D. Carton, A. Turnwald, W. Olszowy, M. Buss, and D. Wollherr, “Using penalized spline regression to calculate mean trajectories including confidence intervals of human motion data,” in *IEEE Workshop on Advanced Robotics and its Social Impacts*, 2014, pp. 76–81.
- [16] D. Berndt and J. Clifford, “Using dynamic time warping to find patterns in time series,” in *KDD workshop*, 1994, pp. 359–370.
- [17] G. ten Holt, M. Reinders, and E. Hendriks, “Multi-dimensional dynamic time warping for gesture recognition,” in *Advanced School for Computing and Imaging*, 2007.
- [18] N. Gillian, R. B. Knapp, and S. OModhrain, “Recognition of multivariate temporal musical gestures using n-dimensional dynamic time warping,” in *Int. Conf. on New Interfaces for Musical Expression*, 2011, pp. 337–342.
- [19] C. Cassisi, P. Montalto, M. Aliotta, A. Cannata, and A. Pulvirenti, “Similarity measures and dimensionality reduction techniques for time series data mining,” *Advances in Data Mining Knowledge Discovery and Applications*, 2012.
- [20] G. Das, D. Gunopulos, and H. Mannila, “Finding similar time series,” in *Principles of Data Mining and Knowledge Discovery*. Springer, 1997, pp. 88–100.
- [21] E. Keogh and M. Pazzani, “Derivative dynamic time warping,” in *SIAM Proc. of Int. Conf. on Data Mining*, 2001, pp. 5–7.
- [22] M. Vlachos, G. Kollios, and D. Gunopulos, “Discovering similar multidimensional trajectories,” in *IEEE Int. Conf. on Data Engineering*, 2002, pp. 673–684.
- [23] S. Schaal, “Dynamic movement primitives -a framework for motor control in humans and humanoid robotics,” in *Adaptive Motion of Animals and Machines*. Springer Tokyo, 2006, pp. 261–280.