Motion Error Classification for Assisted Physical Therapy
A Novel Approach using Incremental Dynamic Time Warping and Normalised Hierarchical Skeleton Joint Data

Julia Richter, Christian Wiede, Bharat Shinde and Gangolf Hirtz
Department of Electrical Engineering and Information Technology,
Technische Universität Chemnitz, Reichenhainer Str. 70, 09126 Chemnitz, Germany
{*julia.richter, christian.wiede, bharat.shinde, g.hirtz*}*@etit.tu-chemnitz.de*

Keywords: Health Care, Medical Training Therapy, Support Vector Machines, Motion Sequence Matching, Incremental Dynamic Time Warping.

Abstract: Preventive and therapeutic measures can contribute to maintain or to regain physical abilities. In Germany, the growing number of elderly people is posing serious challenges for the therapeutic sector. Therefore, the objective that has been pursued in recent research is to assist patients during their medical training by reproducing therapists' feedback. Extant systems have been limited to feedback that is based on the evaluation of only the similarity between a pre-recorded reference and the currently performed motion. To date, very little is known about feedback generation that exceeds such similarity evaluations. Moreover, current systems require a personalised, pre-recorded reference for each patient in order to compare the reference against the motion performed during the exercise and to generate feedback. The aim of this study is to develop and evaluate an error classification algorithm for therapy exercises using Incremental Dynamic Time Warping and 3-D skeleton joint information. Furthermore, a normalisation method that allows the utilisation of non-personalised references has been investigated. In our experiments, we were able to successfully identify errors, even for non-personalised reference data, by using normalised hierarchical coordinates.

1 INTRODUCTION

With regard to our steadily ageing population, the maintenance of each individual's health should be a priority for our society. Beside preventive measures, rehabilitation with the aim of recovering physical capabilities especially after a surgery is an important aspect. Thus, elderly with a total hip endoprothesis, for example, require an appropriate rehabilitation in order to recover from the surgery and to restore their mobility.

In contrast to this aim, rehabilitation facilities are facing challenges in maintaining a high quality during the medical training therapy because of a lack of therapists. Recent evidence suggest that the number of therapists is insufficient to give adequate feedback to patients (Richter et al., 2016). Detailed feedback, however, is necessary to ensure a correct exercise execution. Consequently, there remains an urgent need to solve this problem. Recent research has been carried out to support the rehabilitation process with technical assistance systems. To date, little is known about the recognition of typical errors that occur during certain exercises. This study therefore seeks to investigate the recognition of errors that are characteristic for the exercise hip abduction. Moreover, we investigated the effect of using non-personalised reference and training data on the error recognition accuracy and present a solution to avoid a deterioration of accuracy. This solution encompasses the utilisation of normalised hierarchical joint coordinates.

The paper is organised as follows: Section 2 gives an overview about related work and evinces the research gap. Section 3 describes the methods that were developed to identify patients' motions errors. Moreover, in Section 4, the evaluation methodology is introduced. The results are presented and discussed in Section 5, whereas the paper is concluded in Section 6. Moreover, we give an outlook to future work.

2 RELATED WORK

Capturing and assessing motions in real-time is an emergent field of research. Yurtman et al. distributed wearable motion sensors, i. e. accelerometers, gyro-

Richter, J., Wiede, C., Shinde, B. and Hirtz, G.
Motion Error Classification for Assisted Physical Therapy - A Novel Approach using Incremental Dynamic Time Warping and Normalised Hierarchical Skeleton Joint Data.
DOI: 10.5220/0006108002810288
In *Proceedings of the 6th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2017)*, pages 281-288
ISBN: 978-989-758-222-6

scopes and magnetometers, over the human body in order to evaluate a therapy exercise as correct or incorrect (Yurtman and Barshan, 2013). They detected occurrences of template signals in the sequence acquired during the therapy session using Dynamic Time Warping (DTW). Tormene et al. matched data with the help of DTW for post-stroke rehabilitation (Tormene et al., 2009). The data was obtained from strain sensors worn into a long-sleeve shirt. Tak et al. presented an approach for human abnormality detection in video sequences (Tak et al., 2011). They calculated distance curves from the detected foreground regions in each image and performed a Fourier transform on every distance curve. After that, they deduced motion similarities by comparing a predefined human motion sequence with the motion that was performed by the user in real-time. For this step, they applied DTW and Dynamic Group Warping.

Especially since the launch of the Kinect version 1.0 in 2010, researchers investigated assessment methods using the skeleton that is detected in depth images using the algorithm of Shotton et al. (Shotton et al., 2013). Recent work focused on the evaluation of motions during sports, such as dancing, ballet training and Tai Chi exercises, using the Kinect skeleton. Since our application is similar to this type of training, relevant principles are briefly described in the following. Huang et al. applied DTW to compare a pre-recorded reference sequence from a teacher against the dancing motion performed by a student (Huang et al., 2013). Instead of using the Euclidean distance as a distance metric for the DTW cost matrix calculation, they determined angles derived from body motion vectors. A warping path aligned the two sequences, so that the accumulated distance was minimal. A final score was calculated from the accumulated distances along the warping path. In order to assess the rhythm, the magnitudes of body motion vectors were transformed to the frequency domain. Thus, the feedback given by this system provided the grade of similarity in comparison with the reference motion. However, this feedback did not contain information about specific motion errors. Another approach that aimed at generating feedback for ballet training analyses motion trajectories with spherical self-organizing maps (Muneesawang et al., 2015). The system provided feedback in a virtual environment, i. e. by displaying information on the screens in a 3-D Cave Automatic Virtual Environment (CAVE). The work of Lin et al. focused on assessing Tai Chi exercises (Lin et al., 2013). They calculated the mean Euclidean distance and the mean angle difference between joints of a reference and the current exercise. The comparison with thresholds yielded a final score indicating the grade of similarity. In the field of physical rehabilitation, similar approaches can be found. Several studies aimed at comparing a reference recording of a therapy exercise with an exercise currently performed by a patient. Su et al. captured personalised trajectories from joints of interest and calculated the minimal accumulated Euclidean distance using DTW (Su et al., 2014). By means of fuzzy logic, they generated a textural output from the DTW joint values that informed the patient after the exercise whether the performance was *bad*, *good* or *excellent*. Moreover, they stated that future work should investigate normalisation methods that allow a patient to use the system without personalised pre-recordings. Since it is often helpful for the patient to receive the feedback directly during the exercise, Khan et al. introduced a method for continuous evaluation, which is called Incremental Dynamic Time Warping (IDTW), (Khan et al., 2014). This extension of DTW enables the comparison between the complete reference exercise and an incomplete exercise that the patient currently performs. After every new frame, the IDTW algorithm determines the reference segment that matches best with the currently performed part of the exercise. The DTW value calculated from both partial sequences then represented the similarity between the reference and the current execution. This method, however, is not able do determine specific errors that occur during the exercise performance.

Overall, these studies presented principles and solutions to assess the similarity between a pre-recorded reference and a currently performed motion. Although these studies showed that such kind of feedback is highly beneficial for the user, no study exists so far that recognises and communicates specific errors. The purpose of this investigation is to explore approaches that allow the recognition of errors that typically occur during the performance of hip abduction exercises. In this context, we intent to investigate a normalisation method in order to avoid taking prerecordings and examples of incorrect motions for new patients. Therefore, the performances of two types of coordinate representation, i. e. local and normalised hierarchical coordinates, were compared against each other by using three different scenarios. Since we intend to give continuous feedback to the patient, we employed the IDTW algorithm described by Khan et al. (Khan et al., 2014).

3 METHODS

This section describes the applied and implemented methods for motion error identification. First of all,

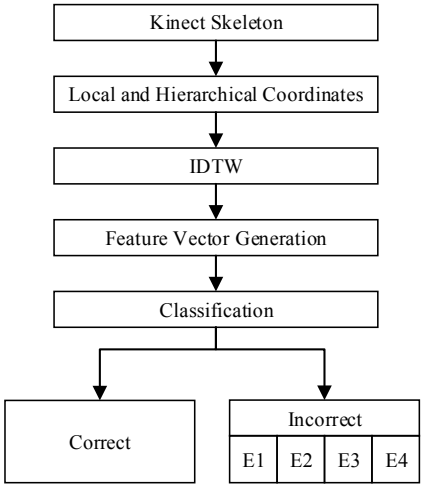

Figure 1: Procedure overview.

the input skeletons were pre-processed, which includes local and hierarchical coordinate transformation. Secondly, IDTW was used to determine the corresponding reference frame for every new incoming frame for certain joints. The coordinates from this determined reference frame and the current frame were thereupon used for feature vector generation and finally for error identification. At this point, a hierarchical SVM was used to classify the motion in the current frame as correct or incorrect in a first level. In a second hierarchical level, the classifier identified which errors the incorrect motion encompassed. Figure 1 gives an overview about this procedure.

3.1 Kinect Skeleton

The input data for the following algorithms were Kinect skeletons. Therefore, we briefly present the Kinect skeleton data structure in this section.

A coordinate of a joint in camera coordinates is given by a 1×3 vector of the form

$$joint_{\mathrm{i}} = \begin{bmatrix} x_{\mathrm{cam,i}} & y_{\mathrm{cam,i}} & z_{\mathrm{cam,i}} \end{bmatrix}, \quad i \in \{1,...,20\} \tag{1}$$

whereas $x_{\mathrm{cam,i}}$, $y_{\mathrm{cam,i}}$ and $z_{\mathrm{cam,i}}$ are the x, y and z components of the joints in the 3-D camera coordinate system S_{cam}. Figure 2 illustrates the twenty obtained joints. The joint indices i are defined in this figure.

For the exercise hip abduction, we defined joints of interest, which are relevant for the error identification. These relevant joints are those with indices $i = \{3,4,17,18,19,20\}$. In the following, all processing steps were performed only for these relevant joints.

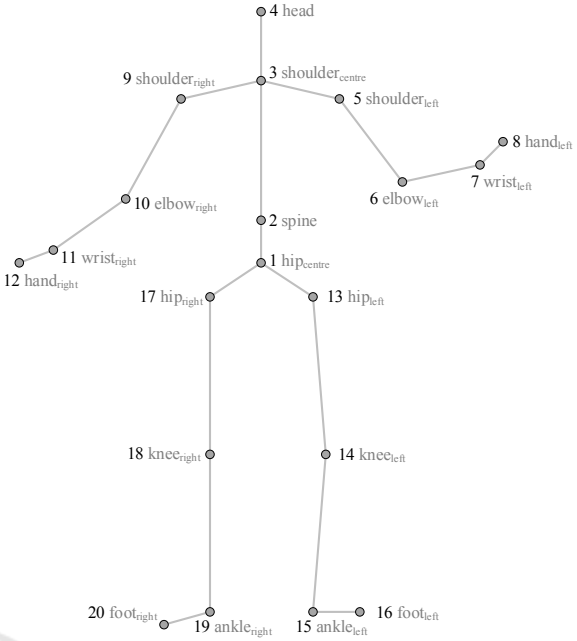

Figure 2: Twenty skeleton joints provided by Kinect version 1.0 with the corresponding indices i (frontal view).

3.2 Local and Normalised Hierarchical Coordinates

The pre-processing step firstly comprised the calculation of a rotation matrix R, so that the coordinates became invariant with respect to the rotation of the Kinect, and secondly the transformation of all joints to local or normalised hierarchical coordinates respectively.

Rotation Matrix. A coordinate transformation was applied to make the data translation and rotation invariant in case of different sensor orientations while capturing the data. The rotation matrix R is calculated using the first frame of a repetition. We thereby assumed that the person is standing straight at the start of an exercise.

The x-axis x of the new coordinate system is defined as the unit vector between the right and the left hip joint, see Equation 2. Then, h is defined as the vector between the shoulder center and the left hip, see Equation 3. The new z-axis z is the cross product of h and the new x-axis x, as formulated in Equation 4. Finally, according to Equation 5, the new y-axis y is the cross product of the obtained z-axis and the x-axis.

$$x = \frac{joint_{17} - joint_{13}}{\|joint_{17} - joint_{13}\|} \tag{2}$$

$$h = joint_{13} - joint_{3} \tag{3}$$

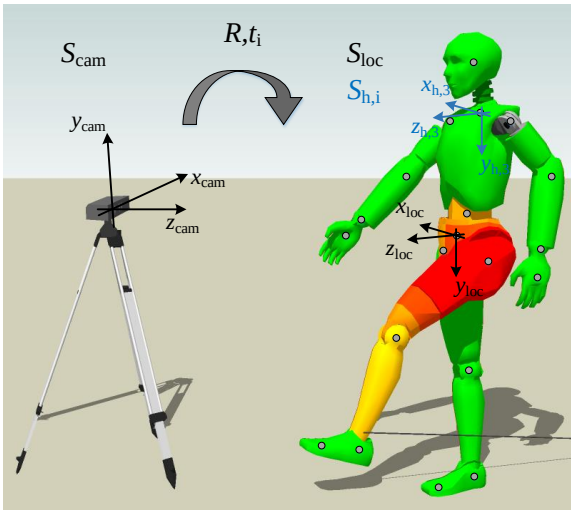

Figure 3: Coordinate transformation from camera coordinate system S_{cam} to local coordinate system S_{loc} (black) and hierarchical coordinate systems $S_{h,i}$ (blue) with corresponding parent joints. Only $S_{h,3}$ is illustrated as an example for the hierarchical representation.

$$z = \frac{x \times h}{\|x \times h\|} \tag{4}$$

$$y = \frac{z \times x}{\|z \times x\|} \tag{5}$$

The resulting rotation matrix is denoted as

$$R = \begin{bmatrix} x^\intercal & y^\intercal & z^\intercal \end{bmatrix}. \tag{6}$$

This rotation matrix was used for the calculation of both local and normalised hierarchical coordinates, which is described in the following.

Local Coordinates. Joints $joint_{loc,i}$ in local coordinate representation were obtained from the joints $joint_i$ in camera coordinate representation according to

$$joint_{loc,i} = R \cdot \left(joint_i^\intercal - joint_1^\intercal \right). \tag{7}$$

Consequently, the coordinates of each local joint are given with respect to the hip centre $joint_1$. The local camera system is denoted as S_{loc} and illustrated in Figure 3. Compared to hierarchical coordinates, the disadvantage of local coordinates is that every joint's position is a sum of all the joints vectors it is connected to. This results in a summation of deviations from the reference. In contrast to that, hierarchical coordinates allow to observe the deviation of every single joint from the reference independent from other joint vectors.

Normalised Hierarchical Coordinates. In the same way as in the study from Khan et al. (Khan et al., 2014), we transformed the joints, which are originally represented in the camera coordinate system S_{cam}, to a hierarchical representation in coordinate systems $S_{h,i}$. Moreover, we normalised the limbs to make the algorithm invariant to different body sizes and proportions.

The hip centre is on the highest hierarchical level. The joints that are connected to the hip centre, specifically both hip joints and the spine, are on the second level and are called child joints of the hip. Vice versa, every child joint has a parent joint on the next higher hierarchical level to which it is connected to. The parent joint of the hip centre is defined as the hip centre itself. This hierarchy is constructed until the hands, the head and the feet are reached. Every child joint is finally represented in the coordinate system of its parent joint.

The single hierarchical coordinate systems $S_{h,i}$ have the orientation R as calculated in Equation 6. In contrast to the local coordinate system, they have their origin in the corresponding parent joints. In other words, the camera coordinate system is translated by the coordinate of every parent joint, which results in the translation vectors described in Equation 8. Thereby, $parent(joint_i)$ denotes the parent joint of every joint. Every joint is then represented in its corresponding single coordinate systems $S_{h,i}$.

$$t_i = parent(joint_i) \tag{8}$$

The hierarchical coordinate of each joint $joint_{h,i}$ is calculated according to Equation 9: Firstly, the child joint $joint_i$ is translated by t_i. Secondly, this translated joint is multiplied by the rotation matrix R.

$$joint_{h,i} = R \cdot \left(joint_i^\intercal - t_i^\intercal \right) \tag{9}$$

Finally, the lengths of the limbs are normalised, see Equation 10. Consequently, the obtained normalised hierarchical joints $joint_{hn,i}$ have now a length equal to one.

$$joint_{hn,i} = \frac{joint_{h,i}}{\|joint_{h,i}\|} \tag{10}$$

Figure 3 illustrates the transformation from camera coordinates to local and hierarchical coordinates respectively.

3.3 Incremental Dynamic Time Warping

In order to identify possible incorrect motions during a patient's exercise execution, we compared a prerecorded reference of one repetition with the patient's current motion execution. This reference defines the correct motion execution. In this study, the reference was captured from the patient, which we call personalised reference, as well as from a therapist, which we denote as non-personalised reference.

Since we aim at giving continuous feedback, the IDTW algorithm by Khan et al. (Khan et al., 2014) was employed to compare the reference against the patient's motion. In the following, the principle of the IDTW algorithm is described. For more details, we refer to the publication by Khan et al.

The IDTW algorithm aligns an incomplete patient's sequence with the complete reference sequence and thereby allows a frame-wise feedback generation. In our study, both the reference and the partial current recordings consisted of a sequence of local or normalised hierarchical 3-D coordinates of a specific joint. In principal, both reference and current sequence can contain any data format.

Khan et al. calculated the IDTW distances joint-wise. For every single joint, a cumulative cost matrix G is calculated, whereas the DTW values in G are the accumulated Euclidean differences between a reference and a current coordinate. The calculation of this matrix is illustrated in Figure 4. Every time a new frame is acquired and new joint coordinates are calculated, a new column is appended to G. The reference frame that fits best to the latest acquired current frame is determined by finding the minimal DTW value in this column. In this way, the algorithm selects the part of the reference that matches best with the current partial sequence.

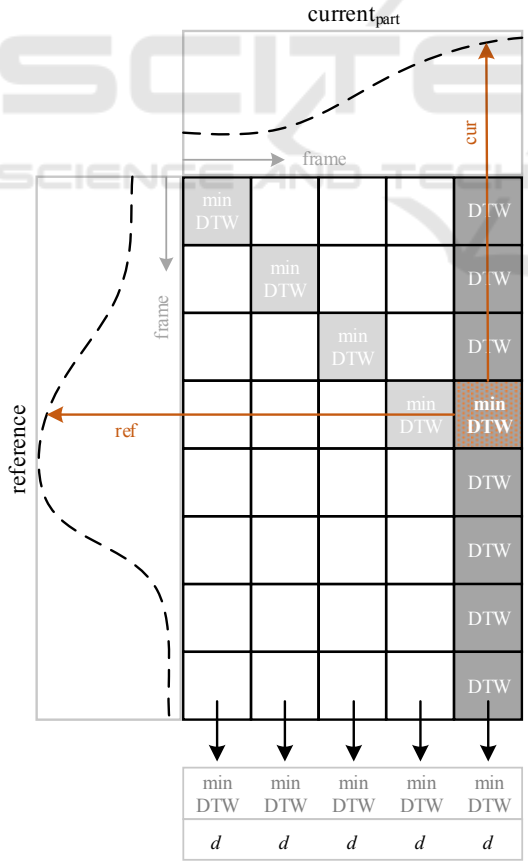

Figure 4: Cumulative cost matrix G for one single joint, which is updated frame-wise.

Khan et al. used the IDTW algorithm to give feedback by colour-coding the single joints according to their minimal DTW values for every incoming frame. In our approach, we extended the IDTW algorithm to determine the reference frame $joint_{\mathrm{i,ref}}$ that corresponds to the currently acquired frame $joint_{\mathrm{i,cur}}$. In this way, we were able to calculate a distance vector $d_{\mathrm{i,frame}}$ that describes the difference between the joint position in the current frame and the corresponding reference joint position. Consequently, instead of minimal DTW values, we obtained a feature vector for every new frame as an output of the IDTW algorithm, as illustrated in Figure 4.

In order to reset the cost matrix after each repetition, we implemented an algorithm to detect the end of a performed repetition. This algorithm is based on the distance between the left and the right ankle. The cost matrix was reset when the end of a repetition was detected, which means that all the elements were deleted. The feature vector calculation is described in the following section.

3.4 Feature Vector Generation

For the error classification, we calculated a feature for every interesting joint, which characterises the difference $d_{\mathrm{i,frame}}$ between the current joint position $joint_{\mathrm{i,cur}}$ and the corresponding reference joint position $joint_{\mathrm{i,ref}}$, see Equation 11.

$$d_{\mathrm{i,frame}} = joint_{\mathrm{i,cur}} - joint_{\mathrm{i,ref}}\,, \quad i \in \{3,4,17,18,19,20\} \tag{11}$$

The concatenation of the relevant joints $d_{\mathrm{i,frame}}$ represents the feature vector FV_{frame} for one frame:

$$FV_{\mathrm{frame}} = \begin{bmatrix} d_{3,\mathrm{frame}} & d_{4,\mathrm{frame}} & \dots & d_{20,\mathrm{frame}} \end{bmatrix} \tag{12}$$

3.5 Classification

In this study, we distinguished between the classes listed in Table 1. According to therapists, the described errors are frequently occurring when patients train without a therapist's feedback. A linear multi-class one-versus-all Support Vector Machine (SVM) was trained with feature vectors that were calculated from recordings of different persons. These persons performed several repetitions of the exercise hip abduction according to the descriptions in Table 1. Depending on the scenario, the training and testing configurations changed. This methodology will be explained in detail in Section 4.

Table 1: Classes with labels L and description for exercise hip abduction.

L	Class	Description
C	Correct	The exercise is performed in a correct way.
BK	Bent Knee	The abducted leg is not straight, but bent while performing the motion. The joints of the leg do not form a straight line.
FO	Foot Outside	The leg is rotated outwards, which results in a rotated position of the toe joint.
UB	Upper Body	The joints of the torso and head are moving although their position should not change.
WP	Wrong Plane	The joints of the abducted leg are not moving in the plane that is spanned by the joints of the supporting leg and the straight torso.

As a result, the whole classifier consists of five single one-versus-all SVMs, whereas the first SVM classifies whether the motion in the current frame is correct or incorrect (first hierarchy). The other SVMs (second hierarchy) decide whether the specific errors were detected if the first SVM predicts the motion as incorrect. If none of the five classifiers responds, we assumed the motion in the tested frame to be correct. In this way, we have designed a hierarchical SVM.

4 EVALUATION METHODOLOGY

In order to evaluate the capability of the system to work with non-personalised data, we distinguished three scenarios, which are summarised in Table 2. The scenarios correspond to the grade of personalisation: Scenario 1 is using only personalised data, while scenario 3 works only with non-personalised data. Consequently, scenario 3 is the most challenging one where we expected the lowest accuracy. The scenarios are presented and explained in detail in the following:

Scenario 1. For every patient, an individual machine was trained and tested with the individual patient's data and his or her own reference ref_{i}. For this purpose, every person's recordings were split in training $P_{\mathrm{i,train}}$ and testing set $P_{\mathrm{i,test}}$. Consequently, in practical applications, training data as well as the patient's reference have to be recorded for every new patient, which is rather impractical. For this scenario, the data of ten probands was used.

Table 2: Scenarios S1-S3 for evaluation.

Scenario	Train	Test
S1	$P_{\mathrm{i,train}}$, ref_{i}	$P_{\mathrm{i,test}}$, ref_{i}
S2	P_{train}, ref_{train}	$P_{\mathrm{i,test}}$, ref_{i}
S3	P_{train}, ref	$P_{\mathrm{i,test}}$, ref

Scenario 2. In contrast to scenario 1, one single machine was trained with the motion data of three persons P_{train} and their individual reference data ref_{train}. These were other persons than the probands in scenario 1. Just as in scenario 1, the test was performed with each test person's individual reference ref_{i} while the data from the ten probands $P_{\mathrm{i,test}}$ mentioned in scenario 1 was used. In practice, only the new patient's reference, but not his or her training data, has to be recorded. Obviously, this would be more practicable than scenario 1.

Scenario 3. The machine was trained with the motion data of the three persons P_{train} mentioned in scenario 2. In contrast to scenario 1 and scenario 2, the reference ref has not been changed according to the person. For testing, the very same reference ref was used for all test samples of the ten persons $P_{\mathrm{i,test}}$. This means that, in practice, neither training data nor a reference has to be recorded for a new patient. This would considerably reduce the effort for both patients and therapists.

When we consider these scenarios, we can conclude that practicability increases from scenario 1 to scenario 3, because the need of taking recordings before the training decreases: scenario 1 needs recordings from the patient's correct motions as well as from the incorrect motions and the patient's reference before a training session can start. In contrast to that, scenario 3 requires neither of these recordings, because it works with pre-recorded data from other persons. The question we want to answer is whether the results of scenario 3, which is the desired implementation, are still equivalent to those of scenario 1 and scenario 2. We moreover investigated whether the results improve for these scenarios if we use normalised hierarchical coordinates instead of non-normalised local coordinates.

To compare the performance when using local or hierarchical coordinates, we determine the overall accuracy η_{all} for all classes N for every scenario according to Equation 13:

$$\eta_{all} = \frac{1}{N} \cdot \sum_{n=1}^{N} \eta_n , \tag{13}$$

whereas η_n is the accuracy for every single one-

versus-all classifier, see Equation 14.

$$\eta_n = (\mathrm{TPR}_n + \mathrm{TNR}_n) \cdot 0.5 \tag{14}$$

These accuracies η_n are calculated for every single one-versus-all classifier to evaluate the detection of single error types. TPR_n and TNR_n are the true positive and the true negative rates for each class.

5 RESULTS AND DISCUSSION

The overall accuracies η_{all} for the three scenarios using the two different coordinate representations are presented in Table 3. It becomes obvious that, when using local, i. e. non-normalised coordinates, the overall accuracy considerably deteriorates from scenario 1 to scenario 3. This complies with our expectations: the less personalised data is used, the higher is the negative influence of factors, such as body size and proportions, on the accuracy. In contrast to local coordinates, the accuracy only slightly decreases when using normalised hierarchical coordinates. These results suggest that normalised hierarchical coordinates should be used instead of local coordinates in order to avoid the recording of personalised data for each new patient without loosing classification performance.

Based on this finding, we elaborated the accuracies η_n of the single classes for scenario 3 using normalised hierarchical coordinates, which is the most challenging scenario but the one with highest practical relevance as well. The results are summarised in Table 4. It is apparent from this table that 84 % of the samples that have been labelled as correct were also classified as correct. In 16 % of the tested frames, an error was detected although the test persons performed the exercise correctly. This is mainly because of the relatively high false positive rate (FPR) of the FO classifier, which results in a detected error in the second stage of the hierarchical SVM. The table also reveals that in 75 % of the frames that have been labelled as incorrect, an error could be identified. Taking the true positive rates (TPRs) of the error classes BK, UB and WP into consideration, we see that in at least 65 % of the frames where the specific error was present this very error could be identified. When we consider the error classes BK, UB and WP, the FPRs are low. This is of high importance in order not to confuse the patient by giving him feedback about errors that he or she actually has not performed. Only the error FO could not be successfully detected. The low TPR in FO is one reason for the low FPR in C, whereas the high FPR in FO is responsible for the high false negative rate (FNR) in C. For this reason, we performed a second test excluding the class FO. The improved results are presented in Table 5. The TPR of C increased to 90 % and the true negative rate (TNR) to 81 % when using normalised hierarchical coordinates in scenario 3.

At this point we would like to stress that the presented results in Table 4 represent a low-level classification, which is based on the single frames in an exercise sequence. Given the low FPRs (below 5 %) and the relatively high TPR (at least 65 %) we achieved for the classes BK, UB and WP, the error can be reliably detected. It is remarkable that the errors could be detected by just using the distance vector between another person's reference and the current patient's ex-

Table 3: Overall accuracies η_{all} for the three scenarios S1, S2 and S3 with local and normalised hierarchical coordinates.

Scenario	η_{all}, local	η_{all}, hierarchical
S1	0.83	0.86
S2	0.75	0.81
S3	0.67	0.82

Table 4: Confusion matrices and corresponding accuracies η_n for scenario 3 using normalised hierarchical coordinates. L: samples that were classified to be C, BK, FO, UB or WP. $\overline{L}$: samples that were classified to be $\overline{\mathrm{C}}$, $\overline{\mathrm{BK}}$, $\overline{\mathrm{FO}}$, $\overline{\mathrm{UB}}$ or $\overline{\mathrm{WP}}$.

	L	$\overline{L}$	η_n
C	0.84	0.16	0.82
$\overline{\mathrm{C}}$	0.25	0.75	
BK	0.81	0.19	0.90
$\overline{\mathrm{BK}}$	0.02	0.98	
FO	0.52	0.48	0.68
$\overline{\mathrm{FO}}$	0.17	0.83	
UB	0.87	0.13	0.91
$\overline{\mathrm{UB}}$	0.04	0.96	
WP	0.65	0.35	0.81
$\overline{\mathrm{WP}}$	0.03	0.97	

Table 5: Overall accuracies η_{all} for the three scenarios S1, S2 and S3 with local and normalised hierarchical coordinates without class FO.

Scenario	η_{all}, local	η_{all}, hierarchical
S1	0.90	0.90
S2	0.78	0.85
S3	0.72	0.87

ercise sequence. A potential error source is the matching of the sequences using IDTW. In case of inaccurate matching, both sequences could possibly not be correctly aligned. This results in an incorrect distance vector and hence in an insignificant feature. A general challenge we encountered is the noise in the obtained 3-D Kinect skeleton. This noise causes misclassifications, because the joints are inaccurately localised. Some of the persons, for example, wore short, wide trousers. The ends of the trousers were frequently recognised as the knee joints because of the working principle of the Kinect. Incorrect localisations especially affect the error FO since the foot joints are often error-prone.

6 CONCLUSIONS AND FUTURE WORK

In this study, we presented a method to detect incorrect motions during therapy exercises. In our experiments, we achieved results of high quality even if we used non-personalised data, i. e. neither the current person's reference nor the person's training data. The most obvious finding to emerge from this study is that by using hierarchical normalised joint representation, a unified model that was trained with other persons' data can be used for new patients. As this method allows the usage of pre-recorded data from other persons, a direct start for new patients without the need of recording their individual error motions and their reference is possible. Moreover, the study has confirmed that simple distance vectors are suitable feature vectors for error classification.

Nevertheless, future research should concentrate on the investigation of more distinctive features to further increase the accuracy and on a subsequent high-level processing using filtering techniques. For practical applications, the generation of synthetic training data would be sensible. In this way, the pre-recording of motion data could be avoided completely, which would simplify the extension with further exercises, such as hip extension and hip flexion. Another investigation will focus on the input for the IDTW. Currently, the trajectories of single joints were evaluated, but we did not contextualise them. Therefore, we plan to fuse all joint information into one single cost matrix. Another aspect for future work will be automatic joint filtering to find joints that are relevant for different exercises.

Taken together, this study created a base for further research that shows high potential for the required assistance in the therapy sector.

ACKNOWLEDGEMENTS

This project is funded by the European Social Fund (ESF). Moreover, we would like to thank all the persons who participated during the exercise recordings.

REFERENCES

Huang, T.-C., Cheng, Y.-C., and Chiang, C.-C. (2013). Automatic Dancing Assessment Using Kinect. In *Advances in Intelligent Systems and Applications-Volume 2*, pages 511–520. Springer.

Khan, N. M., Lin, S., Guan, L., and Guo, B. (2014). A visual evaluation framework for in-home physical rehabilitation. In *Multimedia (ISM), 2014 IEEE International Symposium on Multimedia*, pages 237–240. IEEE.

Lin, T.-Y., Hsieh, C.-H., and Lee, J.-D. (2013). A kinect-based system for physical rehabilitation: Utilizing tai chi exercises to improve movement disorders in patients with balance ability. In *2013 7th Asia Modelling Symposium*, pages 149–153. IEEE.

Muneesawang, P., Khan, N. M., Kyan, M., Elder, R. B., Dong, N., Sun, G., Li, H., Zhong, L., and Guan, L. (2015). A machine intelligence approach to virtual ballet training. *IEEE MultiMedia*, 22(4):80–92.

Richter, J., Wiede, C., Apitzsch, A., Nitzsche, N., Lösch, Christiane amd Weigert, M., Kronfeld, T., Weisleder, S., and Hirtz, G. (2016). Assisted Motion Control in Therapy Environments Using Smart Sensor Technology: Challenges and Opportunities. In *Proceedings of Zukunft Lebensräume Kongress 2016*, pages 90–102.

Shotton, J., Girshick, R., Fitzgibbon, A., Sharp, T., Cook, M., Finocchio, M., Moore, R., Kohli, P., Criminisi, A., Kipman, A., et al. (2013). Efficient human pose estimation from single depth images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 35(12):2821–2840.

Su, C.-J., Chiang, C.-Y., and Huang, J.-Y. (2014). Kinect-enabled home-based rehabilitation system using Dynamic Time Warping and fuzzy logic. *Applied Soft Computing*, 22:652–666.

Tak, Y.-S., Rho, S., and Hwang, E. (2011). Motion Sequence-Based Human Abnormality Detection Scheme for Smart Spaces. *Wireless Personal Communications*, 60(3):507–519.

Tormene, P., Giorgino, T., Quaglini, S., and Stefanelli, M. (2009). Matching incomplete time series with dynamic time warping: an algorithm and an application to post-stroke rehabilitation. *Artificial intelligence in medicine*, 45(1):11–34.

Yurtman, A. and Barshan, B. (2013). Detection and evaluation of physical therapy exercises by dynamic time warping using wearable motion sensor units. In *Information Sciences and Systems 2013*, pages 305–314. Springer.