

Motion Similarity Analysis and Evaluation of Motion Capture Data

Tong Guan and Yee-Hong Yang

Computer Graphics Research Group

Department of Computing Science, University of Alberta

Edmonton, AB T6G 2E8 Canada

guantong02@hotmail.com, yang@cs.ualberta.ca

Abstract

Motion similarity analysis is a critical stage for the successful reuse of motion capture data. Some previous works use one or multiple motion features, such as the difference between joint positions, joint angles, joint velocities and accelerations to capture the similarity information between two frames of different or the same motion streams. In this paper, two typical motion similarity approaches are reviewed and related problems are identified. To address the problems, a novel motion similarity method is proposed and two different features are used to measure the similarity between two motion frames: the curvature of space curve and the difference between joint relative positions. Additionally, we also propose a new general criterion to evaluate and to compare the performance of different motion similarity techniques. The experimental results demonstrate that our new motion similarity method can automatically generate visually acceptable results efficiently and that the other methods can be

improved by using our proposed evaluation criterion to select the required attribute weights appropriately.

Index Terms --- Character animation, motion capture, motion transition, motion similarity measure

1 Introduction

In computer graphics, character animation is an important research area and many research works focus on creating realistic and natural human animations. Recently, the price of motion capture hardware has dropped significantly, and hence, techniques based on motion capture data have become commonly used alternatives to the other two animation techniques, namely, keyframing and physical-based simulation. Because of the ease of use, realistic and natural motion capture data have become widely used in commercial applications; in particular, with the development of practical motion capture and editing techniques [3][4][5][10][11][12][14].

To generate new motion patterns by reusing existing motion capture data is still an interesting open problem. Two major approaches exist. One is to use motion blending by combining two or more motion examples to form a new motion clip [11][12][14]. The other approach is motivated by the video texture technique [2][7][8][9], by which a new motion sequence is generated by stitching the original motion clips in a new order. Because it directly reuses the source motion sequences, it is able to preserve realism and high-level details of the original motion. Moreover, motion blending focuses on creating an individual motion clip, while motion reassembling focuses on generating a new motion sequence. Both of the approaches mentioned above have the same goal to create new motions from existing motion capture data. In this paper, the focus is on techniques

to stitch the original clips in a new order, in particular, in locating the best transition points in connecting two clips.

To use the raw motion capture data directly is difficult because of the unstructured and high complexity nature inherent in the data. Thus, motion analysis tools, especially for analyzing human body motion, are indispensable and have become a very important research topic in the context of motion editing. Motion similarity analysis provides the foundation for many recent research works. For example, in [2][7][8][9], similar frames are detected first to find candidate transition points. Then a new motion sequence is synthesized by reordering motion clips. As discontinuities are introduced at transition points, they must be carefully selected. Thus, a smooth transition between clips is generated. While in motion blending [3][11][12][14], time warping is performed to align the example motions in the time domain (synchronization) according to the motion similarity.

Some motion editing techniques require the user to determine the similarity manually [11] [12]. The animator specifies similar motion frames according to his/her perception and experience. Therefore, the quality of the resulting motion depends significantly on the animator's skills. Moreover, it is a labor-intensive process. In order to reduce the burden on animators and to increase the speed of motion similarity analysis, some methods [2][7][8][9] have been proposed to automatically detect motion similarity.

Two main types of motion similarity metrics based on different motion features are defined in previous works. The first one is based on joint orientations and velocities [2][9] and the other on the distance between sample points (point clouds) [8]. Both of them emphasize on the pose similarity of two frames. In order to capture kinematic information, such as velocity and acceleration, the first approach directly incorporates these features, while the second approach considers the difference between the neighborhoods of the two frames. For the first approach, as the motion capture data are directly represented by the joint angle of a skeleton, the similarity metrics based on joint orientations can be easily computed directly. The disadvantage of this technique is that it uses a weighted sum of multiple joint attributes as a measure. The optimal attribute weights are very difficult to identify and may be dependent on motion patterns. As a result, the dynamic information of the source data may not be incorporated well. The second approach is very time consuming and dependent on the coordinate system. The assumption that a motion is not changed by a rigid 2D transformation restricts its application to the movements on the same ground plane.

In order to tackle the problems mentioned above, the present work introduces two novel similarity features to reuse realistic motion capture data. Curvature, one of the intrinsic properties of space curve, is used to capture the kinematic information. The difference between joint positions in their own parent coordinates is used to evaluate the similarity of body configurations. Both of these two features are coordinate invariant and can be computed efficiently. Most importantly, no attribute weight is needed in our approach.

To evaluate the performance of different motion similarity methods, this paper also presents a general criterion that helps to determine optimal attribute weights used in other approaches. The experimental results demonstrate that our approach for motion similarity analysis can generate visually acceptable results and that the other methods' performance can be improved by using our new evaluation criterion.

This paper is organized as follows. In section 2, three motion similarity metrics are investigated. Then we describe the process of our motion similarity analysis in section 3. Finally we show experimental results to compare our approach with other approaches.

2 Related Work

Since motion similarity analysis is a critical stage for many motion editing techniques, in this section, two typical types of distance functions in recent research work are reviewed, and then the problems related to them are identified.

2.1 Similarity Based on Joint Attributes

Because motion capture data are usually directly represented by the joint angles, many research works prefer using similarity distance functions that are based on joint angles and joint velocities. Two typical examples are given below.

J. Lee et al. [9] present a similarity metric based on joint angles and joint velocities. The similarity between two frames is computed as follows.

$$D_{ij} = w_r \left\| \mathbf{p}_{i,0} - \mathbf{p}_{j,0} \right\|^2 + w_a \sum_{k=1}^m w_k \left\| \log(\mathbf{q}_{j,k}^{-1} \mathbf{q}_{i,k}) \right\| + w_v \sum_{k=1}^m w_k d(\mathbf{v}_{i,k}, \mathbf{v}_{j,k}) \quad (2.1)$$

where $\mathbf{p}_{i,0}$ and $\mathbf{p}_{j,0}$ are the root joint (pelvis) positions at frame i and frame j , respectively. $\mathbf{q}_{j,k}$ is the quaternion that represents the orientation of the k th joint at frame j . $\left\| \log(\mathbf{q}_{j,k}^{-1} \mathbf{q}_{i,k}) \right\|$ is the angle by which the k th joint rotates from the orientation of frame i to the orientation of frame j . $\mathbf{v}_{i,k}$ is the velocity of the k th joint at frame i . $\mathbf{q}_{j,k}, \mathbf{v}_{j,k}$ are, respectively, the joint orientation and velocity at frame j . k is the joint index, m is the number of joints; w_k is the joint weight to control the importance of the k th joint. w_r, w_a and w_v are the attribute weights to accommodate for the relative importance of the different joint attributes.

The similarity function between frame i and j consists of three terms. The first one is the difference between root velocities. In the original paper [9], when the relative coordinate system used, it is the difference between the global root translation with respect to the previous frame; in which case, it is the same as the difference in the root velocities. While in the fixed coordinate system, the root positions are used instead of velocities, this method will become coordinates dependent. Then only the frames that are located nearby in the 3 dimensional space can be selected as the similar frame pairs. Hence, the similarity analysis results largely depend on the root positions. The other two terms are

the difference of joint angles and the difference of joint velocities. The joint angle term captures the pose information. While the other two terms incorporate the kinematic information of the joints. From Equation (2.1), we can see that there are 3 weight parameters each for the 3 terms. In the original work, no information is given on how to determine the values of these parameters. We have tried using different attribute weights and found that the results can be quite different depending on their values (see Figure 1.)

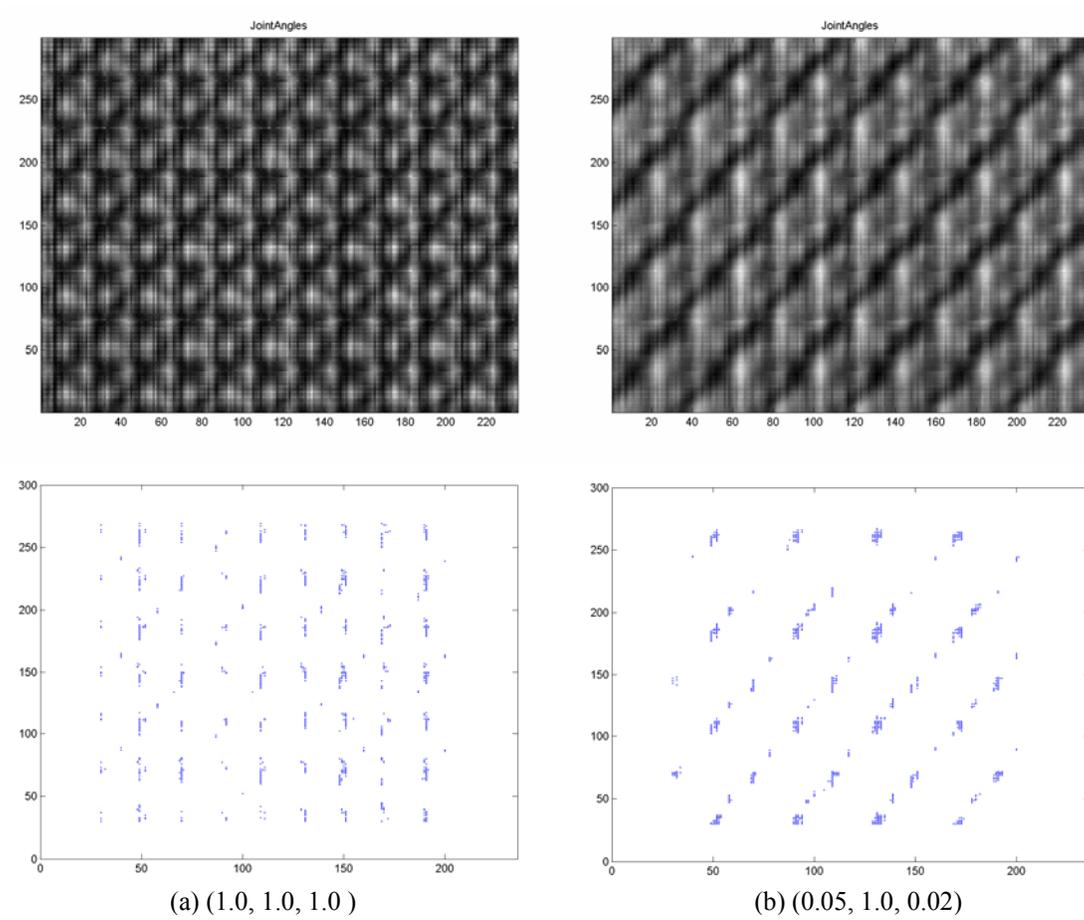


Figure 1: Similarity analysis based on joint angles with different attribute weight sets between left and right (columns correspond to frames from the first clip; rows correspond to frames from the second clip; the sampling rate is 24 frames/second). Top: The distance matrix. Bottom: The resulting similar frame pairs are denoted by dots.

Arikan and Forsyth [2] define a similar distance function in Equation (2.2), except that it incorporates joint accelerations as well. This measure is based on joint positions and velocities in the root joint (pelvis) coordinates. There are four components in the equation, each of which represents a contribution to the similarity measure. $\sum_{k=1}^m w_k dp_k$ and $\sum_{k=1}^m w_k dv_k$ are, respectively, the difference in joint positions and joint velocities with respect to the root coordinates; dv_0 and da_0 are the difference of root velocities and accelerations. Each of the last two terms includes translation and orientation. k is the joint index, m the number of joints. w_k is the joint weight to control the importance of the k th joint; w_p , w_v , w_{v_0} and w_{a_0} are the attribute weights to control the relative importance of the different joint attributes. The difference of joint positions in the root frames defines the pose similarity. The other components incorporate kinematic information like velocities and accelerations.

$$D_{ij} = w_p \sum_{k=1}^m w_k dp_k + w_v \sum_{k=1}^m w_k dv_k + w_{v_0} dv_0 + w_{a_0} da_0 \quad (2.2)$$

The major drawback of this metric is on how to determine the optimal attribute weights. In the original work, the authors provide a clever way to handle the attribute weights. The maximum difference between sequential frames in the database is used to normalize each term. The normalization process could better balance the effects of the four features on similarity analysis. But the attribute weights after normalization might be dependent on different motion databases. In other words, the attribute weights are likely to change when the database includes motion streams with different special patterns. Figure 2

shows the different similarity analysis results generated for the same two clips in two different datasets.

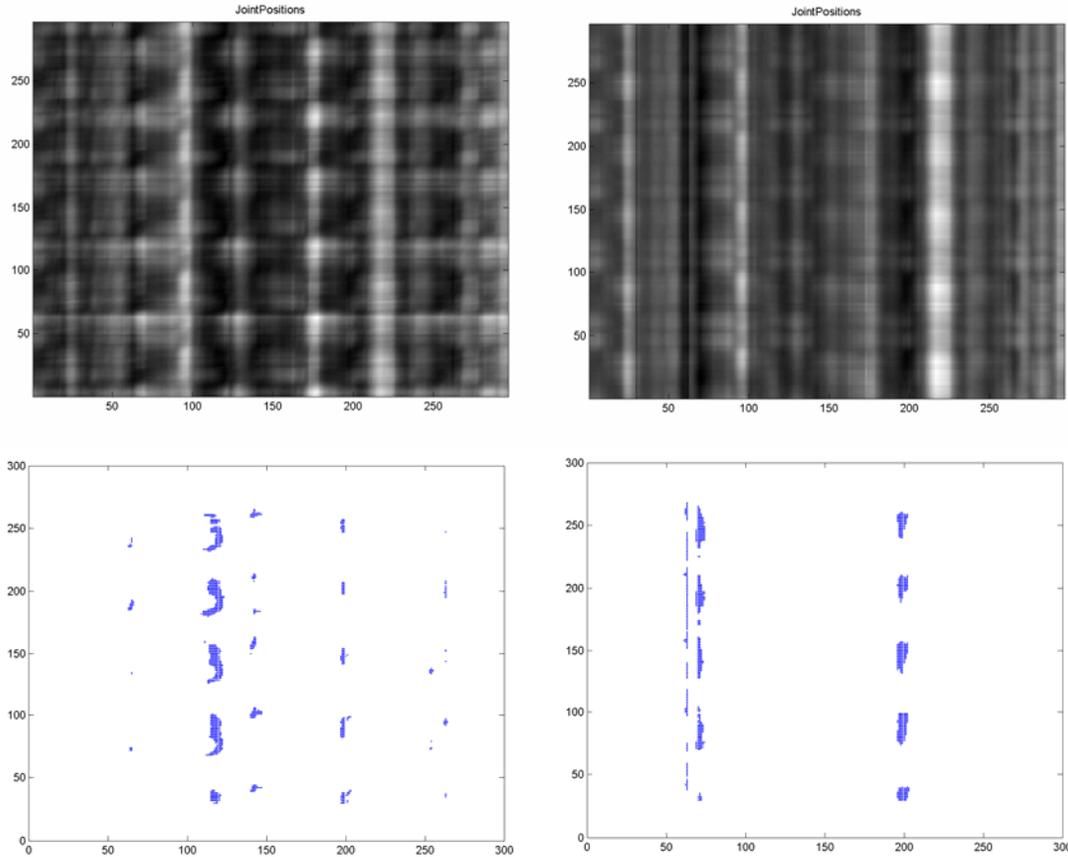


Figure 2: Similarity analysis based on joint velocities and accelerations in different datasets. Left: The dataset has only 2 clips. Right: The dataset has 25 clips. Top: The distance matrix. Bottom: The resulting similar frame pairs are denoted by dots.

2.2 Similarity Based on Point Clouds

In [8], two point sets driven by a skeleton are used to estimate the similarity between two frames. In fact, the point set of a frame can be formed by the joint positions at the

frame and its neighboring frames, which form a window of the current frame. Each frame in the window has a different weight, called frame weight, which controls the relative importance of the frame with respect to the other frames in the window. Typically, the frame weight is tapered off towards the end of the window.

The similarity metric is defined as the minimum difference between the two point sets. In order to calculate the minimum difference in joint positions, an optimal 2D transformation matrix must be computed first. After transformation, the second point set can be aligned with the first point set. The similarity measure is defined in Equation (2.3).

$$D_{i,j} = \min_{\theta, x_0, z_0} \sum_{f=1}^F w_f \sum_{k=1}^m w_k \left\| \mathbf{p}_{f,k}^i - \mathbf{T}_{\theta, x_0, z_0} \mathbf{p}_{f,k}^j \right\|^2 \quad (2.3)$$

where $\mathbf{p}_{f,k}^i$ is the k th joint position in the f th frame of the point set at frame i . $\mathbf{p}_{f,k}^j$ is similarly defined to the point set at frame j . k is the joint index; m is the number of joints; w_k is the joint weight. w_f is the frame weight; f denotes the frame index within the window and F is the number of frames in the window. \mathbf{T} is the transformation matrix for rotating by an angle θ around the vertical axis, translating on the horizontal plane by (x_0, z_0) . In the original paper, a solution is given to find the matrix \mathbf{T} . In this approach, as joint positions are directly used, the similarity between body configurations of the frames can be captured. As well the consideration of a frame's neighborhood (8 frames in length) incorporates the kinematic differences between the two sequences.

The major disadvantage of this metric is that it uses a 2D transformation matrix as the optimal matrix. The method assumes that all the motions happen at the same level as the ground. The optimal transformation matrix in [8] does not consider translation along the y-axis, and rotations around the x-axis or the z-axis. For example, when in one motion clip a character is running on a flat floor; and in the second clip, the character is walking on a ramp (see Figure 3). The similarity result is shown in Figure 4. Obviously, when the vertical distance between the root joints of the two clips increases, their similarity measure also increases. As a result, only the frames with root joints at the same level are selected as similar frame pairs. To solve this problem, one way is to compute the 3D transformation matrix by which the two point sets can be aligned. However, computing the 3D transformation matrix can be very time consuming.

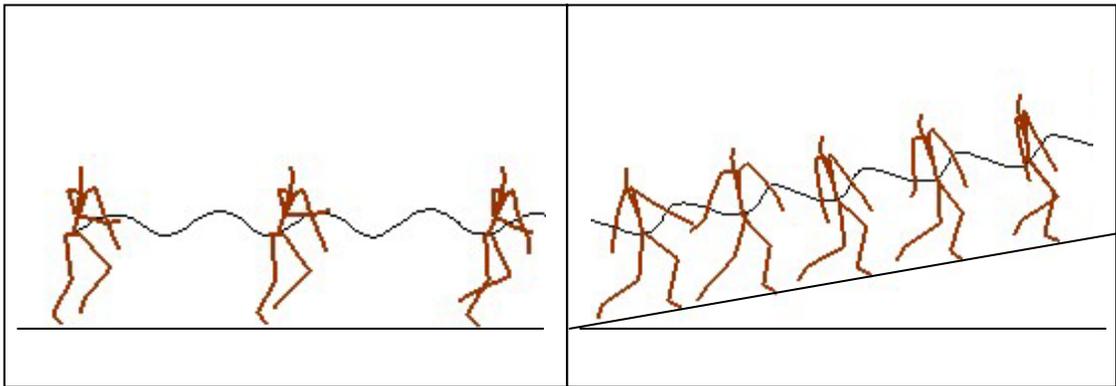


Figure 3: Two motion clips on the different level.

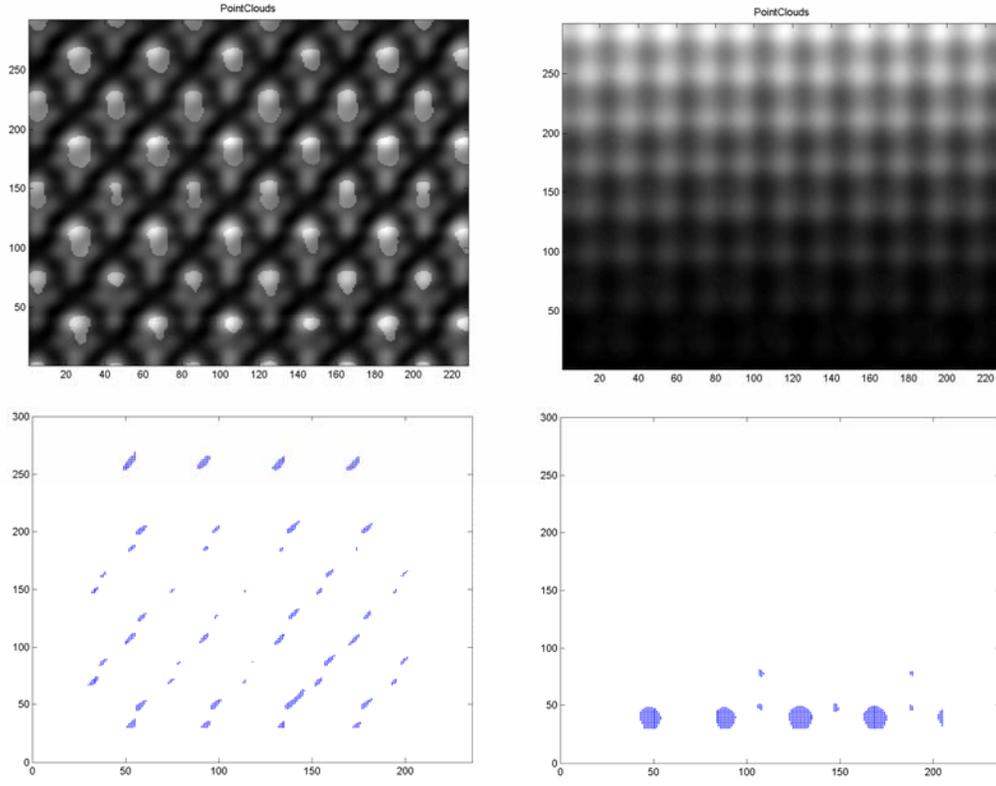


Figure 4: Similarity analysis based on point clouds. Left: two clips on the same level. Right: two clips on the different level. Middle: The distance matrix. Bottom: The resulting similar frames denoted by dots.

2.3 Joint Weight

In the motion similarity metrics discussed above, different joint weights are used to control the importance of different joints. In real life, different joints surely have different visual importance to the perception or understanding of motion. For example, the hip joint, knee joint, shoulder joint, elbow, pelvis and spine are more important than the others. In [9], the weights of the important joints are set to one, while the unimportant ones are set to zero. Wang et al. [13] propose a set of optimal joint weights

(shown in **Table 1**) for the similarity metrics proposed in [9]. They also compare their optimal joint weight set with the one in [9] by running a user study and found that the results using the optimal joint weight set are more robust and superior to the results using the original joint weight set in [9]. However, they indicate that their experimental results may be affected by the motion database and the different transition techniques applied. As all motion similarity analysis techniques have to specify a joint weight set to control the importance of different joints, in the novel approach presented in Section 3, we adopt the joint weight set as specified in [9].

Table 1: The optimal joint weight set in [13]¹.

Joint Name	Optimal Joint Weight
Right and Left Hip	1.000
Right and Left Knee	0.0901
Right and Left Shoulder	0.7884
Right and left Elbow	0.0247

3 Curvature and Relative Position Approach

In this paper, we are concerned with finding visually similar frames in two motion streams automatically. This means that the skeletal poses, joint velocities, and accelerations of these two frames and of their neighborhood frames should be similar. In

¹ Only the joints with non-zero weights are shown.

order to efficiently and accurately identify corresponding similar frames in two motion clips, we select features that satisfy the following criteria.

- Coordinate invariant. This is important because the motion should not change after 3D rigid transformations. It is also called translation and rotation invariants. Here, we only consider the motion data with the same skeletal size.
- Efficient computation. Since similarity analysis is performed on a frame by frame basis, the computation time is at least $O(n^2)$, where n is the number of frames in the motion database. For this reason, the feature computation cost is expected to be lower. Thus, the similarity computation can scale well with large motion databases.
- The similarity of skeletal pose and kinematics information should be incorporated.
- The method should not require determining the attribute weight parameters.

According to the above criteria, we select two features in our similarity analysis method. The first is joint relative position in its own parent coordinate. The second is the curvature of the global curve formed by the movement of each joint. Both of them are coordinate independent and can be easily computed from the motion capture data. In the proposed approach, there are two stages as shown in Figure 5, which are described in details in the following sections.

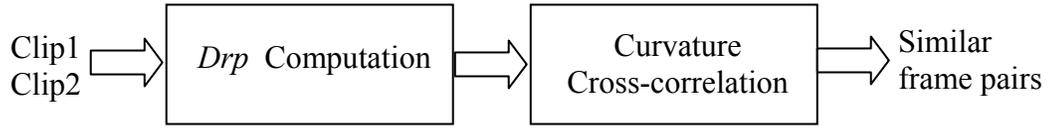


Figure 5: Motion similarity analysis process.

3.1 Difference between Joint Relative Positions

The joint (except for the root joint) relative position and its difference in two motion frames are computed according to Equation (3.1) and (3.2).

$$\mathbf{rp} = R_z R_x R_y \mathbf{v} \quad (3.1)$$

$$D_{rp} = \sum_{k=1}^m w_k \left\| \mathbf{rp}_{i,k} - \mathbf{rp}_{j,k} \right\|^2 \quad (3.2)$$

\mathbf{v} is a vector that represents the offset of a joint. $R_z R_x R_y$ are the rotation matrix of its parent; i and j are the frame indices; k is the joint index, m is the number of joints; w_k is the joint weight to control the importance of the k th joint. Note that in this feature, the root joint (pelvis) is excluded. Since the root joint's parent is the global world, its relative position is the same as the root's global position. If the root joint were included in the difference sum, then a coordinate dependent factor would be introduced. For the same reason, the root's children, i.e. Chest, LeftUpLeg and RightUpLeg, are also excluded. Thus, the difference between joint relative positions is also invariant to global orientations. The joint relative position reduces the effects of its parent's movements on

the joint. The difference of this feature represents the position difference after the origins of their parent coordinates have been aligned and hence, can capture the pose similarity more accurately. After this step only candidate frame pairs that are similar enough will go to the second stage. A threshold is needed to specify the number of candidates. By trying different choices, we found that 20 percent of all the possible frame pairs are sufficient for the second step. Figure 6 shows the D_{rp} matrix and candidate similar frame pairs.

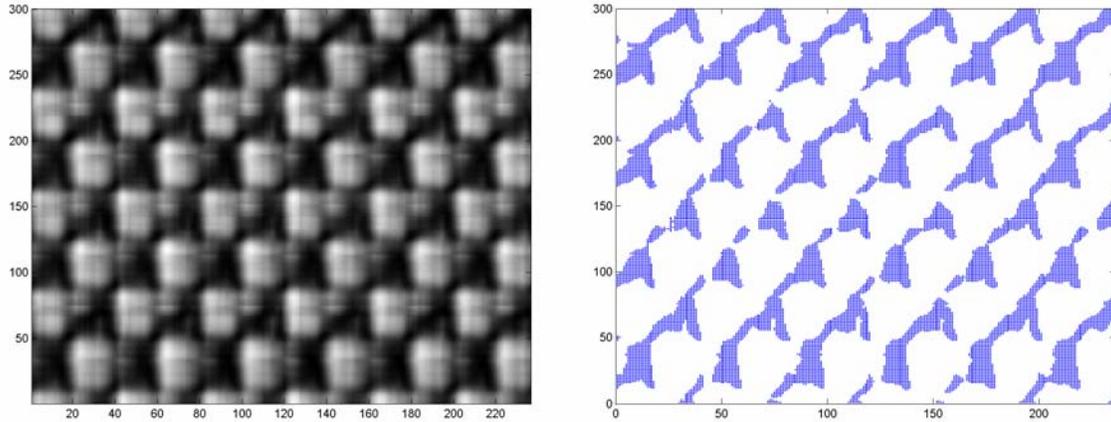


Figure 6: D_{rp} computation between walking and running. Left: D_{rp} matrix. Right: Candidate similar frame pairs.

3.2 Curvature Cross Correlation

In the second stage, the global motion curve of each joint is generated and its curvatures are calculated. Then cross-correlation is used to find the corresponding similar frames based on the curvature information.

The space curve of a joint is calculated as given in Equation (3.3). The offset vector of the joint \mathbf{v} is multiplied by the transformation matrix of its parent, then the transformation matrix of its grandparent and so on, until the transformation matrix of the root is reached. This operation is performed in the same manner as that in determining the joint positions in the global coordinate frame.

$$\mathbf{r}(t) = M_{root} \dots M_{grandparent} M_{parent} \mathbf{v} \quad (3.3)$$

When a space curve is represented in parametric form, its curvature is calculated according to Equation (3.4).

$$\kappa(t) = \frac{\|\mathbf{r}'(t) \times \mathbf{r}''(t)\|}{\|\mathbf{r}'(t)\|^3} \quad (3.4)$$

$\mathbf{r}'(t)$, $\mathbf{r}''(t)$ are, respectively, the first and second order derivatives of the space curve at time t . In practice, the cubic B-spline is used to approximate the trajectory curve, so that the first and second order derivatives can be directly evaluated. Before the B-spline approximation, the curve is smoothed by a Gaussian filter to make the curvature value more stable. Figure 7 shows the results with smoothing and without smoothing.

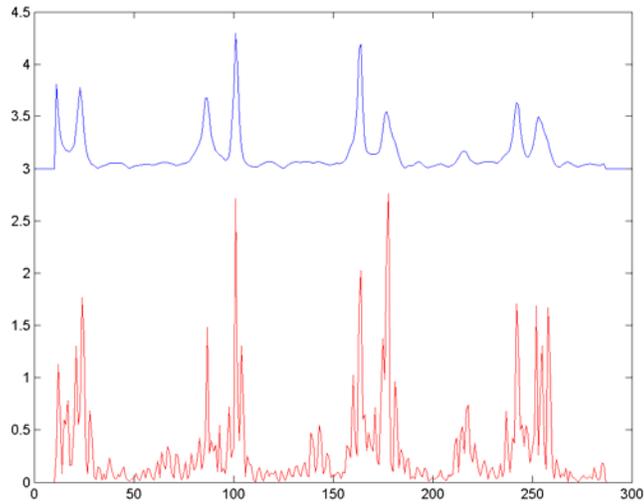


Figure 7: The curvature of the motion curve for the joint leftLowArm in a walking clip, where the top signal is the curvature (shifted up by 3) calculated with a Gaussian filter and the bottom signal is the curvature calculated without a Gaussian filter.

Curvature is an unsigned value that measures how rapidly the curve pulls away from its tangent. Therefore, it is used to represent the kinematic information of joint movements. The curvatures of each joint at frame t form a feature vector for one motion stream. The dimension of the vector is determined by the number of joints.

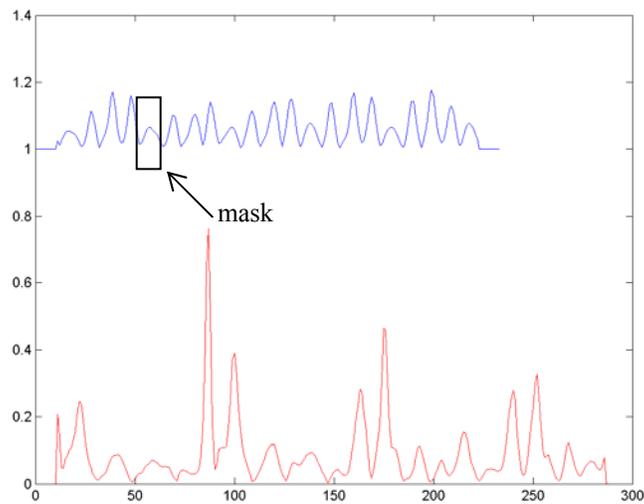


Figure 8: Cross correlation operated on the feature vectors of two motion clips.

To identify similar movements, cross-correlation is applied, which is a standard method of estimating how two signals are correlated. It is a widely used technique to identify similar 2D patterns in computer vision. Here, the feature vector defined above in the neighborhood of a frame forms a pattern mask. The mask slides over the second motion clip. If a similar pattern can be found in the second clip, then the corresponding frame is similar to the frame in the middle of the mask in terms of curvature (see Figure 8). The correlation between two frames is defined as in Equation (3.5).

$$corr_{i,j} = \sum_{k=1}^m w_k \sum_{l=-N/2}^{l=N/2} (\kappa_{i+l,k} - mean_{i,k}) (\kappa_{j+l,k} - mean_{j,k}) \quad (3.5)$$

where N is the size of the neighborhood. w_k is the joint weight to control the importance of the k th joint. Frame i and its eight neighbors form a mask. We have found empirically that 9 frames is suitable for the size of the mask, since most transitions happen in a short interval, ranging from 10 to 30 frames. $\kappa_{i+l,k}$ is the feature's k th dimension value of the l th neighbor of frame i ; $mean_{i,k}$ is the feature's k th dimension mean value of frame i . $\kappa_{j+l,k}$ and $mean_{j,k}$ are similarly defined for frame j . According to the cross correlation result, the corresponding frame pairs with larger correlation values are considered as similar frame pairs. The left image of Figure 9 shows the cross correlation result between a walking and a running sequence, where the whiter areas have larger cross correlation values. Note that only the cross correlation results for the

candidate frame pairs are computed, while the others are assigned the minimum of the computed cross correlation values to have the gray scale drawing. The right image in Figure 9 shows similar frame pairs as the output of our system.

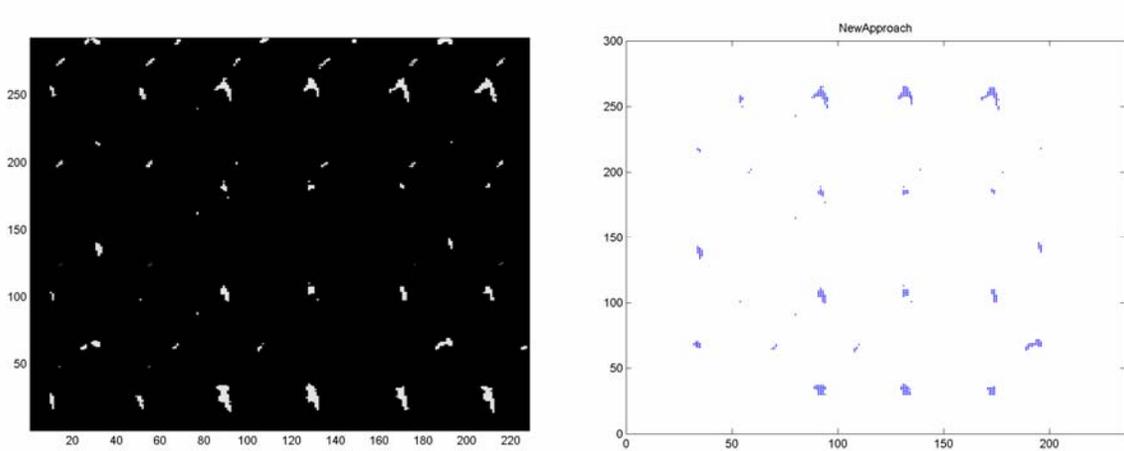


Figure 9: Left: Cross correlation results. Right: Similar frame pairs.

4 Experimental Results

In this section, we present the experimental results on motion similarity analysis using the novel approach described in the previous section. The other 3 previous approaches for motion similarity analysis are also implemented. Then experiments designed to compare the performance of different methods are carried out. We focus on three major performance aspects: efficiency, accuracy and visual acceptability. All approaches are implemented on an average PC (1.16 GHz P4 with 2 GB of memory running Windows XP Professional) using C++ and Maya 5.0. The motion dataset includes a wide variety

of human motions, such as walking, running, dancing with different styles, and on different levels.

4.1 Run Time Comparison

Table 2 shows the run time comparison between different approaches, where the similarity computation time between walking (300 frames) and running (236 frames) is shown. Given a motion dataset involving n clips, the total time needed for motion similarity analysis is *Run Time* (in Table 2) $\times \frac{n(n+1)}{2}$, when the number of frames for each motion clip is in the same range, that is, from 236 to 300 frames. The run time of the different motion similarity analysis methods ordered from fast to slow is: the curvature and D_{rp} approach, the joint positions and velocities approach, the joint angles and velocities approach, and the point clouds approach. The results demonstrate that our approach is very efficient and can significantly improve the run-time performance for motion similarity analysis.

Table 2: Run time to compute the similarity between two motion clips for different motion similarity analysis methods.

Approaches	Run Time (seconds)
Curvature and D_{rp}	1.842
Joint positions and velocities	7.585

Joint angles and velocities	9.844
Point clouds	65.422

4.2 Accuracy Study

4.2.1 Accuracy Criteria

From the proceeding discussion, it is already known that similar frame pairs should match in terms of skeletal configurations, joint velocities and joint accelerations. Thus, the difference in positions, velocities and accelerations can be applied to indicate how similar a frame pair is. For that reason, in our experiments, three criteria are designed to evaluate the qualities of the results between different similarity analysis methods. They are, respectively, the velocity difference, the acceleration difference and the position difference (D_v , D_a and D_p for short), and defined in the following equations.

$$D_v = \sum_{i,j} \sum_{k=1}^m w_k \|\mathbf{v}_{i,k} - \mathbf{v}_{j,k}\|^2 \quad (4.1)$$

$$D_a = \sum_{i,j} \sum_{k=1}^m w_k \|\mathbf{a}_{i,k} - \mathbf{a}_{j,k}\|^2 \quad (4.2)$$

$$D_p = \sum_{i,j} \sum_{f=1}^F w_f \sum_{k=1}^m w_k \|\mathbf{p}_{k,f}^i - M_{i,0} M_{j,0}^{-1} \mathbf{p}_{k,f}^j\| \quad (4.3)$$

where \mathbf{v} , \mathbf{a} and \mathbf{p} represent, respectively, the velocity, acceleration and position. i and j are the indices of the similar frame pair; k is the joint index; m is the number of joints;

w_k is the joint weight. f is the frame index within the neighborhood of the compared frame; F is the number of frames in the neighborhood; w_f is the frame weight. $M_{i,0}$ is the transformation matrix of the root joint at frame i ; $M_{j,0}^{-1}$ is the inverse of the transformation matrix of the root joint at frame j . When the transition happens from frame i to frame j , the matrix $(M_{i,0}M_{j,0}^{-1})$ is used to align the two clips. Therefore, D_p is able to roughly measure the similarity between the two compared frames and their neighborhoods. Actually, the physical interpretation of D_p is similar to that of the minimal position difference in the point clouds approach. But they are different in the following two aspects. In our approach, a 3D transformation matrix is used to align the two point sets, while a 2D transformation matrix is used in the point clouds approach. Additionally, the 3D transformation matrix is directly computed from the transformation matrix of the root joints at the two compared frames, no other joints are involved. In the point clouds approach, a more complex computation scheme involving the important joints is carried out to compute the 2D transformation matrix, and makes the computation cost more expensive. For example, on the same PC, it takes about 65 seconds to compute the point clouds distance matrix between two clips (300frames \times 236frames), while it takes about 40 seconds to compute D_p for the same two clips. Moreover, since the position difference D_p can individually capture the similarity between two frames and their neighborhoods, it is more important than the other two measures, and is used as a major evaluation criterion in our evaluation scheme.

To compute the above three measures, the same number of similar frame pairs are generated between any two clips in the motion datasets for each approach. The quick sort algorithm is used to order the computed similarity measure between any two motion clips. The frame pairs with a higher similarity are kept as the results. Note that the curvature and D_{rp} approach sorts the candidate frame pairs according to the cross-correlation values in the second step.

4.2.2 Optimal Attribute Weights Computation

The measure D_p can provide guidelines to find an optimal attribute weight set needed in the approaches based on the sum of multiple joint attributes. At this point, the computed optimal attribute weights are used to control the relative importance of the different joint attributes, such as joint positions, joint angles, root velocities and accelerations. The optimal attribute weights are different from joint weights which control the importance of different joints. In our experiments, we adopt the joint weight set as that specified in [9]. The weights of the important joints are set to one, while the unimportant ones are set to zero. All the other approaches use the same joint weight set.

The process to compute the optimal attribute weights needed in the approaches based on the weighted sum of multiple joint attributes is described in detail as follows.

- Design an initial attribute weight search space. For example, each weight has 10 possible values evenly distributed in the range $(0, 1]$.
- For each possible attribute weight set, calculate D_p . Then the attribute weight set with the minimum D_p corresponds to the optimal attribute weight set in the current search space.
- Refine the attribute weight search space if necessary.

The exhausting search for an optimal attribute weight set is very time consuming and the optimal attribute weight set depends on the motion patterns in the dataset. For example, in the approach based on joint angles and velocities, when the dataset includes only 2 clips (each has 150 frames), and the search space is $10 \times 10 \times 10$, it takes more than 4 hours to find the optimal attribute weights. When the search space increases to $20 \times 20 \times 20$, it needs about 34 hours. In the first dataset, which includes running and walking only, the optimal attribute weight set is $(0.1, 1.0, 0.1)$. In the second dataset, which includes dancing and running, the optimal attribute weight set becomes $(1.0, 1.0, 0.1)$. If no measures such as D_p are defined to help detect the optimal attribute weights automatically, then the animators have to estimate the values of these attribute weights by trial-and-error, which is obviously a very tedious and labor intensive exercise. Furthermore, the quality of the results depends on the experience and skills of the animator. So compared with these techniques, our approach has the major advantage that it does not require the animator to set any attribute weights.

4.2.3 Results for Accuracy Evaluation

In this section, the results for accuracy evaluation are presented when different motion similarity analysis techniques are applied to different datasets.



Figure 10: Two motion clips on the same floor plane.

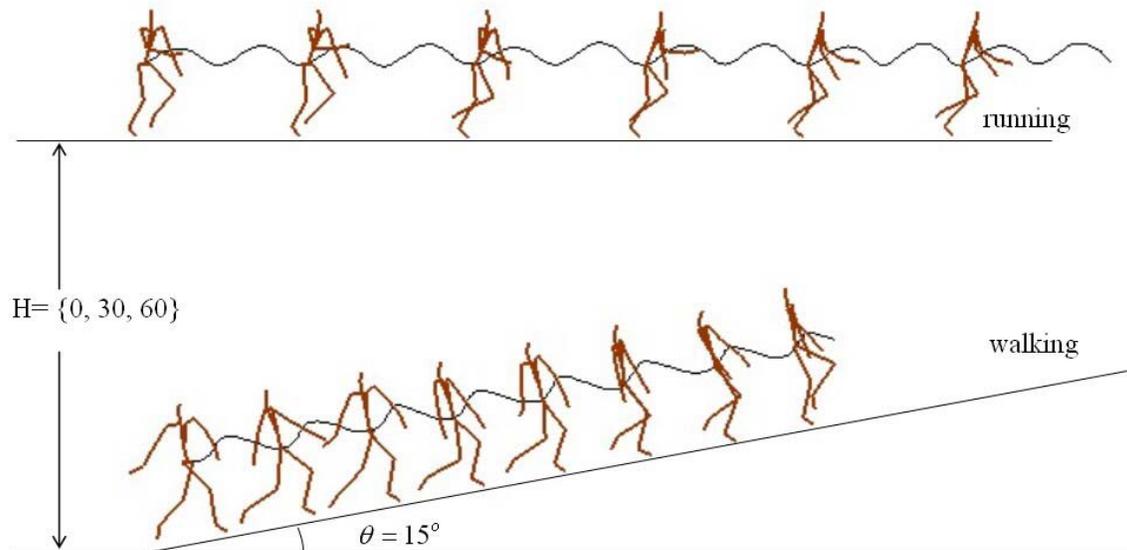


Figure 11: Two motion clips on different levels. Top: Running on the floor plane. Bottom: Walking along a ramp with a slope of 15 degrees. H represents the vertical distance by which the walking clip is shifted down from the floor plane.

4.2.3.1 Motions on Different Levels

The two clips are running and walking on the same floor plane shown in Figure 10. In the different level cases, the walking is modified by rotating the root joint around the z-axis by 15 degrees so that walking on the ground plane is changed to walking on a slope. Then two more walking clips (see Figure 11) on a ramp are generated by shifting the rotated walking clip down by 30 and 60 cm. Thus, the effects of the vertical distance between the root joints of the two compared frames can be investigated.

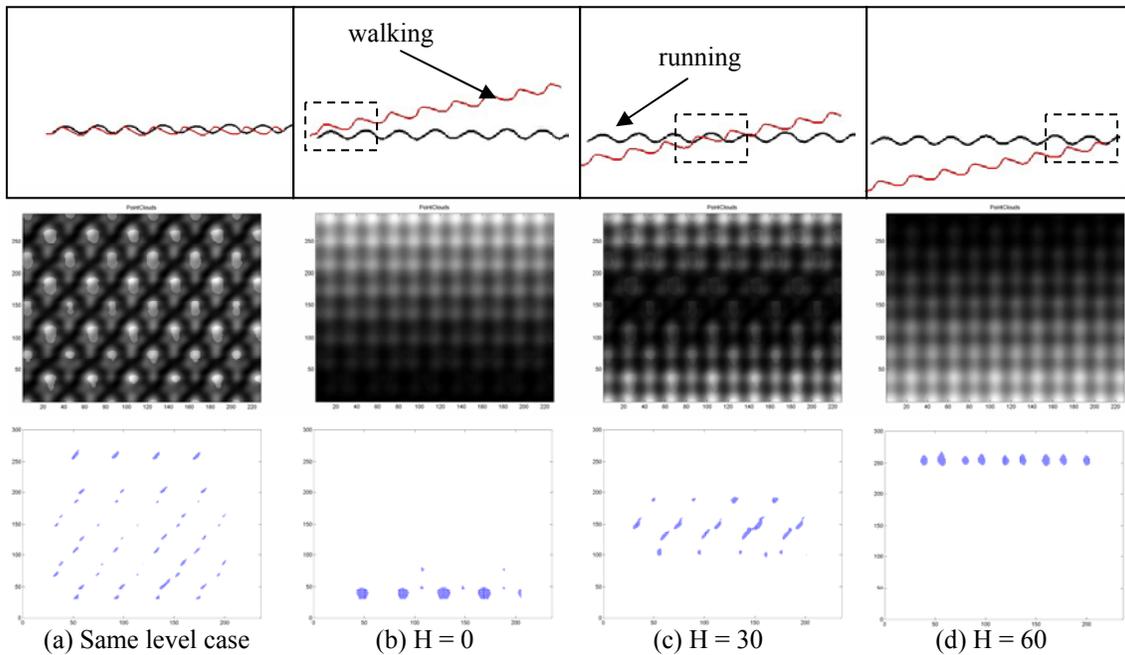


Figure 12: Similarity analysis based on point clouds. (a) Two clips are on the same floor plane; (b)(c)(d) on 3 different level cases. Top: The root trails of walking and running. The dotted boxes mark the areas with a smaller vertical distance of the root joints. Middle: The distance matrix (darker areas imply a lower distance measure). Bottom: The resulting similar frame pairs denoted by dots. The rows represent the frames in the running clip, while the columns represent the frames in the walking clip.

The top images in Figure 12 show the trails of the root joints in walking and in running, where the areas with a smaller vertical distance of the root joints are marked by the dotted boxes. The middle and bottom images show the similarity analysis results in the point clouds approach between the same level case and 3 different level cases. Since the point clouds approach uses a 2D rigid transformation matrix to compute the minimal distance between two point sets, the similarity results obtained by this method are quite different between the four compared cases. When the vertical distance between the root joints of the two clips increases, the difference measure increases, or the similarity measure decreases. While in the other three techniques, since all the motion features are coordinate invariant, the same results are achieved between the same level case and the different level cases for each technique (shown in Figure 13). Figure 14 shows the performance comparison between the same level case and 3 different level cases. The results demonstrate that the accuracy performance of the point clouds approach decreases largely in the different level cases when compared with that in the same level cases; while the other approaches keep the same performance in both cases. Here, in the approaches based on the weighted sum of multiple joint attributes, all the attribute weights are set to one. According to the criterion D_p , the performance of the approach based on curvature and D_{rp} , the approach based on joint positions and velocities, the point clouds approach in the same level case are much better than the performance of the approach based on joint angles and velocities. It is because the unit weight set used in the joint angle approach is not suitable for the test motion clips.

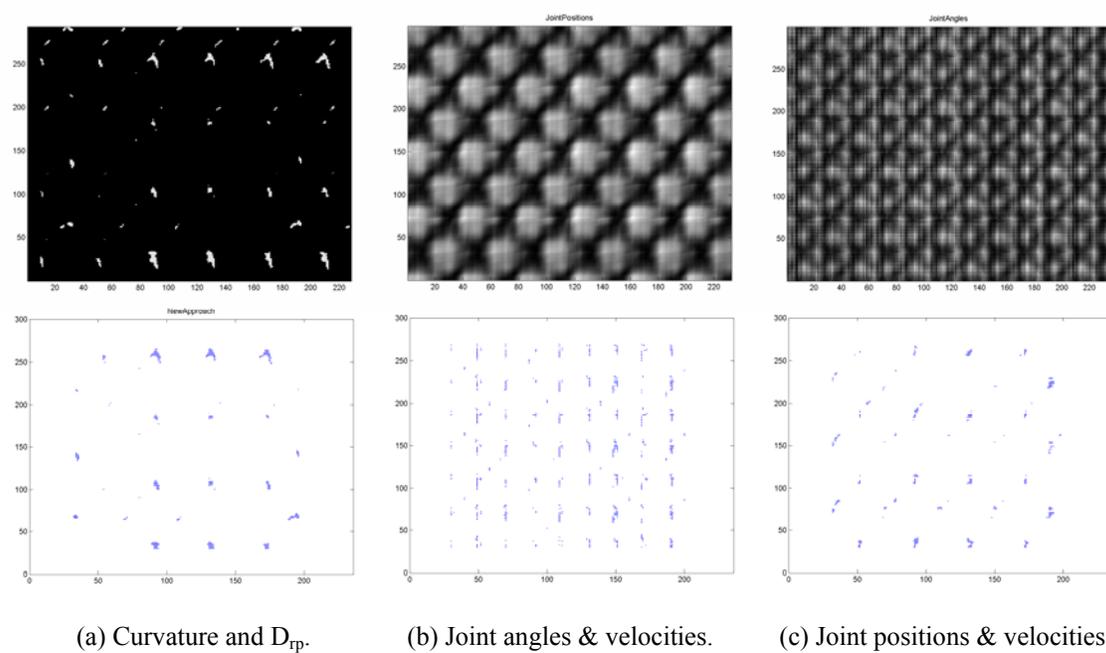


Figure 13: Similarity analysis based on: (a) Curvature and D_{rp} ; (b) Joint angles and velocities; (c) Joint positions and velocities. Top left: Cross-correlation results. Top middle and right: Distance matrix. Bottom: The resulting similar frame pairs.

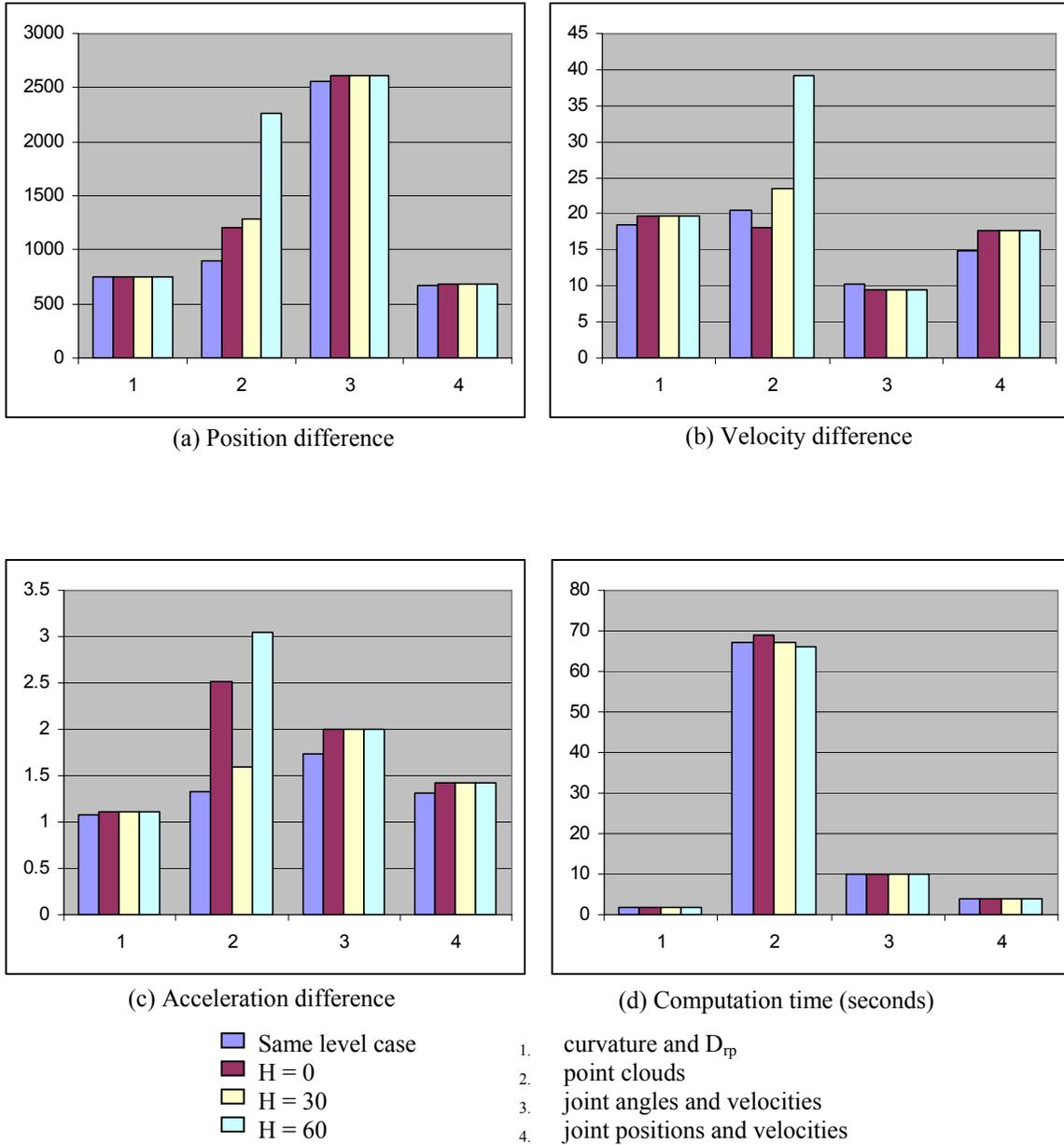


Figure 14: Performance comparison between the same level case and the 3 different level cases. In the different level case, the performance for the point clouds approach decreases largely when compared with its performance in the same level case; while the other approaches perform almost the same in both cases.

4.2.3.2 Optimal Attribute Weight Sets Computation

In the joint angle approach, three attribute weights ($w_p : w_a : w_v$) are used to accommodate for the relative importance of the difference of root velocities, the difference of joint angles and the difference of joint velocities. To get a suitable attribute weight set needed in the joint angle approach, the search process described in 4.2.2 is operated in two search spaces. Both are $10 \times 10 \times 10$, one evenly distributed, and the other non-uniformly distributed. Two optimal attribute weight sets are obtained for the test motions. Figure 15 shows the difference in the results generated by using the joint angle approach with different attribute weight sets. In Figure 16, the performance for the approach based on joint angles and velocities is presented with different attribute weight sets. The results demonstrate that the performance with the optimal attribute weight sets is greatly improved when compared with the user specified attribute weights. Movie 1 also shows the difference in the results by using a user specified attribute weight set and the improved optimal attribute weight set.

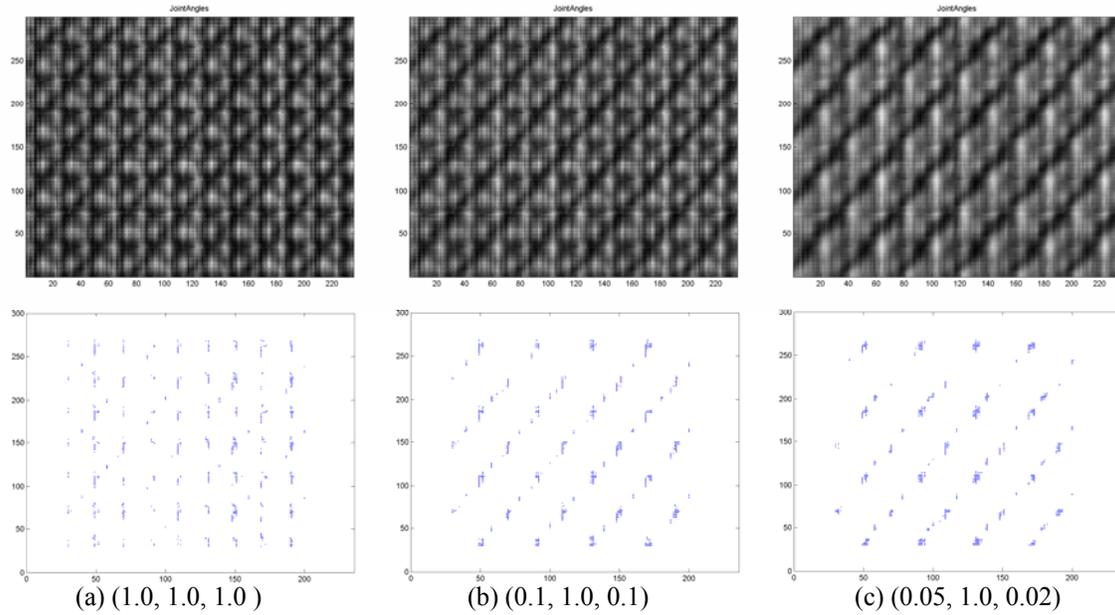
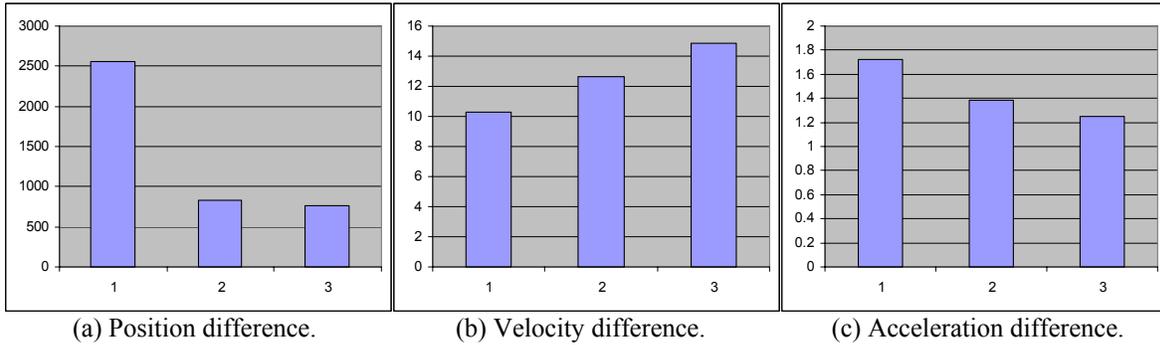


Figure 15: Similarity analysis based on joint angles and joint velocities with different attribute weight sets. Left: A user specified attribute weight set. Middle and right: Two optimal attribute weight sets. Top: The distance matrix. Bottom: The resulting similar frame pairs denoted by dots.



1. A user specified weight set { 1.0 : 1.0 : 1.0 }
2. Optimal attribute weight set One { 0.1 : 1.0 : 0.1 }
3. Optimal attribute weight set Two { 0.05 : 1.0 : 0.02 }

Figure 16: Performance comparison for the joint angle approach by using different attribute weight sets. The performance of the joint angle approach is largely improved by using the computed optimal attribute weight sets according to D_p .

4.2.3.3 Evaluation between Different Datasets

Three different datasets have been tested using different motion similarity analysis approaches. Set A includes a walking and running clip (300×236 frames). Set B includes a rocknroll and a highwire walking (260×300 frames). Set C has 12 motion clips with a variety of human behaviors, such as sneaking, drunk-walking and dancing. For the three datasets, all the motions happen on the same level ground and the joint angle approach is applied with an improved attribute weight set. Figure 17 shows the similarity analysis results for Set B. The performance of different approaches for the 3 datasets is presented in Figure 18. The result demonstrates that the performance of different approaches varies a little when different datasets are applied. According to the position difference criterion, there is not too much difference in their performance in terms of accuracy or quality. This can be verified by the similarities of their performance in terms of visual acceptability presented in the next section. Furthermore, the whole evaluation process shows that the two criteria, the velocity difference D_v and the acceleration difference D_a , cannot be independently used as an evaluation criterion.

Motion Similarity Analysis and Evaluation of Motion Capture Data

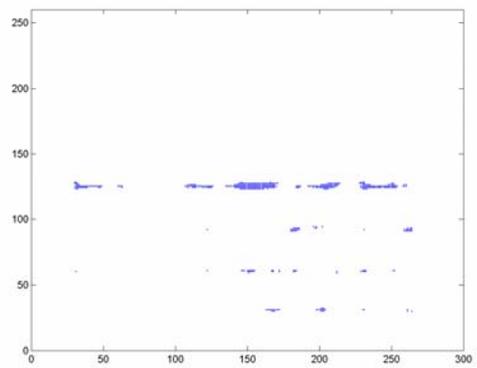
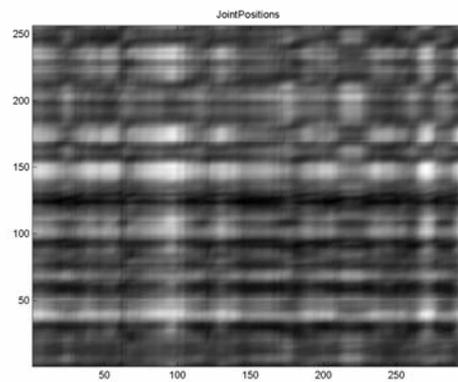
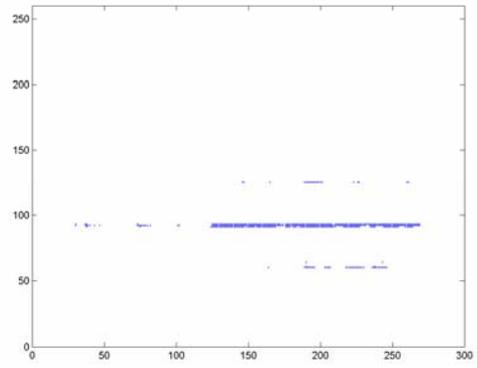
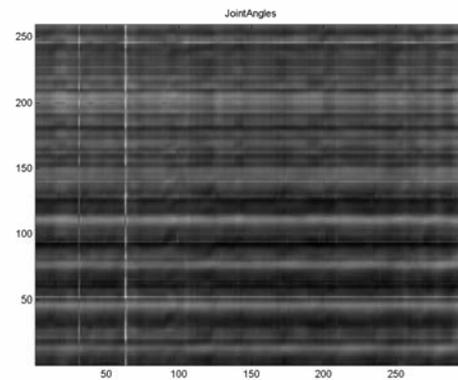
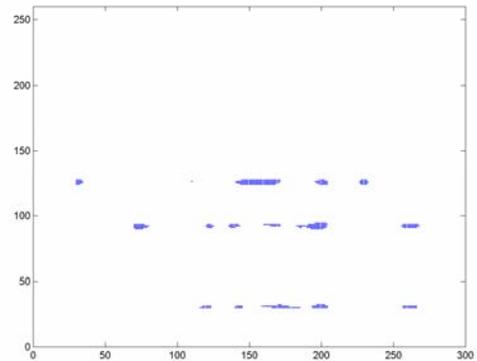
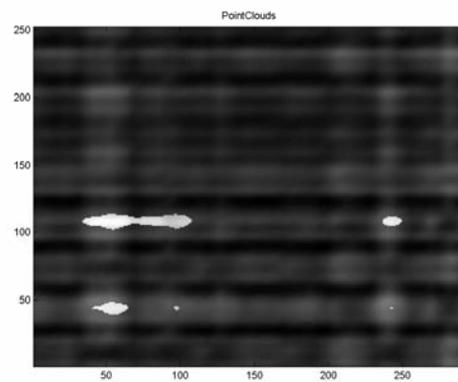
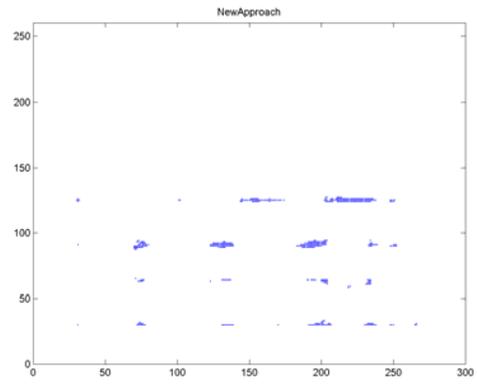
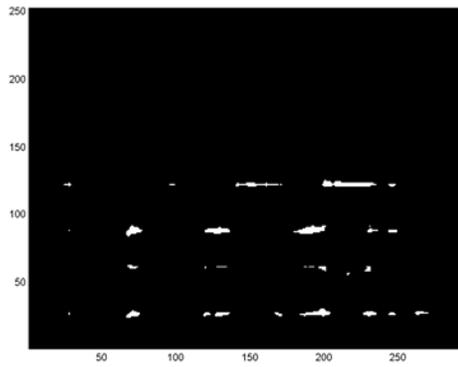


Figure 17: Motion similarity analysis results for Dataset B. Left: The distance matrices. Right: The resulting similar frame pairs. First row: Curvature and D_{rp} approach; Second Row: The point clouds approach. Third Row: The joint angles approach. Fourth Row: The joint positions and velocities approach.

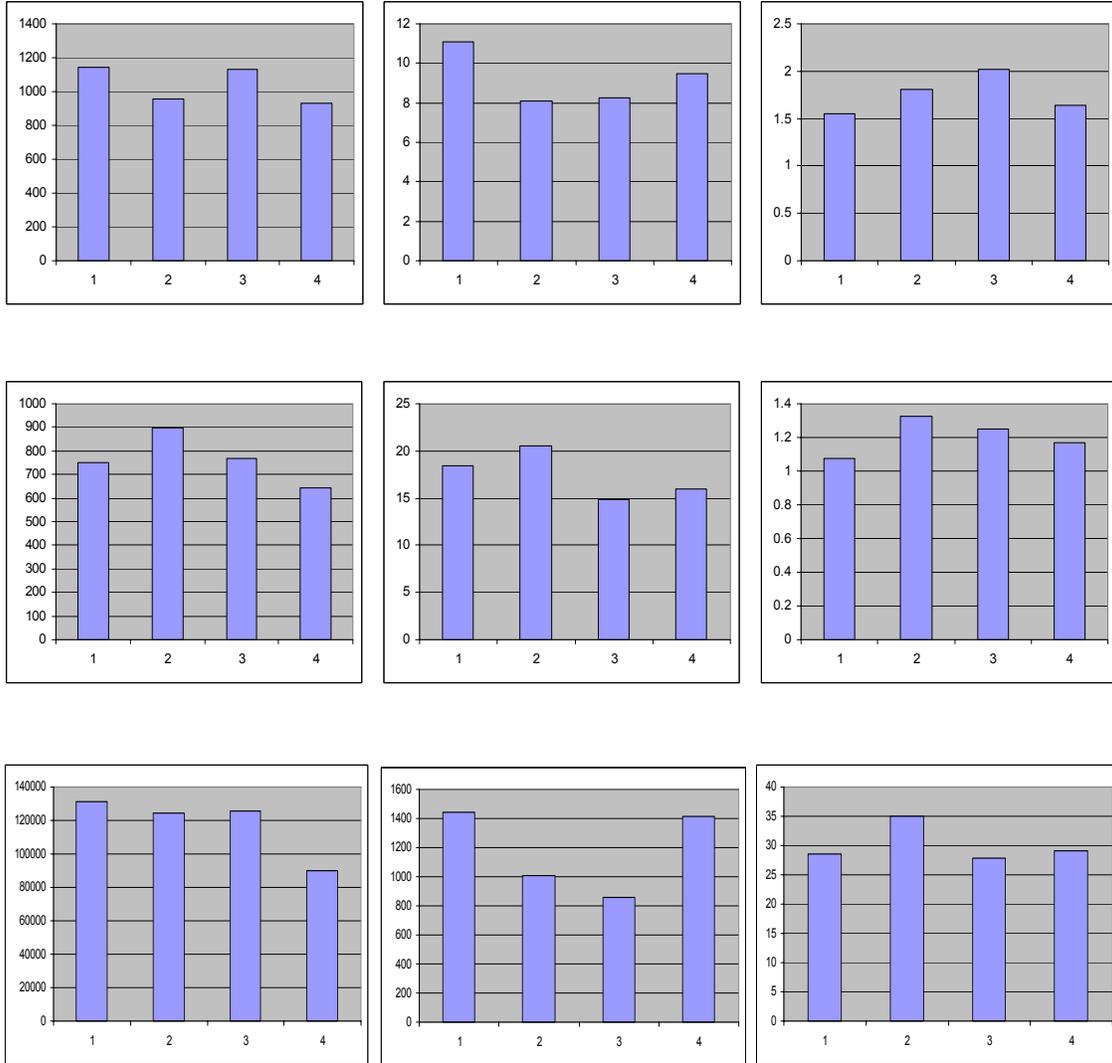


Figure 18: Performance comparison between 3 different datasets. Top row: Dataset A, a walking and a running (300×236 frames). Middle row: Dataset B, a rocknroll and a highwire walking (260×300 frames). Bottom row: Dataset C, 12 different motion clips. Left: the position difference; middle: the velocity difference; right: the acceleration difference.

The results of our evaluation between different approaches are summarized in Table 3.

Table 3: Comparison between different approaches for motion similarity analysis.

	Our Approach	Point Clouds	Joint Angle	Joint Velocity & Joint Acceleration
Running Time	Fast	Slow	Medium	Medium
Optimal Attribute Weight Determination	No	No	Yes	Yes
Coordinate Invariant	Yes	No	Yes	Yes
Velocity Difference	Acceptable	Good	Good	Acceptable
Acceleration Difference	Good	Acceptable	Good	Good
Position Difference	Acceptable	Acceptable	Acceptable	Good

4.3 Visual Acceptability

To visually assess the performance of the similarity analysis results, a hierarchical graph is built based on the results of different motion similarity analysis methods. Then according to the user's specifications, new motion sequences are created by stitching different motion clips together. Since motion transition points are detected based on motion similarity analysis, the quality of the resulting motion sequence directly depends on the similarity analysis results. In our experiments, linear interpolation is applied to distribute the discontinuities in the transition region. The experimental results demonstrate that the smooth and natural transitions can be achieved at the similar frame pairs generated by using the curvature and D_{rp} approach. Figure 19 shows the transitions between walking and running. The left columns show the skeleton poses at selected

similar frame pairs. The middle and right columns are the transition results without and with linear smoothing. Movie 2 to Movie 4 demonstrate that our approach can generate visually appealing transitions. Movie 5 shows the transition results based on the other similarity analysis techniques.

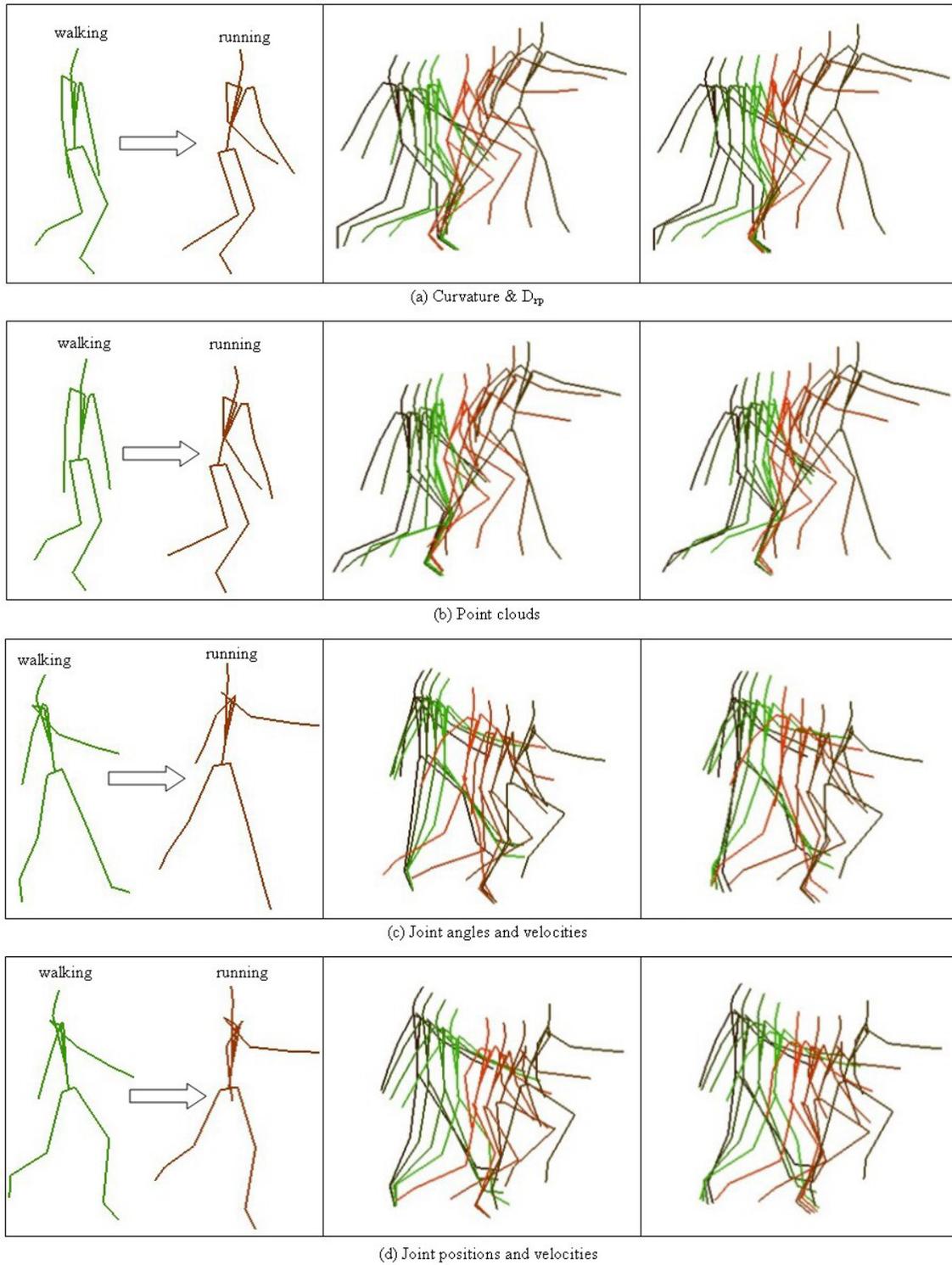


Figure 19: Transitions between walking (green) and running (red). Left: Skeleton poses at the selected similar frame pairs. Middle: Transitions without linear interpolation.

Right: Transitions with linear interpolation. The screen shots are down-sampled by a factor of 3.

5 Conclusions and Future Work

In this paper, two typical motion similarity approaches are investigated and the related problems are identified. The approach based on a weighted sum of multiple joint attributes requires attribute weight detection, while the point clouds approach is coordinated variant and not efficient. To tackle the problems existed in the previous approach, a novel method for motion similarity analysis is designed and developed. Two motion features: the curvature of space curve and the joint relative positions are presented to estimate the similarity between two motion frames. The experimental results show that our approach is very efficient and is completely coordinate invariant. And visually acceptable transition results can be generated. We also introduce a general criterion to evaluate the performance of different methods for motion similarity analysis. By using this criterion, better attribute weight parameters can be found to improve the results of the approach based on the weighted sum of different joint attributes.

For the approach based on joint positions, velocities and accelerations, as the similarity is represented as the weighted sum of three or more components, the appropriate attribute weight set is very difficult to determine. Moreover, for the dataset with different motion patterns, the attribute weight set is likely to change. In our approach, although two motion features are used, the two-step process instead of the weighted sum avoids the attribute weight selection. Though a threshold is specified in the first stage,

we have found a fixed range (20%~25%) empirically. Meanwhile, in the case of the point clouds approach, the minimum distance between two point sets is computed by aligning them with a 2D rigid transformation matrix. As a result, this method cannot correctly deal with motions on different plane levels. The universal criterion D_p presented in this paper is different from the minimum distance in the point clouds approach for the following two reasons. First, D_p is the distance after two point sets are aligned by a 3D transformation matrix instead of a 2D transformation matrix. Second, the 3D transformation matrix is directly computed from the 6 DOF of the root joints of the two frames. Therefore, D_p is more general and easily computed. And according to D_p , an optimal attribute weight set can be found for approaches that require attribute weight determination.

In our approach, two novel motion features are proposed to describe the similarity between two motion frames. The joint relative positions capture the skeletal pose information, while the curvature of space curve formed by the joint movements capture the kinematic information. Compared with the previous approach for motion similarity analysis, the new curvature and D_{rp} approach has three major advantages. First, it has a high efficiency. It is because the two features can be computed directly from motion capture data. Second, it is coordinate invariant. The two selected features are coordinate independent. Thus, the proposed approach can correctly handle a wide range of motions,

particularly, the motions on different levels. Finally, no attribute weight determination is required in our approach. A two-step process does not require attribute weight selection.

However, there are several limitations in our work. When processing stationary motions, the curvature calculation becomes unreliable that may ruin our similarity analysis results. So our approach is restricted to the motions where most of the important joints keep continuously movement. But the accuracy of the curvature computation can be improved by using the ENO schemes (Essentially Non-Oscillatory) that are introduced in [6]. Secondly, only a brute-force sampling method is employed to illustrate how to use the more general measure D_p to find the optimal attribute weight set. In the future, we can try some classic optimization methods to improve the computation process for optimal attribute weight set. Another issue needs to consider is the design of Gaussian filter used to smooth the space curve of joint movement. Technically, it depends on amount of noise in the raw motion data.

Additionally, there are several areas left open for future work. One possibility is to improve the efficiency of the curvature and D_p approach by performing the cross-correlation in the Fourier domain by using the FFT. As well, we would like to apply our work to much larger datasets with various motion patterns and to test the effects of using larger data sets on the performance of different approaches. Moreover, we will investigate if there are other motion features that are more suitable to represent and to identify motion similarities between two motion frames.

Acknowledgments

The authors would like to thank NSERC and the University of Alberta for financial support and Alias|Wavefront for the Maya licenses.

References

- [1] A. Ahmed, F. Mokhtarian, and A. Hilton. Parametric motion blending through wavelet analysis, short presentations. In *Proceedings of Eurographics*, pages 347-353, 2001.
- [2] O. Arikan and D. A. Forsyth. Interactive motion generation from examples. In *Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques*, pages 483-490. ACM Press, 2002.
- [3] A. Bruderlin and L. Williams. Motion signal processing. In *Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques*, pages 97-104. ACM Press, 1995.
- [4] M. Gleicher. Motion editing with spacetime constraints. In *Proceedings of the 1997 Symposium on Interactive 3D Graphics*, pages 139-148. ACM Press, 1997.
- [5] M. Gleicher. Retargetting motion to new characters. In *Proceedings of the 25th Annual Conference on Computer Graphics and Interactive Techniques*, pages 33-42. ACM Press, 1998.
- [6] A. Harten, B. Engquist, S. Osher and S. R. Chakravarthy. Uniformly high order accurate essentially non-oscillatory schemes, 111. *J. Comput. Phys*, 71(2):231-303, 1987.

- [7] L. Kovar and M. Gleicher. Flexible automatic motion blending with registration curves. In *Proceedings of the 2003 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, pages 214-224. Eurographics Association, 2003.
- [8] L. Kovar and M. Gleicher. Motion graphs. In *Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques*, pages 473-482. ACM Press, 2002.
- [9] J. Lee, J. Chai, P. S. A. Reitsma, J. K. Hodgins and N. S. Pollard. Interactive control of avatars animated with human motion data. In *Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques*, pages 491-500. ACM Press, 2002.
- [10] J. Lee and S.Y. Shin. A hierarchical approach to interactive motion editing for human-like figures. In *Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques*, pages 39-48. ACM Press/Addison-Wesley Publishing Co., 1999.
- [11] C. Rose, M. F. Cohen and B. Bodenheimer. Verbs and Adverbs: Multidimensional Motion Interpolation. *IEEE Comput. Graph. Appl.*, 18(5): 32-40, 1998.
- [12] M. Unuma, K. Anjyo and R. Takeuchi. Fourier principles for emotion-based human figure animation. In *Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques*, pages 91-96. ACM Press 1995.
- [13] J. Wang and B. Bodenheimer. An evaluation of a cost metric for selecting transitions between motion segments. In *Proceedings of the 2003 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, pages 232-238. Eurographics Association, 2003.

- [14] D. J. Wiley and J. K. Hahn. Interpolation Synthesis of Articulated Figure Motion. *IEEE Comput. Graph. Appl.*, 17(6): 39-45,1997.