Deformation-Based Abnormal Motion Detection using 3D Skeletons

Renato Baptista, Girum Demisse, Djamila Aouada, Björn Ottersten

Interdisciplinary Centre for Security, Reliability and Trust (SnT), University of Luxembourg, Luxembourg

{renato.baptista, girum.demisse, djamila.aouada, bjorn.ottersten}@uni.lu

Abstract— In this paper, we propose a system for abnormal motion detection using 3D skeleton information, where the abnormal motion is not known a priori. To that end, we present a curve-based representation of a sequence, based on few joints of a 3D skeleton, and a deformation-based distance function. We further introduce a time-variation model that is specifically designed for assessing the quality of a motion; we refer to a distance function that is based on such a model as *motion quality distance*. The overall advantages of the proposed approach are 1) lower dimensional yet representative sequence representation and 2) a distance function that emphasizes time variation, the motion quality distance, which is a particularly important property for quality assessment. We validate our approach using a publicly available dataset, SPHERE-StairCase2014 dataset. Qualitative and quantitative results show promising performance.

Keywords— 3D Skeleton, Temporal Analysis, Abnormal Detection, Deformation.

I. INTRODUCTION

Human motion analysis received a considerable attention from the computer vision community over the last few years. This is mainly due to a wide-range of applications that relies on effective motion analysis, *e.g.*, surveillance and human-computer interaction [1], [2]. In most applications, the main effort is on action representation [1], [3] and action recognition [4]–[6]. Consequently, the research is focused on the modelling and representation of an action sequence properties that are invariant to temporal and geometric variation. On the contrary, in applications like abnormality detection [2], the main focus is on analyzing the quality of the motion, given a predefined action. Hence, a motion quality assessment research emphasizes modeling temporal and geometric variations; such systems have several important real-world applications, *e.g.*, in sports [7], and in healthcare [8].

In this work, we are interested in a motion quality assessment for healthcare related applications, *e.g.*, home-based rehabilitation for stroke patients [9]. In the case of stroke survivors, a regularly monitored exercise is particularly important to recover some autonomy in their daily life activities [10]. Post-stroke patients are initially subjected to physical therapy sessions under the supervision of a healthcare professional, who usually suggests activities for a home-based rehabilitation [11]. In such scenarios, having an automated tool that assesses the quality of the patient's exercise could be of great importance. This would facilitate the monitoring of the patients' progress by the physiotherapist, enabling an intervention in case of continuously less satisfactory reports.

In general, the quality of a given motion is estimated by measuring the deviation from what is normal. Such approaches may fall into the topic of abnormal motion detection. Abnormalities are usually detected by comparing a given motion with respect to a model of a normal motion [12]–[14]. In order
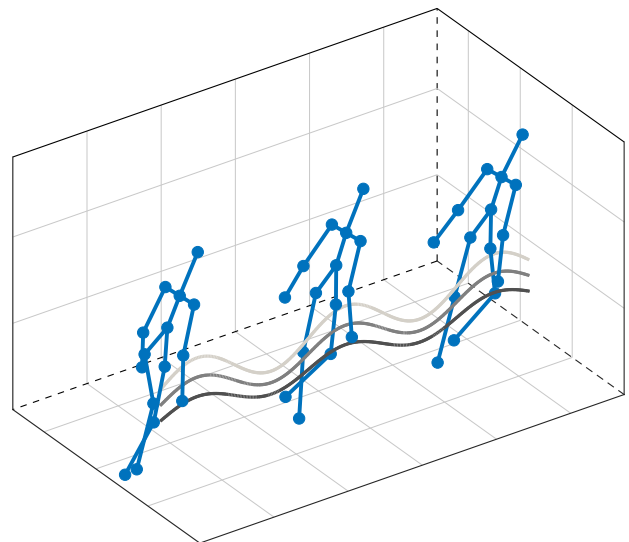

Fig. 1: Illustration of the proposed representation: The figure shows the proposed curve-based sequence representation derived from the knee joints only.

to detect abnormalities, two main approaches exist: 1) where the abnormalities are known a priori; and 2) where there is no information of what an abnormality is. In [15], a binary classifier is trained to classify an observation into either normal or abnormal. In contrast, works such as [13], [14], [16], [17] presented methods where the model of a normal motion is learned and the abnormalities are classified as deviations, given some threshold, from the model.

In this work, similar to [13], [14], [16], [17], we build a model from a set of normal motions and detect abnormalities based on a predefined threshold. We begin by representing a skeleton sequence as a curve by tracking few selected joints of the skeletons throughout the sequence. Contrary to [13], [14], we do not apply a manifold-based dimensionality reduction method. Instead, similar to [8], we select few joints using problem-specific knowledge to build a 2-dimensional representation from a $3 \times N$-dimensional data, for N skeleton joints, see Fig. 1. With this curve representation, a distance function on curves could then be used to measure the dissimilarity between two sequences. We propose to adapt the *deformation-based distance* function, introduced in [18], for this purpose. Subsequently, we model normal motion sequences by computing the average of their representation.

We note that the speed, or latency variation in between sequences, contributes significantly to a distance function–mainly due to mismatching frames. Hence, the representativeness of a model depends highly on how time-variation is addressed. We thus introduce the concept of a *motion-quality* distance which is based on detecting matching frames (key-frames), then dividing

the observed sequences into matching sub-sequences. The latter are then used to compute a series of sub-distances. This approach is different from a simple linearization as proposed in [19]. We define the motion-quality distance such that the variation in time is emphasized, while the optimal alignment proposed in [19] emphasizes invariance to variation in time.

We show that in assessing the quality of motion, our approach leads to high performance in comparison to optimal alignment, hence the name *motion-quality* distance. The following list summarizes our main contributions:

1) We propose to represent a 3D skeleton sequence as a curve, reducing the dimensionality of the data significantly. Subsequently, we show that the curve-based representation has a comparable performance to a related motion quality assessment approach [13], [14].

2) We propose to detect key-frames using the curve-based representation in order to quantify variation in time. However, we emphasize time-variation as opposed to emphasizing invariance. We show that emphasizing time-variation, in motion quality assessment, leads to high performance.

The rest of the paper is organized as follows: in Section II a brief review of related works is presented. Section III describes the problem formulation of an abnormal motion detection system followed by the proposed approach in Section IV. Experimental results are presented in Section V, and finally, Section VI concludes the paper.

II. RELATED WORK

The interest in the area of assessment of the quality of human actions has been increasing very rapidly over the last years. In this section, we briefly review vision-based methods to assess the quality of an action and methods related to the abnormality detection in human motion analysis.

Assessment of the quality of actions: Assessing the quality of an action can be thought as a problem of measuring how close an action is from a reference action. In other words, giving a score measured by the matching between an action and a pre-trained model [13], [14]. Pirsiavash *et al.* [7] proposed a framework for learning how to assess the quality of actions from RGB videos. In their work, a regression model is trained from spatio-temporal pose features such that it predicts scores of actions. Training a regression model requires a large number of annotated data. Using a different type of input data, Wang *et al.* [8] presented a method to address the problem of automated quantitative evaluation of muskulo-skeletal disorders of patients who suffer from Parkinson's disease using a 3D sensor. They selected a reference motion by choosing one cycle of the motion and performing temporal alignment with the remaining motion cycles. The feature selection is movement dependent (step size, stepping time and the swing level of the hands), which makes it undesirable to be generalized to other motions. The authors of [14] proposed a framework to assess the quality of actions, where they evaluate the deviation of an observation from a learned model of normal human motion. They used Hidden Markov Models (HMM) to model the dynamics of the human motion from skeleton-based samples of healthy individuals. This can be seen as an abnormal motion detection system, where the deviation from a model of normal motion is estimated.

Abnormality Detection: Abnormality detection refers mainly to the problem of finding patterns in data that do not conform to expected behavior [12]. There are two main

approaches for abnormality detection: 1) the abnormality is known as prior knowledge, and 2) the abnormality is not known beforehand. In the first approach, the work of Parra-Dominguez *et al.* [15] presented a supervised learning method to learn the abnormalities during stair descent (fall detection). A binary classifier is then trained on annotated data in order to decide whether the motion is normal or abnormal, while in the second approach there is no knowledge of what an abnormality is. Nater *et al.* [16] proposed to learn a model of normal human behavior in an unsupervised way. The model uses a hierarchical representation of appearance and action level of normal movements to detect abnormalities (fall detection) from silhouettes. Snoek *et al.* [17] trained an HMM using sequences of normal staircase motion to detect abnormal motion during stair descent from RGB data. The closest work that is most related to ours is [13]; there, the authors presented an approach that detects abnormal events and provides an assessment of the quality of the motion on a frame-by-frame basis. The work is based on a continuous statistical model that is built from a set of normal human motion using 3D human skeleton data. A non-linear manifold technique was used in order to reduce the dimensionality of the skeleton information.

III. PROBLEM FORMULATION

In this section, we briefly describe the problem formulation of an abnormal motion detection system.

Let $\mathsf{S} = \{\mathbf{j}_1, \cdots, \mathbf{j}_N\}$ be the skeleton pose of a human subject with N joints, where each $\mathbf{j}_i \in \mathbb{R}^3$ is an estimate of the joint positions in 3D space. Subsequently, a skeleton-based abnormal motion detection system classifies a sequence $\mathsf{M} = \{\mathsf{S}_1, \cdots, \mathsf{S}_k\}$ as either normal or abnormal. In general, such a problem is formalized as

$$f(\mathsf{M} \mid \mathsf{M}_t) = \begin{cases} 0 & , \text{ if } \|\mathsf{M} - \mathsf{M}_{t_0}\| < \|\mathsf{M} - \mathsf{M}_{t_1}\| \\ 1 & , \text{ otherwise} \end{cases}, \quad (1)$$

where $\|\cdot\|$ denotes a distance function, and M_t denotes the set of parameters of the abnormal and normal movement models, *e.g.*, simple kernel based binary classifier [20].

In this paper, however, we aim to estimate an abnormal motion detection system from a set of normal motion sequence examples, without any prior knowledge of what an abnormal motion is. As a result, instead of solving for a binary classifier that maximizes a decision margin as given in (1), we introduce a motion representation approach $\mathcal{G}(\cdot)$ such that the distance between two normal motions is less than a given δ under a distance function $D(\cdot, \cdot)$ of the motion representation. Hence, (1) may be reformulated as

$$f(\mathcal{G}(\mathsf{M}) \mid \mathcal{G}(\mathsf{M}_t)) = \begin{cases} 0 & , \text{ if } D(\mathcal{G}(\mathsf{M}), \mathcal{G}(\mathsf{M}_t)) > \delta \\ 1 & , \text{ otherwise} \end{cases}, \quad (2)$$

where $\mathcal{G}(\mathsf{M}_t)$ represents the model of a normal motion.

In the following section, we describe our proposed approach for representing a normal motion such that performance of (2) is maximized.

IV. PROPOSED APPROACH

In this section, we describe the proposed approach for estimating (2) and discuss data variation due to motion velocity. In Section IV-A, we first detail the proposed motion representation function $\mathcal{G}(\cdot)$ and a descriptive distance function $\mathfrak{D}(\cdot, \cdot)$ followed by the description of the normal motion model, $\mathcal{G}(\mathsf{M}_t)$. Later in Section IV-B, we detail time-variation and its correspondence with the defined distance function.

A. Motion Representation

In specific problems like abnormal motion detection, descriptive movements of a motion sequence are captured by tracking fewer joints than is needed for general problems, *e.g.* action recognition [21], [22]. Hence, a given skeleton $S \in \mathbb{R}^{3N}$ is a high dimensional data point with redundant information. As a result, we define a data representation approach that reduces the dimensionality of the data by using a priori problem-specific knowledge.

In abnormal motion detection problems, the motion that is analyzed is conditioned on a specific and fixed action, i.e., every subject, in both training and testing dataset, is expected to perform a similar action. Hence, there is no variation in action class but in the manner that an action is performed. As a result, we select a subset of the joints that shows the largest variation in performing a specific action as representative joints. In this work, we select the left and right knee joints, which we denote by $\mathbf{j}_{lk}$ and $\mathbf{j}_{rk}$, respectively, as representative joints. Thereafter, a skeleton sequence is represented as

$$\mathcal{G}(\mathsf{M}) = \{p_i : \forall_{i \in [1, \cdots, k]}, p_i = \|\mathsf{P}\mathbf{j}_{lk_i}\|_2 - \|\mathsf{P}\mathbf{j}_{rk_i}\|_2\}, \quad (3)$$

where P is defined as the projection matrix on the main direction of the motion variation. Although, the main direction can be estimated using conventional dimensionality reduction methods like Principal Component Analysis (PCA) [23], in this work we select the y-z plane as the main movement direction using a data normalization process. This ensures the joint's main variation to lie on y-z plane, see Subsection V-A. Hence, P is defined as

$$\mathsf{P} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \quad (4)$$

Subsequently, we use a deformation-based distance function, as defined in [18], [19], for measuring the difference between two given motion representations; a brief description of the considered distance function is given below.

Distance Function: Consider a curve defined by (3) and described by a set of q points that are sampled from $k \gg q$ set of points using φ defined as follows

$$\varphi : [0, k] \rightarrow [0, q]. \quad (5)$$

We denote such a curve as $\varphi \circ \mathcal{G}(\mathsf{M}) = (p_1, \cdots, p_q)$, where $p_i \in \mathbb{R}^2$. In [18], [19], such a curve is represented by a set of rigid-transformation matrices as $(g_1, \cdots, g_q)$, so that $\varphi \circ \mathcal{G}(\mathsf{M}) = (g_1 p_1, \cdots, \prod_{i=1}^{q} g_i p_1)$. Henceforth, we denote rigid-matrix based representation of a curve $\varphi \circ \mathcal{G}(\mathsf{M}_a)$ as $\mathcal{C}_{a,\varphi} = (g_1, \cdots, g_q)$.

Thereafter, following [18], [19], the distance between two curve representations $\mathcal{C}_{a,\varphi}$ and $\mathcal{C}_{b,\varphi}$ is defined as

$$\mathfrak{D}(\mathcal{C}_{a,\varphi}, \mathcal{C}_{b,\varphi}) = \sqrt{\sum_{i=1}^{q-1} d(g_a^i, g_b^i)^2}, \quad (6)$$

where $d(\cdot, \cdot)$ is the geodesic distance given by

$$d(g_a, g_b) = \sqrt{\|\log(\mathsf{R}_a^\mathsf{T}, \mathsf{R}_b)\|_F^2 + \|\mathbf{v}_b - \mathbf{v}_a\|_F^2}, \quad (7)$$

where T denotes the matrix transpose, $\| \cdot \|_F$ denotes the Frobenius norm, and

$$g_i = \begin{pmatrix} \mathsf{R}_i & \mathbf{v}_i \\ 0 & 1 \end{pmatrix}, \quad (8)$$

such that R_i and $\mathbf{v}_i$ are rotation matrices and translation vectors in $\mathbb{R}^3$, respectively. Hence, the distance between two curves $\mathcal{G}(\mathsf{M}_a)$ and $\mathcal{G}(\mathsf{M}_b)$ that are represented by $\mathcal{C}_{a,\varphi}$ and $\mathcal{C}_{b,\varphi}$, respectively, can be computed using (6), see [18], [19] for further details.

Normal Motion Model: We model a normal motion sequence using the sample mean of the training dataset. Consequently, we compute the sample mean of a curve representation dataset $\{\mathcal{C}_{1,\varphi}, \cdots, \mathcal{C}_{n,\varphi}\}$, following [18], [19], as

$$\mathcal{C}_{t,\varphi} = \underset{\mathcal{C}_{e,\varphi}}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^{n} \mathfrak{D}(\mathcal{C}_{e,\varphi}, \mathcal{C}_{i,\varphi})^2, \quad (9)$$

where $\mathcal{C}_{t,\varphi}$ represents the estimated mean of the dataset and $\mathfrak{D}(\cdot, \cdot)$ is as defined in (6). Finally, the binary decision of classifying a motion as either normal or abnormal is made by assuming a symmetric distribution of the training data points and fixing a δ-radius range, as defined in (2).

B. Time-Variation in Motion Analysis

Although motion variation due to speed or latency is considered as irrelevant information for tasks like action recognition, it is one of the distinctive characteristics studied in problems like motion quality assessment, since the focus is on performance variability under a predefined action. Consequently, to compare two given sequences, time-variation needs to be taken into account in both action recognition like problems and motion analysis.

In action recognition, Dynamic Time Warping (DTW) [24] is one of the widely used techniques to filter time-variation between two given motion sequences with respect to one another [25]. In principle, DTW selects frames from both sequences such that a given cost function is minimized, which usually is the L_2-norm. Meanwhile, in [19] an objective function that is similar to DTW, yet flexible, is proposed and used to align curves based on deformation cost. Nevertheless, our goal is not to filter time-variation in motion but to use it for motion quality analysis. Hence, in this subsection, we detail an adaptation of the objective function introduced in [19] to define a distance function in the context of motion analysis.

We begin by defining the deformation-based curve alignment function introduced in [19]. Let $\bar{\varphi}$ be a function that samples uniformly spaced points as defined in (5), then a curve $\mathcal{C}_i$ is aligned to a fixed curve $\mathcal{C}_{t,\bar{\varphi}}$ by solving for

$$\varphi^* = \underset{\varphi}{\operatorname{argmin}} \ \mathfrak{D}(\mathcal{C}_{t,\bar{\varphi}}, \mathcal{C}_{i,\varphi}), \quad (10)$$

where $\mathfrak{D}(\cdot, \cdot)$ is as defined in (6).

The solution for (10), φ^*, returns matching points which we will refer to as key-points. In other words, the distance between the two curves is minimized if $\mathcal{C}_i$ is sampled according to φ^*. Hence, the aligned curves are defined as $\mathcal{C}_{t,\bar{\varphi}}$ and $\mathcal{C}_{i,\varphi^*}$. However, we are not interested on filtering time-variation from the motion arguments– i.e., minimizing the distance between two curves based on the key-points φ^*. In fact, we aim to analyze the time-variation in between key-points; in a sense, using the solution φ^* for the opposite effect than it was intended. Figure 2 shows an example of the distinction between the key-points and the time-variation between key-points (magnified part in Fig. 2).

As a result, given q key-points that are identified by φ^*, we summarize the motion between two sequential key-points ℓ, j of the curve $\mathcal{C}_i$ as
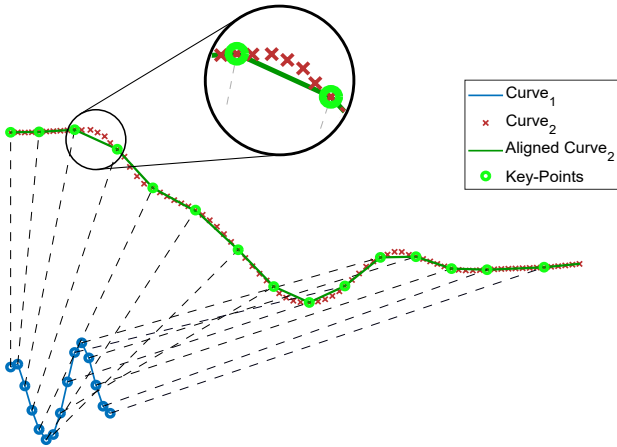
Fig. 2: Deformation-based alignment between two curves. The blue curve is described as a set of uniformly sampled points $\bar{\varphi}$. The red curve represents the original curve to be aligned. The green line shows the linearization using the key-points φ^*. The magnified part in the figure highlights the distinction between the key-points and the time-variation between key-points.

$$\mathcal{F}_j(\mathcal{C}_{i,\varphi^*_{(j,\ell)}}) = \prod_{r=j}^{\ell} g_r, \tag{11}$$

where $j < \ell$. In contrast, the approach in [19] linearizes the deformation between the key-points (refer to Fig. 2) while (11) attempts to preserve the observed deformation due to time-variation. Consequently, (6) is redefined to reflect the time-variation between two curves as follows

$$\mathcal{D}(\mathcal{C}_{t,\bar{\varphi}}, \mathcal{C}_{i,\varphi^*}) = \sqrt{\sum_{j=1}^{q-1} d(\mathcal{F}_j^t, \mathcal{F}_j^i)^2}, \tag{12}$$

where $d(\cdot,\cdot)$ is as defined in (7). Henceforth, we will refer to (12) as motion-quality distance. For a better understanding of the proposed approach, Fig. 3 shows an overview of the pipeline.

V. EXPERIMENTS

In this section, we present the dataset used to evaluate the proposed approach. Furthermore, we show how the skeleton data is normalized to address variations due to scale and absolute locations in Subsection V-A. In Subsection V-B we present an analysis of the time-variation of the skeleton sequences, and finally, in Subsection V-C we show quantitative results of the abnormal motion detection system.

To evaluate the proposed approach for detecting abnormal motion using a curve-based representation from skeleton data, we use the publicly available dataset SPHERE-Staircase2014 introduced in [13]. This dataset includes 48 sequences performed by 12 subjects while walking upstairs. The sequences were captured using an RGB-D sensor placed at the top of the stairs. In the dataset, the abnormal motion events were performed by subjects under the guidance of a physiotherapist. Such scenario is very interesting to test our approach for the reason that we are interested in home-based rehabilitation. We follow the same protocol as [13] and we use 17 sequences to build a normal motion model. The rest of the sequences were used in the testing part, where 14 sequences are considered as normal motions and 17 contain abnormalities.

A. Data Normalization

In order to make the skeleton information invariant to absolute location and scale, each skeleton S is spatially registered

without modifying the joint angles. First, the world coordinate system is placed at the joint corresponding to the hip center. Then, the skeleton is rotated such that the projection of vector from the left hip to the right hip onto the x-y plane is parallel to the x-axis [25]–[27]. Furthermore, each skeleton is normalized such that the body-parts lengths match the corresponding lengths of the reference skeleton [25]–[27]. Note that, a body-part is a rigid segment that connects at least two human body joints. With such normalization, the main direction of the motion variation happens to be on the y-z plane. Consequently, the direction of the x-axis is perpendicular to the main movement direction, which has an irrelevant impact on the walking pattern. Figure 4 shows an example of a normalized skeleton with the corresponding coordinate system and also the plane concerning the main direction of the motion. Before any other step, we apply an averaging window filter with size 11 on the joint coordinates in order to reduce the amount of noise from the skeleton data.

B. Time Variation Analysis

Generally, when comparing two temporal sequences, temporal alignment techniques are used in order to find the points that minimize a given cost function. With this in mind, we performed an analysis on the detection of the key-points, which are the points that minimize the deformation-based alignment between two curves (refer to Fig. 2) as defined in (10). With the purpose of analyzing the impact of the time-variation between key-points, we selected a uniform sampling function $\bar{\varphi}$ for the model $\mathcal{C}_{t,\bar{\varphi}}$. Then, we tested the following scenarios for the testing curves:

1) Uniform Sampling, US: using uniformly sampled points to compute the distance (6);
2) Optimal Sampling, OS: using the key-points estimated from (10) to compute the distance (6);
3) Motion-Quality distance, MQ: using the key-points estimated from (10) to compute the motion-quality distance (12).

In order to have a clear understanding, we applied a multidimensional scaling (MDS) method to the data based on the corresponding cost function for each scenario. This was done to visualize the relationship between the distance functions and the ground truth labels. Figure 5 shows the visualization of the MDS, where Fig. 5a refers to the US scenario, Fig. 5b to the OS scenario, and Fig. 5c to the proposed MQ scenario.

The conclusion that we obtain by looking at the Fig. 5 is that for the case of the OS scenario (see Fig. 5b), the information about time-variation is filtered. This means that while computing the distance between two curves that are sampled according to the key-points, the distance will always be the minimum possible. Consequently, the distinction between normal and abnormal motion is not that discriminative. Figure 5b shows that it is not possible to separate the normal from the abnormal motion for such scenario. Hence, for our purpose, the classification using the linearization of the key-points would not be possible at all. Such approaches could be applied to key-pose skeleton based detection applications, e.g., for action detection/recognition applications [28], [29]. On the contrary, for the US scenario (refer to Fig. 5a), the information about time-variation is lost. Meaning that the sampling is happening independently from the model, the sampling rate is chosen a priori. In this case and due to the small amount of variation in the data, the classification would be considerably good. We believe that with the increase of the variation in the data, this
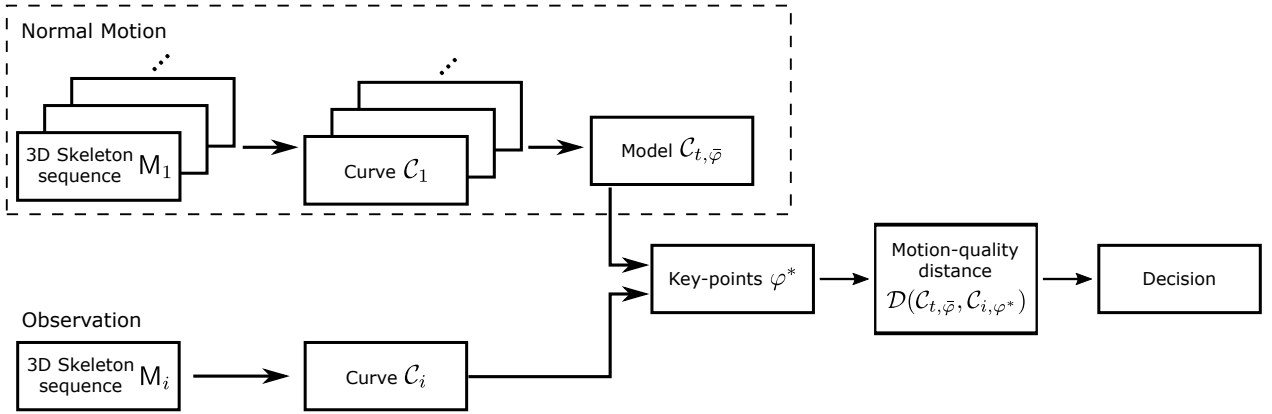
Fig. 3: Overview of the proposed approach. In the upper part, the dashed line rectangle regards the process of achieving the normal motion model. The bottom part concerns the testing scenario, where an input 3D skeleton sequence is represented as a curve. Then, the key-points φ^* are obtained by employing a deformation-based alignment between the model and the observed curve. Consequently, the motion-quality distance is computed in order to decide if whether normal or abnormal motion.
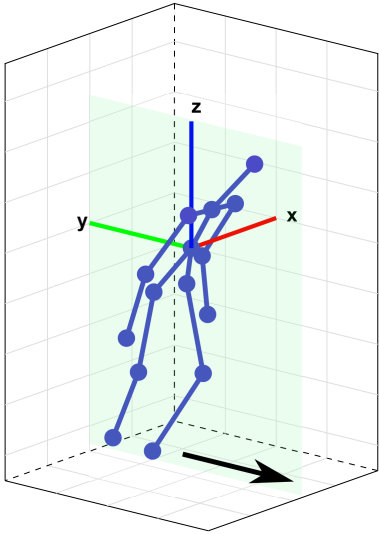

Fig. 4: Skeleton normalization. The word coordinate system is placed at the hip center. The plane colored in green represents the y-z plane where the main direction of motion variation is happening. The arrow shows the main direction of the movement.

scenario would not perform such satisfactory results. Figure 5a depicts that the normal and the abnormal motion are separable from one another. In this work, we are interested in analyzing the time-variation between key-points. Unlike [19], we do not linearize the deformation between the key-points. Instead, we preserve the observed deformation due to time-variation. With such approach, preserving the time-variation allow us to estimate the deviation of an observation from the model. Consequently, using the motion-quality distance, that emphasizes the time-variation between key-points, the classification of normal or abnormal motion produces encouraging results. Figure 5c shows that the data have similar behavior comparing to the OS scenario, but in this case, it is possible to separate the normal from the abnormal motion.

C. Abnormal Motion Detection

To demonstrate the performance of the proposed abnormal motion detection system, we compare our results with the work of [13]. In their work, they evaluate the detection per abnormal motion, while in our case we are evaluating per sequence. With this, we assume that a sequence containing at least one abnormal motion is considered as an abnormal sequence. For the classification results, we evaluated the proposed scenarios

described in the previous subsection. For the normal motion model we used 30 as the number of uniformly sampled points. For the classification of normal or abnormal motion detection, we used an empirically tested threshold $\delta = 30.65$. Table 1 shows the results for the abnormal motion detection using the SPHERE-Staircase2014 dataset. Note that, as mentioned before, the results are in correspondence with the previously described scenarios (refer to Fig. 5). In the OS scenario, the system could not identify any kind of abnormal motion, detecting only the normal motion. Subsequently, for the other scenarios, we achieved promising results, wherein the case that we consider the motion-quality distance (MQ scenario), the results are better than the US scenario.

Table 1: Results for the abnormal motion detection using the SPHERE-Staircase2014 dataset.

Methods	Accuracy (%)
Paiement *et al.* [13]	93
Uniform Sampling, US	91
Optimal Sampling, OS	50
Motion-Quality distance, MQ	**97**

VI. CONCLUSION

In this paper, we presented an approach for an abnormal motion detection system. For that purpose, we proposed to represent a 3D skeleton sequence as a curve to reduce its dimensionality, yet representative. Hence, based on this curve representation, we defined a motion-quality distance, which emphasizes time-variation between key-points. The proposed method showed that highlighting time-variation, we were able to handle small variations in the data due to the nature of the dataset (no variation in the action class, but only in the manner that an action is performed). We also presented an analysis on the time-variation for abnormal motion detection. This was done to present qualitatively the performance of the motion-quality distance, which is an important property of the quality assessment.

As future work, we intend to explore in more detail the properties of the deformation-based alignment in motion analysis. For that purpose, a penalty term would be added in the minimization problem, such that it would incorporate problem-related constraints, *e.g.*, 3D skeleton-specific geometry or noise-free skeleton dynamics [30]. We also intend to expand the proposed approach to represent a skeleton sequence as a point in the deformation space, without any prior knowledge. Such

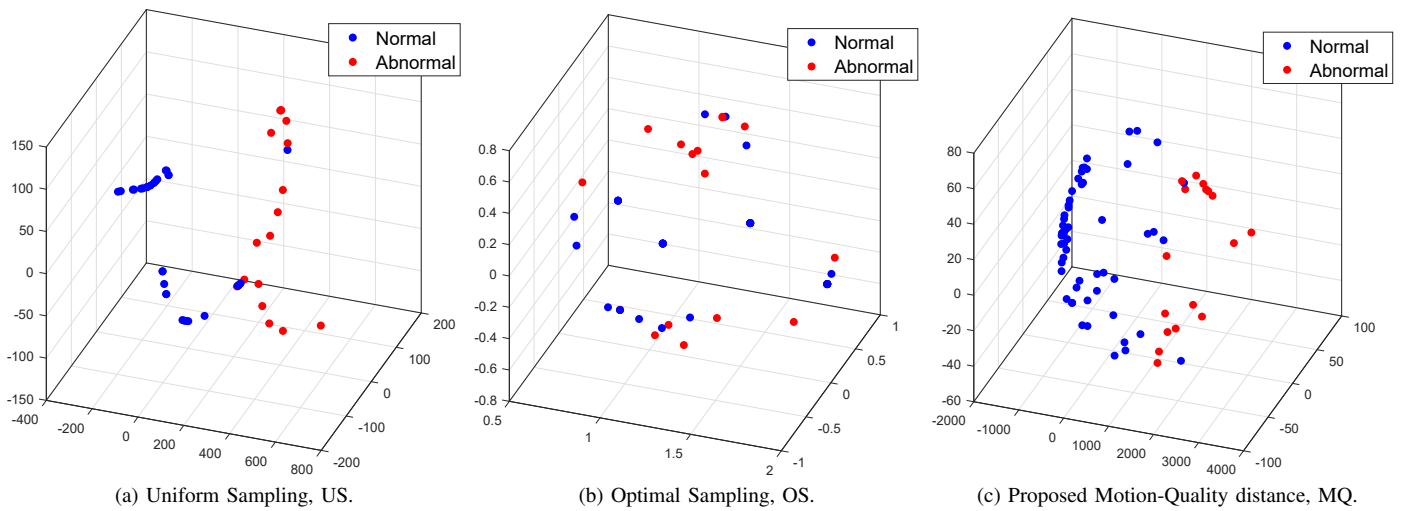| (a) Uniform Sampling, US. | (b) Optimal Sampling, OS. | (c) Proposed Motion-Quality distance, MQ. |

Fig. 5: Multidimensional scale method applied to the data on the representation space. Each point represents a curve. Figure 5a shows the case of the US scenario, Fig. 5b for the OS and Fig. 5c using the proposed MQ scenario.

approach could be used in action detection/recognition applications. Finally, we plan to test the approach with data acquired from real post-stroke patients.

ACKNOWLEDGMENT

REFERENCES

[1] J.K. Aggarwal and M.S. Ryoo, "Human activity analysis: A review," *ACM Comput. Surv.*, vol. 43, no. 3, pp. 16:1–16:43, Apr. 2011.

[2] O. P. Popoola and K. Wang, "Video-based abnormal human behavior recognition - a review," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 42, no. 6, pp. 865–878, Nov 2012.

[3] R. Baptista, M. Antunes, D. Aouada, and B. Ottersten, "Anticipating suspicious actions using a small dataset of action templates," in *13th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISAPP)*, 2018.

[4] R. Poppe, "A survey on vision-based human action recognition," *Image and Vision Computing*, vol. 28, no. 6, pp. 976 – 990, 2010.

[5] K. Papadopoulos, M. Antunes, D. Aouada, and B. Ottersten, "Enhanced trajectory-based action recognition using human pose," in *Image Processing (ICIP), 2017 IEEE International Conference on*. IEEE, 2017, pp. 1807–1811.

[6] K. Papadopoulos, M. Antunes, D. Aouada, and B. Ottersten, "A revisit of action detection using improved trajectories," in *IEEE International Conference on Acoustics, Speech and Signal Processing, Calgary, Alberta, Canada, 15–20 April 2018*, 2018.

[7] H. Pirsiavash, C. Vondrick, and A. Torralba, "Assessing the quality of actions," in *Computer Vision–ECCV 2014*, pp. 556–571. Springer, 2014.

[8] R. Wang, G. Medioni, C. Winstein, and C. Blanco, "Home monitoring musculo-skeletal disorders with a single 3d sensor," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2013, pp. 521–528.

[9] R. Baptista, E. Ghorbel, A. E. R. Shabayek, D. Aouada, and B. Ottersten, "Key-skeleton based feedback tool for assisting physical activity," in *2018 Zooming Innovation in Consumer Technologies Conference (ZINC)*. IEEE, 2018, pp. 175–176.

[10] G. Kwakkel, B. J. Kollen, and H. I. Krebs, "Effects of robot-assisted therapy on upper limb recovery after stroke: a systematic review," *Neurorehabilitation and neural repair*, 2007.

[11] P. Langhorne, G. Taylor, G. Murray, M. Dennis, C. Anderson, E. Bautz-Holter, P. Dey, B. Indredavik, N. Mayo, M. Power, et al., "Early supported discharge services for stroke patients: a meta-analysis of individual patients' data," *The Lancet*, 2005.

[12] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Comput. Surv.*, vol. 41, no. 3, pp. 15:1–15:58, July 2009.

[13] A. Paiement, L. Tao, M. Camplani, S. Hannuna, D. Damen, and M. Mirmehdi, "Online quality assessment of human motion from skeleton data," in *Proceedings of the British Machine Vision Conference*. 2014, BMVA Press.

[14] L. Tao, A. Paiement, D. Damen, M. Mirmehdi, S. Hannuna, M. Camplani, T. Burghardt, and I. Craddock, "A comparative study of pose representation and dynamics modelling for online motion quality assessment," *Computer Vision and Image Understanding*, 2016.

[15] G. S. Parra-Dominguez, B. Taati, and A. Mihailidis, "3d human motion analysis to detect abnormal events on stairs," in *2012 Second International Conference on 3D Imaging, Modeling, Processing, Visualization Transmission*, Oct 2012, pp. 97–103.

[16] F. Nater, H. Grabner, and L. Van Gool, "Exploiting simple hierarchies for unsupervised human behavior analysis," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, June 2010, pp. 2014–2021.

[17] J. Snoek, J. Hoey, L. Stewart, R. S. Zemel, and A. Mihailidis, "Automated detection of unusual events on stairs," *Image and Vision Computing*, vol. 27, no. 1-2, pp. 153–166, 2009.

[18] G. G. Demisse, D. Aouada, and B. Ottersten, "Similarity metric for curved shapes in euclidean space," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

[19] G. G. Demisse, D. Aouada, and B. Ottersten, "Deformation based curved shape representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 6, pp. 1338–1351, June 2018.

[20] B. Schölkopf and A. J. Smola, "A short introduction to learning with kernels," in *Advanced lectures on machine learning*, pp. 41–64. Springer, 2003.

[21] J. Wang, Z. Liu, Y. Wu, and J. Yuan, "Mining actionlet ensemble for action recognition with depth cameras," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, June 2012, pp. 1290–1297.

[22] M. Antunes, D. Aouada, and B. Ottersten, "A revisit to human action recognition from depth sequences: Guided svm-sampling for joint selection," in *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, March 2016, pp. 1–8.

[23] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*, Springer-Verlag, Berlin, Heidelberg, 2006.

[24] Meinard Müller, "Dynamic time warping," *Information retrieval for music and motion*, pp. 69–84, 2007.

[25] R. Vemulapalli, F. Arrate, and R. Chellappa, "Human action recognition by representing 3d skeletons as points in a lie group," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014.

[26] M. Antunes, R. Baptista, G. Demisse, D. Aouada, and B. Ottersten, "Visual and human-interpretable feedback for assisting physical activity," in *European Conference on Computer Vision (ECCV) Workshop on Assistive Computer Vision and Robotics Amsterdam,*, 2016.

[27] R. Baptista, M. Antunes, A. E. R. Shabayek, D. Aouada, and B. Ottersten, "Flexible feedback system for posture monitoring and correction," in *IEEE International Conference on Image Information Processing (ICIIP)*, 2017.

[28] L. Miranda, T. Vieira, D. Martínez, T. Lewiner, A. W. Vieira, and M. F. M. Campos, "Online gesture recognition from pose kernel learning and decision forests," *Pattern Recognition Letters*, vol. 39, pp. 65–73, 2014.

[29] F. Ofli, R. Chaudhry, G. Kurillo, R. Vidal, and R. Bajcsy, "Sequence of the most informative joints (smij): A new representation for human skeletal action recognition," in *2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, June 2012, pp. 8–13.

[30] G. G. Demisse, K. Papadopoulos, D. Aouada, and B. Ottersten, "Pose encoding for robust skeleton-based action recognition," *CVPRW: Visual Understanding of Humans in Crowd Scene, Salt Lake City, Utah, June 18-22, 2018*, 2018.