An Objective Evaluation Method for Rehabilitation Exergames

1st Reza Haghighi Osgouei
Dept. of Surgery and Cancer
Imperial College London
London, UK
r.haghighi-osgouei@imperial.ac.uk

2nd David Soulsby
Acute Team Lead Paediatric Physiotherapist
Chelsea and Westminster Hospital
London, UK
david.soulsby@chelwest.nhs.uk

3rd Fernando Bello
Dept. of Surgery and Cancer
Imperial College London
London, UK
f.bello@imperial.ac.uk

***Abstract*—The aim of this work is to objectively evaluate the performance of patients using a virtual rehabilitation system called MIRA. MIRA is a software platform which converts conventional therapeutic exercises into games, enabling the user to practice the given exercise by playing a game. The system includes a motion sensor to track and capture user's movements. Our assessment of the performance quality is based on the recorded trajectories of the human skeleton joints. We employ two different machine learning approaches, dynamic time warping (DTW) and hidden Markov modeling (HMM), both widely used for gesture recognition, to compare the user's performance with that of a reference as ground truth.**

***Index Terms*—Rehabilitation, exergame, objective evaluation, DTW, HMM, Kinect**

I. Introduction

Rehabilitation is essential to regain the lost or weakened functionality after injury or surgery. While it is initiated in a clinic and supervised by a clinician, the therapeutic exercises must be practiced at home by the patients themselves. The lack of motivation and compliance hinder the healing process and even in some cases worsen the injury. With the advances in virtual reality technologies, various virtual rehabilitation platforms have been introduced addressing this issue [1]–[3]. They are generally equipped with some sensory devices to track and monitor user's motions. Among them are systems based on the concept of *exergaming*, exercise gaming. These are interactive video games, with some simple scenarios, that enable a user to perform a therapeutic exercise by playing the game. Accessibility, not needing the constant presence of a physician, and entertain-ability, turning repetitive tasks to playful activities, are the two main advantages of such systems. However, they tend to lack objective, clinically meaningful evaluation of the user's performance. At best, game scores, the extent to which the player achieved the game goals, are reported in the end. This work aims to fill this gap by introducing an approach to compare the patient's performance with that of a reference using MIRA (Medical Interactive Recovery Assistant) rehabilitation platform. We apply two different machine learning techniques, dynamic time warping (DTW) and hidden Markov model (HMM), which have been widely used for gesture recognition [4]–[9], to study acquired motion trajectories. Comparing two trajectories, user's vs. reference, each approach outputs an objective similarity score indicating how similar the performances were.

The paper is organized as follows. In section II we review some of the related work. We then introduce MIRA platform and our data collection using Kinect in section III. In section IV, the two machine learning approaches are outlined. The results and discussion are given in section V, followed by conclusion and future work remarks in section VI.

II. Related Work

A. Rehabilitation Platforms

The concept of exergaming enables users exercise while playing games. For players, it is an opportunity to play games in a more active and less passive way. For patients, it offers the chance to practice therapeutic tasks in a more playful and less repetitive manner. Exergames offer various activities, such as aerobic exercises and dancing, balance and stretching workouts, and recreational simulations such as golf, skiing, and more. However, they exert additional requirements on hardware and software. Hardware-wise, they require a proper sensory equipment to track user's motion. In terms of software, the game scenario must fit into whole body interaction. There are various commercially available game consoles enabling exergames, including Xbox (Microsoft), PlayStation (Sony), and Wii (Nintendo). Each comes with its own dedicated input device for enabling user interaction with the games, i.e., Kinect for Xbox, Move for PlayStation, and Remote Plus for Wii.

Among them, the Kinect has gained higher popularity due to its acceptable performance and versatility [10]. Kinect enables interaction with virtual environments using gestures rather than conventional controllers. Since its launch, researchers began to use Kinect for various applications including rehabilitation [11]–[16]. Kinerehab is introduced in [11] to assist therapists in rehabilitating students in a public school. They showed a significant increase in patients motivation and hence improved exercise performance. The Kinect serious game for physiotherapy (KSGphysio) is proposed in [12] with a mobile interface to facilitate analysis of patients progress by generating relevant statistics. A web-based serious game called Therasoup is developed in [13] to improve patients motivation and to provide technical data to the physiotherapist. In [14] they developed a Kinect-enabled home-based rehabilitation system (KEHR) to assist patients in conducting safe and effective off-hospital rehabilitation without the immediate supervision of a physician. A serious game framework for therapy (Theragame) providing options to imitate the actions performed by an avatar or to play a game that trains specific parts of the body is introduced in [15]. A web-based platform for physical tele-rehabilitation for patients after hip replacement surgery is described in [16] having two goals in mind: making use of a low-cost motion capture device (Kinect) with a real-time automatic assessment of the execution correctness. While Kinect-based rehabilitation systems are accepted by both patients and therapists, the lack of objective, clinically meaningful evaluation of the user's performance raises questions regarding their effectiveness.

Commerciallly available rehabilitation platforms based on Kinect and exergames include MIRA [1], VirtualRehab [2], and REHABILITY [3]. MIRA, a class I medical device, uses games, which are built based on best clinical practice and expertise from specialist physiotherapists, to keep the patient engaged and motivated throughout the therapy. VirtualRehab, a CE certified class I medical device, is a product that can be used in clinics and hospitals, as well as in the patient's homes, allowing them to continue their rehabilitation treatment, which leads to improved patient outcomes. With REHABILITY, patients can carry out their rehabilitation exercises, from either the centre in which they are hospitalized, or remotely, autonomously, but with constant medical supervision.

B. Performance Evaluation

Automatic performance evaluation of a user carrying out a task has been always a challenge among researchers in both medical and non-medical domain. Such evaluations are less problematic in the real world when they are subjectively assessed by judges who are experts in the given field. For example, evaluation of the quality of a dance, a gymnastic performance, or a physiotherapy/rehabilitation exercise, is carried out by expert dancers, sport masters, and professional therapists, respectively. Such evaluations require the presence of human specialists, who are not easily accessible or affordable to everybody. In addition, the fact that the assessment is subjective, means that a different expert might have a different opinion. The benefits are evident if such evaluation is done automatically by a computer or machine, given that human performance is properly captured by sensory tools. A real world example is a video game called *Just Dance* developed by the French company Ubisoft for Microsoft Xbox. Using the Kinect sensor, the players must mimic the on-screen dancer's choreography to a chosen song. The system then continuously evaluates in real-time the quality of a user's dance movements in terms of being 'Ok', 'Good', 'Super', or 'Perfect', and reports a total numeric score at the end.

Regarding automated assessment of therapy motions, the studies in the literature are scarce [17]–[20], and not much attention has been paid to the development of metrics for perfromance evaluation [21]. As a common scheme, a reference model is captured as the ground truth first. Then, a user's performance is compared to the reference using machine learning approaches. A comprehensive taxonomy of the metrics for evaluation of patient performance in physical therapy is given in [21]. The metrics are classified into quantitative and qualitative categories. Further the quantitative metrics are divided into model-less (based on raw measurements of motions) and model-based (based on a mathematical model of the motions). Utilizing KEHR, the authors used DTW and fuzzy logic to provide real-time subjective discrepancies between the model exercise and patient's performance [14]. Applying HMM and defining an accpet/reject interval, a method to detect deviations from normal repetitions in theraputic activities is presented in [17]. They later compared the performance of their HMM-based technique with that of DTW in [22]. A similar approach using HMMs to assess the correctness of tele-rehabilitation exercises is employed in [18], whereas a cloud-based physical therapy monitoring and guidance system is proposed in [19], which applies DTW to produce subjective assessments in terms of being too slow/fast or overdone/incomplete. Lastly, [20] presents a method based on incremental DTW to classify the incorrectness of the user's performance for a hip abduction exercise into four discrete categories: bent knee, foot outside, upper body, and wrong plane. What is common between all these efforts is that they focused on evaluating the incorrectness of the user's performance on the basis of some subjective terms. In most cases, the developed method was used to sort multiple erratic performances with respect to a reference template. This motivated us to employ the common practice techniques, DTW and HMM, but to generate a similarity score between a user's performance and a reference.

III. Materials

A. MIRA

MIRA [23] is a software platform that turns physiotherapy exercises into clinical exergames, increasing engagement levels by converting the rehabilitation sessions into entertaining activities, making therapy more convenient and easier to follow, offering greater accessibility and improving uptake of exercises. In turn, this has the potential of better recovery times for patients, as well as supporting the therapists and reducing the workload and waiting times at clinics. MIRA has been used in several clinical studies [24]–[27].

The MIRA system includes a Kinect sensor (Microsoft Corp.) connected to a computer running the MIRA program (Fig. 1). Currently, 32 exercises and 25 games are supported. Each rehabilitation session requires selecting an exercise and a suitable game (Fig. 2). Shoulder abduction, elbow flexion and side strides are among the exercises, whereas *Firefly*, *Fishing* and *Football* are examples of the available games. Once an exercise has been selected, adequate game options to choose are presented. The selected combination of exercise and game is then added to the session and can subsequently be executed. Each execution starts with a process to calibrate the user's position in front of the Kinect. A short video tutorial explaining the exercise is followed by another video tutorial describing the game mechanics. As the game starts, the user must play it by moving the intended body part (i.e. left arm, right leg, or neck) in the manner shown in the video. At the end of each session, game scores are reported. Depending on the game, the score reflects to what extent the player followed the game's objectives, for example, the number of fish caught and taken to the boat, or the number of times the spaceship is safely passed through the fire rings. While the scores can be an indication of how well the user played the game, they do not have much value in a clinical context. The aim of this work is to introduce an objective evaluation method that is more meaningful and suitable for clinical evaluation.

B. Data collection

We developed a program in Unity 3D game engine (Unity Technologies - unity3d.com) to capture and store raw 3D position coordinates of the select joints using the Kinect. This was needed as the MIRA program does not allow accessing joint data while playing an exergame due to regulations imposed on class I medical devices. Using Kinect V2 (the second and last version of Kinect), 3D position coordinates of 25 different human skeleton joints can be tracked (Fig. 3(a)) with an update rate of 30 fps. However, it is not necessary to track all the joints, but only those that are involved in the exercise. For this initial study, we have chosen shoulder abduction of the left arm as the exercise. The correct or reference execution requires keeping the arm fully stretched while moving from 0° to 180° as shown in Fig. 3(b). We also devised two incorrect executions to objectively compare with the reference. The first one keeps the arm stretched without making the full range of motion (Fig. 3(c)), whilst the second one does not keep the arm stretched and does not make the full range of motion (Fig. 3(d)).

The four cardinal joints involved in this exercise are spin-shoulder (X_1), shoulder-left (X_2), elbow-left (X_3), and wrist-left (X_4) as shown in Fig. 3(a). The 3D position of joint j ($1 \leq j \leq 4$) at time stamp t is denoted by vector $[x_j(t), y_j(t), z_j(t)]$. As position coordinates are dependent on the user size and location in front of the camera (Kinect), we extract two invariant features, namely shoulder angle (θ_2)

Fig. 1: MIRA system including a Kinect sensor to detect user's motion and software to match an exercise with a game.

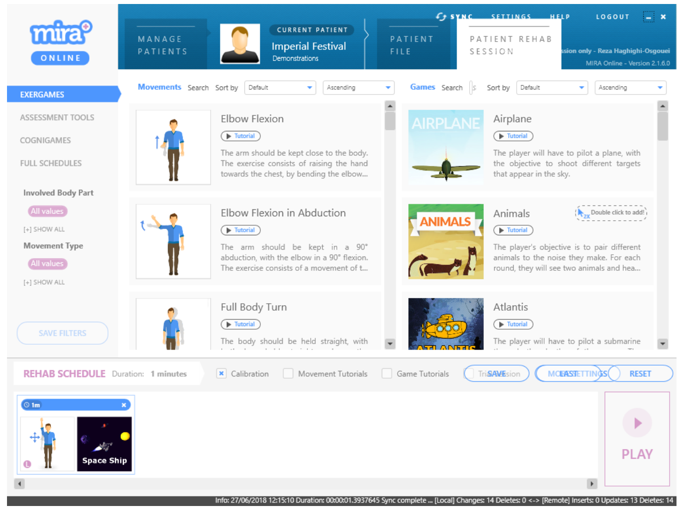


Fig. 2: A snapshot of the MIRA program. The exercises are listed on the left and the games are shown on the right. Selecting an exercise, the user is given multiple game options, enabling to practice the same exercise playing different games. This greatly encourages the user to cope with the prescribed exercise by discovering the various game scenarios.

and arm angle (θ_3) as shown in Fig. 3(a). These two scalar features are sufficient to represent the three executions since θ_2 reflects the range of motion and θ_3 indicates if the arm is being stretched or not. For each execution, a motion trajectory $T(l)$ is formed by the sequence of feature values within the time frame $0 \leq t \leq l$, with l being the execution time.

$$T(l) = \begin{bmatrix} \theta_2(0), \theta_3(0) \\ \theta_2(1), \theta_3(1) \\ \vdots \\ \theta_2(l), \theta_3(l) \end{bmatrix} \tag{1}$$

For simplicity, the reference trajectory is denoted by T_0 and the two incorrect trajectories by T_1 and T_2, with execution times l_0, l_1, and l_2, respectively. The data of a single user was collected repeating each exercise five times. A sample plot is given in Fig. 4.

IV. Methods

A. Dynamic Time Warping

Dynamic Time Warping (DTW) [28] is a technique to align two time series and find the minimum Euclidean distance between them. It is a frequently used approach in speech recognition to classify sound waves of the same word spoken in different accents and duration. DTW is sensitive to both signal pattern and amplitude. If two signals have the same patterns, for example, the same number of peaks, but different amplitude, then the alignment cannot be done perfectly, thus yielding a large distance between them. If they have the same amplitude, but different patterns, the alignment will also result in a large distance. Therefore, the output distance is a measure

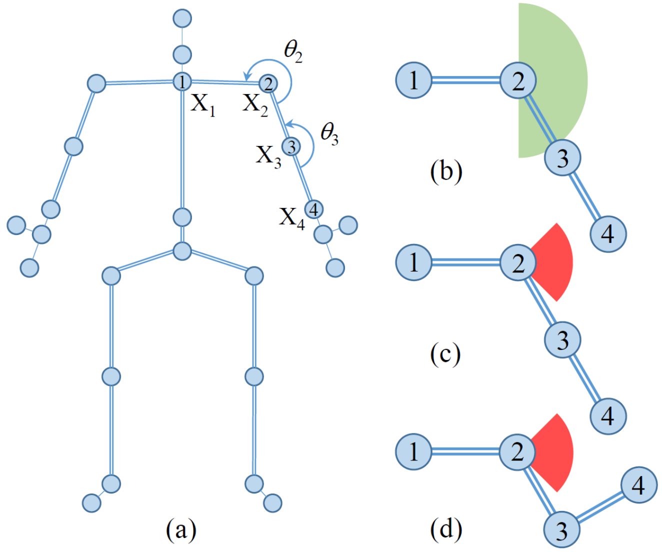
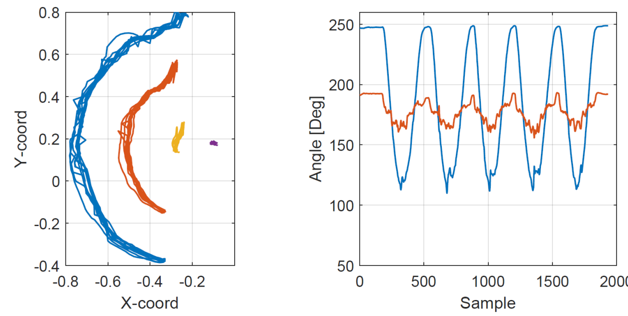
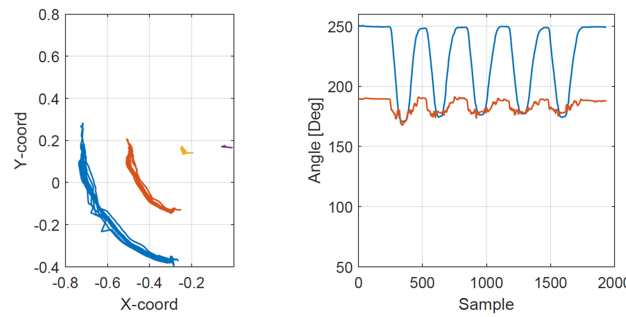


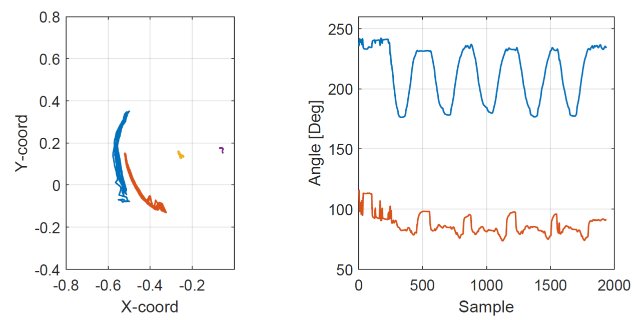
Fig. 3: The 25 skeleton joints tracked by the Kinect V2 (a). Cardinal joints involved in shoulder abduction of the left hand: 1) spin-shoulder, 2) shoulder-left, 3) elbow-left, and 4) wrist-left. From the variant position data, two invariant features are extracted: θ_2 and θ_3. Shoulder abduction is executed in three different ways: a reference with fully stretched arm and full range of motion (b), an incorrect execution with fully stretched arm but half range of motion (c), and a second incorrect execution with a closed arm and half range of motion (d).



(a) Reference execution with open arm and full range of motion.



(b) Incorrect execution with open arm and half range of motion.



(c) Incorrect execution with half close arm and half range of motion.

Fig. 4: Collected position trajectories of four joints (left) and extracted features (right). The joints X_1, X_2, X_3, and X_4 are colored in purple, orange, red, and blue, respectively. Angles θ_2 and θ_3 are colored in blue and red.

of similarity between the two time series. The higher the distance, the greater the deviation.

Whilst it was initially applied to speech recognition, it has also been widely used in gesture recognition due to the similarities between the two domains [5], [7], [29]. In this work, the motion trajectories, each representing a gesture, are classified into the most similar gesture group (i.e. the one with the smallest distance) by converting the distance between two trajectories into a similarity measure.

We define $D_{01} = \text{DTW}(T_0, T_1)$ and $D_{02} = \text{DTW}(T_0, T_2)$ as distances between the correct and two incorrect trajectories. We used Matlab function `dtw` to implement DTW. While the lower limit of this distance is zero, the upper limit is unknown and can be any large value. By estimating an upper limit, it is possible to convert the distance measure into a similarity score. Associating the upper bound with the worst possible performance, an upper limit can be approximated. For the shoulder abduction, the worst performance would be to fully close the arm and to not move at all.

A series of trajectories T_3 were collected for a single user executing this worst performance. Calculating its distance to the reference $D_{03} = \text{DTW}(T_0, T_3)$ allowed us to establish the upper bound for shoulder abduction. Knowing both lower and upper bounds, a given distance $D_l \leq D \leq D_u$ can be transformed into a similarity measure or percentage score S_D:

$$S_D = 100 \times \frac{D - D_l}{D_u - D_l} \tag{2}$$

B. Hidden Markov Modeling

A Hidden Markov Model (HMM) [30] is a stochastic model that considers an observed signal as the result of the transition of a system between several states, each of which has a probability that a particular symbol might be observed. HMMs are useful for recognition of temporal patterns such as speech, handwriting, and gestures. An HMM with discrete observations is mainly specified by the state transition matrix A and the observation matrix B, assuming that the system goes through N different possible states S_1, S_2, ..., S_N and

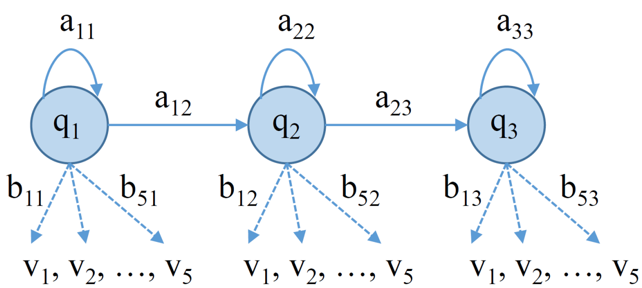


Fig. 5: Structure of a hidden Markov model with three states (q_1 to q_3) and five observation symbols (v_1 to v_5). The relationship among states is described by transition matrix $A = [a_{ij}]_{3\times3}$, and between states and discreet symbols by observation matrix $B = [b_{ij}]_{3\times5}$. It is assumed that the system under study evolves through certain states and we are interested in discovering their relationship.

in each state one of M different symbols v_1, v_2, ..., v_M can be observed (Fig. 5).

HMMs need to be trained to optimize their parameters, matrices A and B. For that purpose, the number of hidden states (N) and the number of symbols (M) should be determined. In addition, the data of multiple repetitions of the same task need to be collected and quantized into single-column symbol vectors. After examining several different values, an adequate consistency in the results was observed setting $N = 5$ and $M = 10$. We used Matlab function `kmeans` to quantize observation vectors and functions `hmmtrain` and `hmmdecode` to train and evaluate our HMM. Once trained, the HMM can be used to calculate the probability (or likelihood) that the given sequence of observations are to be generated by the same process. In our case, the higher the likelihood, the higher the similarity.

For performance evaluation, a single HMM, λ_0, is trained based on the reference motion trajectory T_0. We then calculate log-likelihood of T_1 and T_2 given the trained model by $L_{01} = \log(P(T_1|\lambda_0))/l_1$ and $L_{02} = \log(P(T_2|\lambda_0)/)l_2$. Similar to DTW, lower and upper limits of the log-likelihood need to be calculated. The upper limit is known to be zero since the highest probability is one. However, the lower limit is unknown and can be any small value less than zero. Same as before, we assumed that this lower limits reflects the worst possible performance captured by T_3. Our initial examination showed that, during the worst performance, the system stays in only one state and does not transit between states. This in turn results in a local minimum likelihood value. Observing that from a trained HMM we can find both the most likely state sequence generating a given observation, as well as the least likely state sequence, it is possible to associate the corresponding likelihood value to the lower limit. Letting the lower limit be L_{03}, a similarity score S_H corresponding to log-likelihood $L_l \leq L \leq L_u$ is obtained by:

$$S_H = 100 \times \frac{L - L_l}{L_u - L_l} \tag{3}$$

V. Results and Discussion

In this section we report similarity scores obtained by applying DTW and HMM. As mentioned earlier, data of a single user performing the given exercise under three conditions, one correct (T_0) and two incorrect ones (T_1 and T_2) was collected. In addition, we asked the user to perform under the worst condition, closing arm and not moving (T_3), and also randomly (T_4). Each condition was repeated five times.

Similarity scores obtained by applying DTW are listed in Table I. The distance between each pair of five repetitions of the reference trajectories was calculated first. The minimum distance was then used as the lower limit D_l and the trajectory which was closest to the other four as the minimum reference trajectory. The upper limit D_u was established by obtaining the maximum distance between this trajectory and each of the five worst trials. Similarity scores were then calculated using equ. 2. On average, the first incorrect performance (full arm with half range of motion) is 88% similar to the reference execution. The second incorrect performance (half closed arm and half range of motion) is 42% similar to the reference. The worst trials received the minimum score (0.6% on average). However, random trials achieved a similarity score between 28% and 60%.

TABLE I: The similarity scores obtained applying DTW.

S_D (%)	tr1	tr2	tr3	tr4	tr5	avg, std
D_{00}	98.3	98.6	100.0	98.7	98.7	98.9, 0.66
D_{01}	89.6	88.3	87.9	87.7	88.2	88.3, 0.78
D_{02}	43.5	41.3	39.8	42.5	43.9	42.2, 1.70
D_{03}	0.0	0.8	0.6	0.6	1.2	0.6, 0.43
D_{04}	60.4	59.6	57.3	28.4	41.6	49.5, 14.04

Similarity scores obtained applying HMM are listed in Table II. The log-likelihood between each pair of five repetitions of the reference trajectories was calculated first. The maximum log-likelihood was considered as the upper limit L_u, with the trajectory more similar to the other four used as the maximum reference trajectory. The lower limit L_l was established by finding the less likely state sequence given the maximum reference trajectory. Similarity scores were then calculated using equ. 3. On average, the first incorrect execution is only 64% similar to the reference and the second incorrect is only 32%. For the worst performance, the score is the same for all trials (47%) and the similarity of random trials are between 18% and 56%.

TABLE II: The similarity scores obtained applying HMM.

S_H (%)	tr1	tr2	tr3	tr4	tr5	avg., std
L_{00}	88.3	91.3	100.0	94.9	95.6	94.0, 4.46
L_{01}	75.6	70.7	58.0	63.2	54.3	64.4, 8.79
L_{02}	29.7	22.2	26.0	35.2	49.1	32.4, 10.48
L_{03}	47.4	47.4	47.4	47.4	47.4	47.4, 0.00
L_{04}	56.4	32.1	49.5	18.1	37.1	38.6, 15.00

A comparison between the average similarity scores obtained by DTW and HMM is given in Fig. 6.

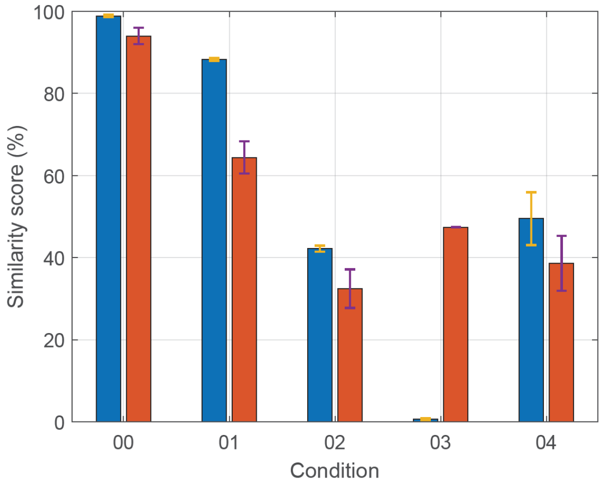


Fig. 6: Average similarity scores obtained by DTW (blue) and HMM (red). Error bars indicate standard error.

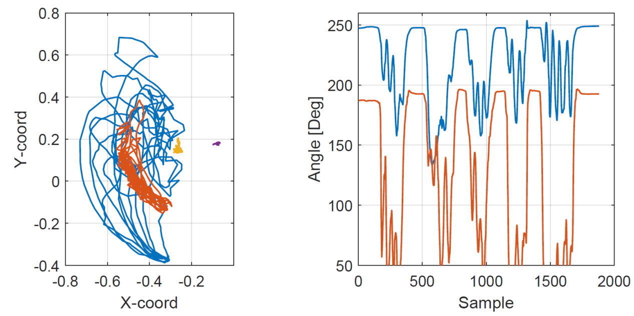


Fig. 7: Random trajectories of four joints (left) and corresponding extracted features (right). Most of the random movements include some form of shoulder abduction built-in, this can be seen by comparing the elbow position (X3, colored in red) and shoulder angle (θ_2, colored in blue) with those of the reference (Fig. 4(a)). The color code is same as Fig. 4.

Several observations can be made from the results. Smaller similarity scores were expected for T_2 than for T_1, since T_2 included two deviations from the reference and T_1 only one. In that respect, both approaches show similar trends. For T_2 vs. T_1, DTW reported 42% $<$ 88%, while HMM reported 32% $<$ 64%. However, the absolute scores are different. The similarity for T_1 is reported as 88% by DTW and 64% by HMM. DTW found T_1 more similar to the reference than HMM. It would appear that DTW is less sensitive to range of motion deviations than HMM. The similarity for T_2 is reported as 42% by DTW and 32% by HMM. In this case, the scores are closer and both are less than 50%.

Variations between the scores obtained for different trials by DTW are smaller than for HMM. For example, the standard deviation for D_{00} is significantly smaller than L_{00} (0.66 vs. 4.46). The lowest score reported between reference trials by DTW is 98.3% while it is 88.3% by HMM. It would appear that HMM tends to be more sensitive to small-scale differences compared to DTW.

This different level of sensitivity can be advantageous. Assuming a patient just started a rehabilitation process, highlighting small deviations might not be a good idea. So, in the early phases, scores reported by DTW might be more helpful and encouraging. Once the patient's performance has improved, reporting small deviations could better assess his/her performance.

Regarding the random trajectories, small similarity scores were initially expected given the randomness of the movements. However, the obtained scores were relatively high, with an average of 49% reported by DTW and 38% by HMM. We initially concluded that both approaches have failed to report meaningful values, and that an additional process would be required to detect outliers such as random movements in advance. But, having a closer look at the recorded data (Fig. 7), we realized that the arm movements were not as random as first thought. In four out of five random trials, the initial movement of the arm was still shoulder abduction with minimal shoulder adduction or extension present, resulting in a higher similarity score than expected. If the first shoulder movement was flexion, then a likely natural occurrence for the next movement would be abduction, again producing a relatively high similarity score. This realization confirms the ability of both approaches to adequately capture the inherent similarity between the random movements and the reference trajectory.

The reference score for the movement could have some inaccuracies as different age groups would have different abilities. For example, a 5 year old would have less ability to stand on one leg than a teenager or adult, but this would still be age appropriate. There would be a psychological/competitive aspect to the movement when put in a game situation, which may encourage people to do the movements at different speeds. For example a child performing shoulder abduction may speed up playing games to try and get a better score, but in turn reduce the quality of their movement. It is therefore important to consider age-specific reference movements, as well as the impact of the game itself and other psychological aspects on the resulting similarity scores.

Computationally, DTW is simpler to implement than HMM since the latter needs to be trained in advance to optimize its parameters, requiring a fairly large amount of data to be collected for this purpose. DTW does not require any training, being able to directly compare unknown and reference trajectories. On the other hand, since DTW is sensitive to both signal pattern and amplitude, it does need the extraction of invariant features, as well as segmenting the right part of the trajectories to compare.

VI. Conclusion

We presented the results of applying two machine learning approaches, DTW and HMM, to objectively evaluate a patient's performance with respect to a reference conducting a therapeutic exercise using the MIRA rehabilitation system. The motivation behind this study was to introduce a more clinically relevant measure than the currently used game scores. Tested on a shoulder abduction exercise, we have reported that the similarity scores obtained by both approaches, well reflect the level of inconsistency between the correct and incorrect performances. The scores obtained by each technique for the same task were different, indicating their different level of sensitivity. For example, for the given exercise, DTW is less sensitive than HMM to deviations on range of motion than on the arm not being fully stretched. In addition, HMM is more sensitive than DTW to subtle variations from the reference. This suggests that each method might be more suitable at certain stages of rehabilitation or indeed for certain exercises. In early phases of a rehabilitation process, DTW-based evaluation might be more effective not focusing on the details, whereas later on, HMM-based evaluation could be advantageous to highlight subtle inconsistencies.

We intend to improve the current work in two aspects. First, by conducting a more substantial human user study with actual patients. Second, by correlating a physician's evaluation of the performances with the proposed objective similarity score.

Acknowledgment

The project was funded through MedCity's Collaborate to Innovate programme and received funding from the European Regional Development Fund and the Higher Education Funding Council for England. We also acknowledge the support of MIRA Rehab Ltd.

References

[1] "MIRA Rehab." [Online]. Available: www.mirarehab.com (last accessed 28/06/2018)

[2] "REHABILITY." [Online]. Available: rehability.me (last accessed 28/06/2018)

[3] "VirtualRehab." [Online]. Available: www.virtualrehab.info (last accessed 28/06/2018)

[4] M. Moni and A. S. Ali, "Hmm based hand gesture recognition: A review on techniques and approaches," in *Proceedings of the IEEE 2nd International Conference on Computer Science and Information Technology (ICCSIT)*. IEEE, 2009, pp. 433–437.

[5] R. Ibanez, Á. Soria, A. Teyseyre, and M. Campo, "Easy gesture recognition for kinect," *Advances in Engineering Software*, vol. 76, pp. 171–180, 2014.

[6] D. Antón, A. Goñi, A. Illarramendi *et al.*, "Exercise recognition for kinect-based telerehabilitation," *Methods of Information in Medicine*, vol. 54, no. 2, pp. 145–155, 2015.

[7] J. Raheja, M. Minhas, D. Prashanth, T. Shah, and A. Chaudhary, "Robust gesture recognition using kinect: A comparison between dtw and hmm," *Optik-International Journal for Light and Electron Optics*, vol. 126, no. 11-12, pp. 1098–1104, 2015.

[8] A. D. Călin, "Gesture recognition on kinect time series data using dynamic time warping and hidden markov models," in *Proceedings of the 18th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC)*, Sept 2016, pp. 264–271.

[9] A. D. Călin, H. F. Pop, and R. F. Boian, "Improving movement analysis in physical therapy systems based on kinect interaction," in *Proceedings of the 31st British Computer Society Human Computer Interaction Conference*. BCS Learning & Development Ltd., 2017, p. 87.

[10] H. Mousavi Hondori and M. Khademi, "A review on technical and clinical impact of microsoft kinect on physical therapy and rehabilitation," *Journal of medical engineering*, vol. 2014, 2014.

[11] Y.-J. Chang, W.-Y. Han, and Y.-C. Tsai, "A kinect-based upper limb rehabilitation system to assist people with cerebral palsy," *Research in Developmental Disabilities*, vol. 34, no. 11, pp. 3654–3659, 2013.

[12] N. Duarte, O. Postolache, and J. Scharcanski, "Ksgphysio-kinect serious game for physiotherapy," in *Proceedings of the International Conference and Exposition on Electrical and Power Engineering (EPE)*. IEEE, 2014, pp. 606–611.

[13] F. Cary, O. Postolache, and P. Girão, "Kinect based system and serious game motivating approach for physiotherapy assessment and remote session monitoring," *Proceedings of the International Conference on Sensing Technology (ICST)*, pp. 2–4, 2014.

[14] C.-J. Su, C.-Y. Chiang, and J.-Y. Huang, "Kinect-enabled home-based rehabilitation system using dynamic time warping and fuzzy logic," *Applied Soft Computing*, vol. 22, pp. 652–666, 2014.

[15] D. Ferreira, R. Oliveira, and O. Postolache, "Physical rehabilitation based on kinect serious games," in *Proceedings of the 11th International Conference on Sensing Technology (ICST)*. IEEE, 2017, pp. 1–6.

[16] Y. Rybarczyk, J. K. Deters, C. Cointe, and D. Esparza, "Smart web-based platform to support physical rehabilitation," *Sensors*, vol. 18, no. 5, 2018.

[17] C. Palma, A. Salazar, and F. Vargas, "Hmm based evaluation of physical therapy movements using kinect tracking," in *Proceedings of the International Symposium on Visual Computing*. Springer, 2015, pp. 174–183.

[18] J. K. Deters and Y. Rybarczyk, "Hidden markov model approach for the assessment of tele-rehabilitation exercises," *International Journal of Artificial Intelligence*, vol. 16, no. 1, pp. 1–19, 2018.

[19] W. Wei, Y. Lu, E. Rhoden, and S. Dey, "User performance evaluation and real-time guidance in cloud-based physical therapy monitoring and guidance system," *Multimedia Tools and Applications*, pp. 1–31, 2017.

[20] J. Richter, C. Wiede, B. Shinde, and G. Hirtz, "Motion error classification for assisted physical therapy-a novel approach using incremental dynamic time warping and normalised hierarchical skeleton joint data." in *Proceedings of the 6th International Conference on Pattern Recognition Applications and Methods (ICPRAM)*, vol. 1, 2017, pp. 281–288.

[21] A. Vakanski, J. M. Ferguson, and S. Lee, "Metrics for performance evaluation of patient exercises during physical therapy," *International journal of physical medicine & rehabilitation*, vol. 5, no. 3, 2017.

[22] C. Palma, A. Salazar, and F. Vargas, "Hmm and dtw for evaluation of therapeutical gestures using kinect," *arXiv preprint arXiv:1602.03742*, 2016.

[23] A. Cantea, C. Mihaiu, A. Dascălu, and A. Călin, "Mira," *Proceedings of the 31st International BCS Human Computer Interaction Conference (HCI)*, 2017.

[24] I. Moldovan, A. Călin, A. Cantea, L. Dascălu, C. Mihaiu, O. Ghircău, S. Onac, O. Rîză, R. Moldovan, and L. Pop, "Development of a new scoring system for bilateral upper limb function and performance in children with cerebral palsy using the mira interactive video games and the kinect sensor," in *Proceedings of the International Conference on Disability, Virtual Reality and Associated Technologies*, 2014, pp. 2–4.

[25] M. McGlinchey, A. Curtis, R. Ritchie, and G. D. Jones, "Feasibility of using mira with adult patients presenting with upper limb motor dysfunction post neurological damage," *Studies in Health Technology and Informatics*, vol. 217, pp. 696–702, 2015.

[26] N. Czakó, C. Silaghi, and R. S. Vizitiu, "Rehabilitation of a 7-year-old boy with brachial plexus palsy using mira exergames and vojta therapy," in *Proceedings of the E-Health and Bioengineering Conference (EHB)*. IEEE, 2017, pp. 349–352.

[27] I. Moldovan, L. Tric, R. Ursu, A. Podar, A. Călin, A. Cantea, L. Dascălu, and C. Mihaiu, "Virtual rehabilitation programme using the mira platform, kinect and leap motion sensors in an 81 years old patient with ischemic stroke," in *Proceedings of the E-Health and Bioengineering Conference (EHB)*. IEEE, 2017, pp. 325–328.

[28] H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 26, no. 1, pp. 43–49, 1978.

[29] M.-C. Hu, C.-W. Chen, W.-H. Cheng, C.-H. Chang, J.-H. Lai, and J.-L. Wu, "Real-time human movement retrieval and assessment with kinect sensor," *IEEE Transactions on Cybernetics*, vol. 45, no. 4, pp. 742–753, 2015.

[30] L. R. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, 1989.