Home-based Physical Therapy with an Interactive Computer Vision System

Yiwen Gu, Shreya Pandit, Elham Saraee, Timothy Nordahl, Terry Ellis, and Margrit Betke
Boston University
{yiweng,shreyap,saraee,nordahlt,tellis,betke}@bu.edu

Abstract

In this paper, we present ExerciseCheck. ExerciseCheck is an interactive computer vision system that is sufficiently modular to work with different sources of human pose estimates, i.e., estimates from deep or traditional models that interpret RGB or RGB-D camera input. In a pilot study, we first compare the pose estimates produced by four deep models based on RGB input with those of the MS Kinect based on RGB-D data. The results indicate a performance gap that required us to choose the MS Kinect when we tested ExerciseCheck with Parkinson's disease patients in their homes. ExerciseCheck is capable of customizing exercises, capturing exercise information, evaluating patient performance, providing therapeutic feedback to the patient and the therapist, checking the progress of the user over the course of the physical therapy, and supporting the patient throughout this period. We conclude that ExerciseCheck is a user-friendly computer vision application that can assist patients by providing motivation and guidance to ensure correct execution of the required exercises. Our results also suggest that while there has been considerable progress in the field of pose estimation using deep learning, current deep learning models are not fully ready to replace RGB-D sensors, especially when the exercises involved are complex, and the patient population being accounted for has to be carefully tracked for its "active range of motion."

1. Introduction

Home-based exercising can become an essential part of any physical therapy program. With enough motivation and correct execution of the exercises at home, a patient can achieve faster recovery than with conventional in-clinic-only physical therapy. To help patients improve their experience during home-based physical therapy by providing them with guidance and feedback, the design of an ideal assistive system needs to tackle two difficult challenges: First, the system has to be capable of performing accurate, real time analysis of the movements of the patient during his or her exercise and provide the patient with motivating feed-

Figure 1: Overview of task

back. Second, it has to summarize the patient performance consistently so that the physical therapist (and patient) can evaluate the patient's performance both quantitatively and qualitatively both at the clinic and patient's home. Both challenges require a computer vision module to compute and analyze human pose estimates from video input.

We designed and developed ExerciseCheck with the goal to overcome the two challenges described above. ExerciseCheck is an interactive computer vision system that tracks human body movement and provides analysis. It serves as a rehabilitation platform that remotely monitors users and evaluate their performance at home. It allows data to be captured in a modular fashion. The data source can be joint coordinates estimated through a deep learning framework or can be provided through a depth camera. By leveraging the motion data captured, it performs quantitative analysis based on the recorded trajectories of a patient's movements and provides the patient and the therapist with visual and quantitative feedback. The analysis aims to address problems such as inaccurate movements, improper speed, and inadequate range of motion.

In this paper, we address two tasks. First, we evaluate whether current deep learning models designed for Human Pose Estimation (HPE) on RGB input can provide estimates of joint positions that are sufficiently accurate and granular so that they can be incorporated into ExerciseCheck (Section 3). Second, we describe our continued work on ExerciseCheck in Section 4 (an earlier version is described in [31, 34]) and then report our new experiments and results using ExerciseCheck at patients' homes in Section 5. We present our results both quantitatively and qualitatively based on working with Parkinson's disease patients. The

patients used ExerciseCheck at their homes for a period ranging from two weeks to a month. Finally, in Section 6, we discuss the potential long-term effects of our system in benefiting patients in performing their home-based exercises and our future work based on the feedback we received from the participants.

2. Related work

With studies reporting that physical therapy benefits the treatment of neurodegenerative diseases [5, 14, 40], efforts has been made to design monitoring systems that support home-based exercise and rehabilitation. These systems track human movement to perform analysis and provide feedback to the users. They have been reviewed for physical therapy in either therapeutic settings or daily life [1, 9, 23, 29, 37, 44, 45]. They are designed to address the inherent problems of home-based physical therapy: poor adherence to the correct exercise movement and the lack of motivation in exercising alone at home.

As for motivating patients, gamification as an occupational therapy intervention [2, 6, 20] is attaining research interest. There exists evidence showing that games could to be powerful motivators for engaging users in physical activity, especially games for clinical or rehabilitation purposes. However, there is relatively less public research on how effective the motion sensing capabilities of commercially available gaming devices are [10, 12, 19]. Accurate performance measurements and real time feed back also helps alleviate the problem of poor exercise adherence. Systems that work directly with patients lack quantitative analysis to evaluate the patients' performance to ensure a good adherence to a correct exercise [10, 16, 17, 18, 19, 20, 36] and those that do performance analysis are not evaluated by patients [13, 41, 43, 49]. Contact-based motion tracking devices, although have been proven useful in patients' rehabilitation process, raise concerns for it can cause burden and discomfort to patients with physical injury [7]. In addition, efficacy of such systems may also suffer from patients' low motor control or the lack of synchronization between the device and body parts involved [47]. Hence, in recent years, non-intrusive and contact-less rehabilitation techniques using computer vision have started to become more popular. *Gesture Therapy* [37] is a vision-based system using a gripper with a pressure sensor to facilitate hand and finger rehabilitation after stroke. A classifier based on computer vision that can discriminate rehabilitation-relevant hand postures has also been proposed for upper limb rehabilitation system [49]. In addition, *MotionTalk* [43] and *VERA* (Virtual Exercise Rehabilitation Assistant) [17] are examples of systems that employ a Microsoft (MS) Kinect sensor for tracking a patients movements. Among other notable RGB-D cameras[1], the Intel RealSense[TM] camera has been used for rehabilitation [4, 8, 11]

Pushing forward on the contact-free and low-cost devices, the computer vision community enjoys the challenge of using RGB cameras only to analyze three-dimensional (3D) body pose [24, 25, 26, 28]. In fact, accurate human pose estimation is considered one of the most challenging tasks in the field of computer vision [21, 42], because images of people have large in-class variations caused by the intrinsic deformation of the shape of the human body and high variability in human clothing and environmental factors [22]. For these reasons, deep learning has been considered as a primary candidate for visual recognition tasks. In 2014, Toshev and Szegedy formulated the human pose estimation problem as a regression problem for estimating the position of body joints to be solved with a deep neural network [39]. Since then, a variety of models using deep learning have been proposed to address this problem. For example, a network named "stacked-hourglass" has been proposed that consecutively and repeatedly applies down-pooling and up-sampling so the information is captured at every scale [30]. After that, a "Pyramid Residual Module" has been proposed that can be added as a building block to the architecture of other deep convolutional neural networks to enhance their scale invariance [48].

Other work exploited the temporal information in videos and proposed a spatial ConvNet for human pose estimation [32]. Instead of providing estimates of joint coordinates, the model regresses to a heatmap of joints. Optical flow was used for the heatmap alignment, and spatial fusion layers were added to implicitly learn the dependencies between human body parts.

How does our research improve previous work?
Our evaluation of deep learning models on exercise data and their comparison with estimates obtained from a depth camera (Kinect) provides insights on how well the models are posed to provide accurate estimates for assessment of physical therapy, where not only joint locations but orientation and depth accuracy are important. ExerciseCheck provides quantitative analysis to measure patients' performance. ExerciseCheck was evaluated by patients who used it at home for a period of time. We collected their feedback and made improvements accordingly, from the user interface to the technical functionality.
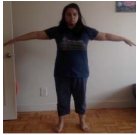
3. Evaluation of Deep Learning Models for Human Pose Estimation

There are many available computer vision and deep learning pipelines available for Human Pose Estimation. These models [27, 30, 32, 38, 39, 46, 48, 50] provide a set of 2D or 3D coordinates for various joints in the human

---

[1]Other notable RGB-D cameras include BlasterX Senz3D (Creative Labs), Xtion Pro Live (AsusTek Computer Inc.), ZED Stereo Camera (Stereolabs Inc.)

Table 1: Comparing the loss between four deep learning models against the MS Kinect. Averages are taken over joint positions estimated by both the deep learning method and the Kinect. The 2D Loss represents the percentage of difference between pixel estimates provided by deep models and Kinect. 3D estimates were reported if available.

M O D E L	Pose 1		Pose 2		Pose 3	
	2D	3D (cm)	2D	3D (cm)	2D	3D (cm)
[27]	5.11	4.86	5.42	8.19	11.40	8.03
[38]	3.76	9.45	3.12	7.62	9.54	20.80
[46]	5.54	-	6.22	-	12.22	-
[33]	3.81	-	3.95	-	10.96	-

body. Most of these models have been trained and tested on datasets having a large number of varied images, such as the Human 3.6M dataset [15] and the MPII dataset [3]. Our approach demonstrates how to compare the pose estimates as provided by depth cameras such as the MS Kinect to those provided by research pipelines based on convolutional and deep learning techniques. This would allow us to check the feasibility of potentially replacing the additional piece of camera hardware that assistance systems often require and perform computations using just the built in device camera such as a webcam on a computer system.

Experimental setup to gather pose data: (1) The webcam and the Kinect were carefully kept at the same height and angle so as to remove positional bias in the image. (2) The user stands in a particular pose in front of the setup, from a simple one such as a Shoulder Abduction to a more complex one such as a Hip Abduction. (3) Coordinates from the depth camera (Kinect in our case) were stored directly with timestamps. (4) The image frames from the webcam were fed into the deep learning models of choice (as listed in Table 1) to obtain corresponding coordinates for the same pose. (5) For a given pose, comparison between the coordinates provided by the depth camera and the model estimates were evaluated using a multi-step approach, which we discuss in the next section.

Algorithm for Evaluation:
- We make the height of the skeletons of both models 100 units so that they can be compared.
- We take the pelvis joint to be the origin for both images (in their own separate coordinate systems). The pelvis is chosen because deep learning models also select it to be the origin if they provide 3D predictions.
- The pairwise Euclidean (L2) distance between each joint is the chosen loss metric. We report the mean

loss across all joints for the 2D analysis, defined as:

$$\text{Loss} = \frac{1}{N} \sum_{i=1}^{N} \left|\left|(x_{im}, y_{im}), (x_{ik}, y_{ik})\right|\right|_2 \quad (1)$$

where the (x_{im}, y_{im}) are the normalized coordinates from the model and (x_{ik}, y_{ik}) are the normalized coordinates from the Kinect. N=16 is the number of joints being evaluated upon, namely: Left/Right Ankle, Left/Right Knee, Left/Right Hip, Pelvis, Thorax (SpineMid), Neck, Head, Left/Right Wrist, Left/Right Elbow, Left/Right Shoulder.
- Since the heights of skeletons are 100, this can be viewed as a difference in percent - e.g., we can now say that the head position from the model was 5% different from the Kinect prediction.
- For comparing the depth estimates, we first convert predictions from all models and the Kinect into metric units (cm). Since most models align the z data with respect to the pelvis, we perform the same alignment for the Kinect, and then compute the mean loss (Eq. 1) based on pairwise distance in the z dimension.
- We also perform a **joint-wise analysis** to find the joint with the most inaccurate location estimates for each pose.

Evaluation Results and Conclusions:
- The availability of code and ease of its use guided our model selection. We evaluated four systems [27, 33, 38, 46]. We report the overall comparison in Table 1 and the results for joint-wise analysis in Table 2.
- Important joints such as SpineMid, Knee and Wrist seem to suffer the most from inaccurate estimation.
- We found that for simple poses like Pose 1 and 2, the 2D predictions are manageable. However, the error increases substantially with a complex pose like Pose 3.
- Fewer models provide 3D estimates, which however is crucial to measuring accuracy in exercises such as squat, lunge, and etc.
- The average error in 3D estimation is almost 10 cm, which would be misleading in understanding patient's performance, especially patients with neurodegenerative diseases where even small distances count.

To summarize, we suggest that although deep learning models have made progress, there are still issues to be resolved before we can use them in a clinical setting to provide accurate performance measure on patients exercises.

4. ExerciseCheck

4.1. Workflow and Hardware Components

The workflow of using the platform is as following. First, during the patient's clinic visit the physical therapist (PT) and the patient are registered into the system and paired. Then the PT shows the patient an exercise, and the

Table 2: Joints with the most inaccurate location estimated by deep learning models compared to the Kinect

Model	Pose 1 - 2D	Pose 1 - 3D	Pose 2 - 2D	Pose 2 - 3D	Pose 3 - 2D	Pose 3 - 3D
[27]	Spine Mid	Right Knee	Spine Mid	Right Knee	Left Wrist	Left Knee
[38]	Pelvis	Right Hip	Right Wrist	Left Ankle	Right Shoulder	Left Ankle
[46]	Spine Mid	-	Spine Mid	-	Left Wrist	-
[33]	Spine Mid	-	Spine Mid	-	Left Wrist	-

patient performs the exercise under the PT's supervision. Once the patient can perform the exercise correctly, the trajectory of the movement is saved into our ExerciseCheck as the *referenceExercise*. After that, when the patient goes home and does the exercise on his/her own, the patient can follow the previously recorded reference to practice the exercise. If satisfied with his/her practice, the patient can save it into the database as a *practiceExercise* for further review. Performance feedback on accuracy and speed is given to the patient at the end of the practice session and is saved so that the patient and the PT can track the performance over time. Since ExerciseCheck evaluates the patients' performance based on their own *referenceExercise*, an individualized performance measure is ensured. Moreover, the *referenceExercise* can always be updated during the patient's clinic visit. This enables updates in the expectation of the therapist for a particular patient over the course of his or her physical therapy.

ExerciseCheck requires two hardware components, a motion sensor (webcam or depth camera) and a computer with internet connection (Fig. 1). The motion sensor is to capture the movement trajectories of the performed exercises, and the computer to visualize, process, and transmit them. For therapist to monitor patients and review their performance, any device with a web-browser will suffice, including a cell phone. As discussed above, we use the MS Kinect as the visual sensor, as it is the most accurate (non-wearable) sensor available. ExerciseCheck captures the data of all the joints for the *referenceExercise* and *practiceExercise*. Given the large volume of the data, compression is employed before storing the data. All data are stored in MongoDB on a server. For more detailed information, please refer to our previous work [31].

4.2. Design and Development

Research and development of ExerciseCheck were done in three phases, where each phase contains design and development of the system accompanied by a set of experiments. The experiments in each phase provided us feedback and directions on what needed to be added in the next phase.

Phase 1: Building a prototype that consists of designing the initial architecture and data storage on a server, implementing a process for quantitative analysis of an exercise, including accuracy and speed, and designing the initial graphical user interface [35]. In our experiment, two

physical therapists and two users without physical disability evaluated our platform.

Phase 2: Major development in the system architecture, data storage, and graphical user interface [31]. The speed analysis was revisited. Exercise repetition counting and analysis of the patient's active range of motion were added [34]. Parkinson's disease patients evaluated our system in their clinic.

Phase 3 (this paper): System updates were made based on the feedback of the participants of the clinic experiments. We prepared ExerciseCheck for the final deployment at the patients' homes. Updates included methods to stop the recording of exercise data automatically and with a minimum amount of error. We tested ExerciseCheck in the patients' homes and performed an analysis of the experimental results.

The user interface of ExerciseCheck is shown in Fig. 2: During the practice exercise, the user can see, in real time, a stick figure of themselves in the *Live View* on the right and the recorded reference trajectory in the *Reference* on the left of the screen. This side-by-side visual feedback is designed so that the patients can easily follow the reference exercise that has been previously approved by their physical therapist. In addition to the stick figures, the interface provides other information such as the name of the exercise, number of sets and repetitions required, and the actual number of the repetitions completed (Fig. 2D).

ExerciseCheck provides the patient and the therapist with an analysis of the accuracy and speed of the patient's movements during the exercise. To do that, ExerciseCheck first normalizes the recorded trajectories. Both the reference trajectory and the practice trajectories are normalized by the corresponding neck point captured in the beginning of the exercise, as well as the "body width" and "body height" variables [34]. Then Dynamic Time Warping is employed to calculate accuracy, and the average over the speed measurements per frame above a given threshold to obtain the user's speed during the exercise excluding the rest time.

Our previous results (based on phases 1 and 2) show that our quantitative analysis is accurate when our system is used in the clinic. All the patients liked the repetition counter, as well as the performance feedback provided to them in the interface. Moreover, we found that the real-time visual feedback helped improve the patients' performance in executing the exercise in a way that their physical therapist recommended. In addition, patients all reported

that the system was easy to use. In the following, we focus on phase 3 of our work, in which our system was used by patients in their homes during the course of their physical therapy.

4.3. Challenges of Phase 3

Our experiments in phase 2 were conducted in the clinic and involved both the physical therapist and the patient. In those experiments, the therapist controlled the computer and told patient when to start. The therapist also stopped the recording when a set of exercise was completed by the patient. However, our phase-3 experiment needs to consider the fact that patients do not always have a companion when performing exercises at their homes. The start of the recording is not a problem because it occurs automatically: After somebody presses the start button, the actual recording will not begin until 5 seconds after the patient's neck is detected in the alignment circle. In this way, the patient can take time to move to the right position. But the design for the stopping of the recording is challenging. Without a second person to click the stop button, the patient has to take several steps to reach the computer and then click the button him/herself, after he/she has finished the practice exercise. These extra movements introduce noise in the data of the practice exercise, which makes measuring the performance accurately difficult. In order to address this problem, we implemented two versions of the stopping procedure; one version introduces a virtual stop button that the user can press with their hands after taking one step forward (Figure 2). The other version automatically stops the recording 3 seconds after the desired number of repetitions are completed (Figure 3).

Both versions were tested with users without physical disabilities, and the accuracy of both methods were acceptable. When asking the patients to try the two approaches, we realized that the virtual stop button adds an additional step for patients to remember and is not visually pleasing. Furthermore, in some cases, it took a long time for the patients to stop the exercise, which compromised the movement accuracy and the speed calculations. On the other hand, the automatic stop button worked perfectly and requires no extra effort from the patient. Thus, for the patients' home-based experiments, we decided to use the version of ExerciseCheck with automatic stopping. Note that the automatic stopping works based on the number of repetitions, thus it cannot be used for the reference exercise. However, this is not causing any problems, since the reference exercise is recorded in the clinic, and, there, the therapist is always present to stop the recording of the exercise.

5. Experiment at Patients' Home

For the third phase of our experiments, we tested our system at patients' homes, so that they can evaluate our system in a real scenario. The goals of this set of experiments are:
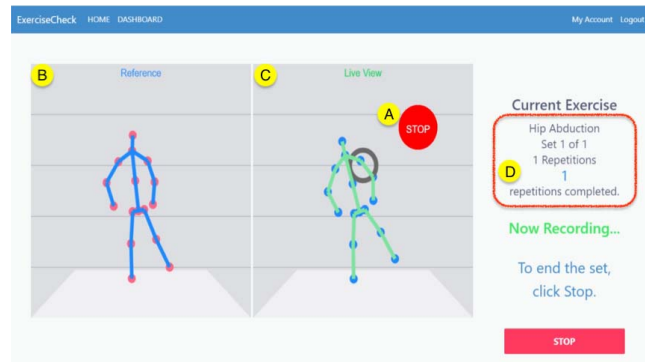

Figure 2: Version of visual stop button. The red visual stop button (A) in the live view (C), where current motion is shown, is placed in a way that takes advantage of the Kinect using a depth camera. The patient must take one step toward the camera in order to activate the visual stop button. The reference view (B) displays the previously recorded *referenceExercise* performed by this patient in clinic. (D) The exercise information includes a repetition counter (blue), which is updated in real time.
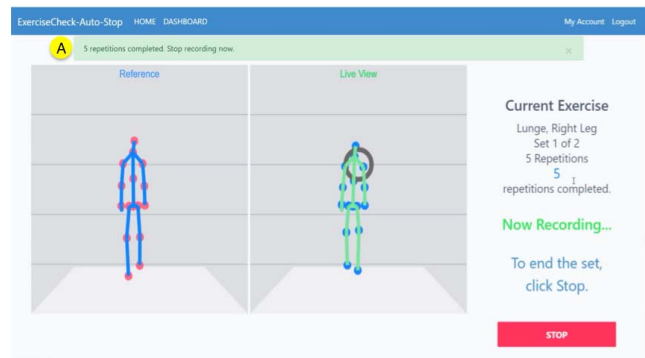

Figure 3: Version of auto-stop. (A) The green pop out message indicates that the program is stopping the recording since the required repetition number is met.

1. To study how patients can work with the system and perform the exercises on their own.
2. To explore the possible scenarios at home that may confuse the system and lead to inaccurate results.
3. To investigate a longer-term impact of the system on the patient during the physical therapy period. For example, to examine the motivational aspect of exercising with a remote monitoring system and receiving performance results after each exercise.

5.1. Design of the Experiment

Similar to our previous experiments, patients were first prescribed an exercise in the clinic, and the reference exercise was recorded under the supervision of the physical therapist. During an exercise session, the user stood approximately 0.75 m from the wall, and the motion capture

device was placed approximately 3.3 m from the wall and 0.9 m from the floor. Then patients were asked to repeat the same exercise at home, at least three times per a week for the duration of a minimum of two weeks. To facilitate the procedure and ensure that the home-based exercise space is correctly set up, we, as investigators of this study, set up ExerciseCheck at each patient's home, following the same layout to the extent possible in the patient's home.

5.1.1 Recruitment Criteria

The recruitment procedure followed the Institutional Review Board (IRB) approved protocol. The **inclusion** criteria identified by our physical therapists for recruiting the participants in our experiments were: (1) age range = 40 to 80 years; (2) Mini-Mental State Exam > 23; (3) able to sit and stand for at least 5 minutes independently; (4) able to understand, communicate with and be understood by recruitment personnel; (5) diagnosed with Parkinson's disease; (6) interested in participating and provide informed consent; and (7) able to perform exercises for 10 minutes.

5.1.2 Study and Exercise Duration

The patients were asked to repeat their prescribed exercises at least three times a week for the minimum of five times total, to the extent it is compliant with their physical condition. If interested, they continued the experiment for a longer period. The first patient worked with our system for 24 days for a total of eight times. Due to a physical problem, the therapist advised him to stop the exercises for a week time between day 10 and 17. He performed two exercises, the hip abduction and lunges. He performed both exercises on both left and right sides.

The second patient used our system for the total of 41 days, thirteen times. Due to a physical problem, she stopped for two separate weeks between day 13 to 20 and 23 to 37. She performed two exercises: (1) marching with one leg (or knee raise) and (2) lunges; both exercises on the left and right sides.

The third patient used our system for two weeks and performed arm raise and lunges for both right and left sides, each five times.

5.2. Results and Discussions

Here, we present the results of our experiment both quantitatively, based on the recorded trajectories, and qualitatively by reviewing the feedback we received from the patients in form of written responses to the interview questions and verbal discussions.

5.2.1 Quantitative Results

The performance scores of the three patients over the course of the home-based experiment are shown in Fig. 4. This figure presents the accuracy and speed ratio values for the right and left sides (leg or hand) of the body. According to the physical therapist's evaluation, all three patients performed well in this experiment. Patient 1 was able to perform the lunge exercise well, where the accuracy score was always above 0.8 (top graph of Fig. 4a). In terms of speed, however, he tended to perform the lunge exercise faster than expected in some cases, especially with the left side (bottom graph of Fig. 4a). Using our system, he was able to correct himself and perform the exercise at the right speed. The first patient found the hip abduction exercise challenging. His accuracy score went down as low as 0.6, and the speed ratio ranged between 0.9 and 1.4 (Fig. 4b). He needs to put more effort on practicing this exercise.

Patient 2 maintained a good performance for the lunge exercise, both in accuracy and speed (Fig. 4c), except in a few cases. For example, on day 4, she performed the lunge exercise, but the results were not satisfactory, so she decided to repeat the same exercise again to obtain a better score. This result was very motivating for us, as it shows that the patients cared about their performance and, using our system, they attempted to perform their daily exercises accurately. Our results also show that patient 2 found the marching exercise more challenging, especially for the left side, shown in red in Figure 4d, where the accuracy score is fluctuating and the speed ratio is above 1 in most cases.

The third patient performed both exercises well, as represented in Figures 4e and 4f. The accuracy and the speed ratio scores are very close to one for both the lunge and the shoulder abduction exercises.

The daily analysis, in addition to informing the patient about the performance scores right after each exercise, highlights how the physical therapist can benefit from our system in order to gather an overall evaluation of the patient's performance and progress over time.

In order to provide the patient and the physical therapist with a better understanding of the performance and the underlying reasons for a given score, we also present them with further analysis regarding one important joint angle that is engaged significantly during a specific exercise. For example, for the hip abduction exercise, ExerciseCheck analyzes the angle between the two legs during the exercise. Figure 5 shows such an analysis for the four exercises. On the right of each graph, you can see the exercise illustration with the corresponding angle highlighted. For each graph, the reference exercise, shown in blue, is compared against two randomly selected practice exercises shown in red and green respectively.

For the hip abduction exercise in Figure 5a, the patient managed to maintain the desired amplitude for the angle, which is around 40 degrees for the reference as well as the two practice exercises. The marching exercise in Figure 5b however, seems to be challenging for the second patient. By comparing the reference (blue line) and the second practice exercise (green line), one can see that the amplitude of the
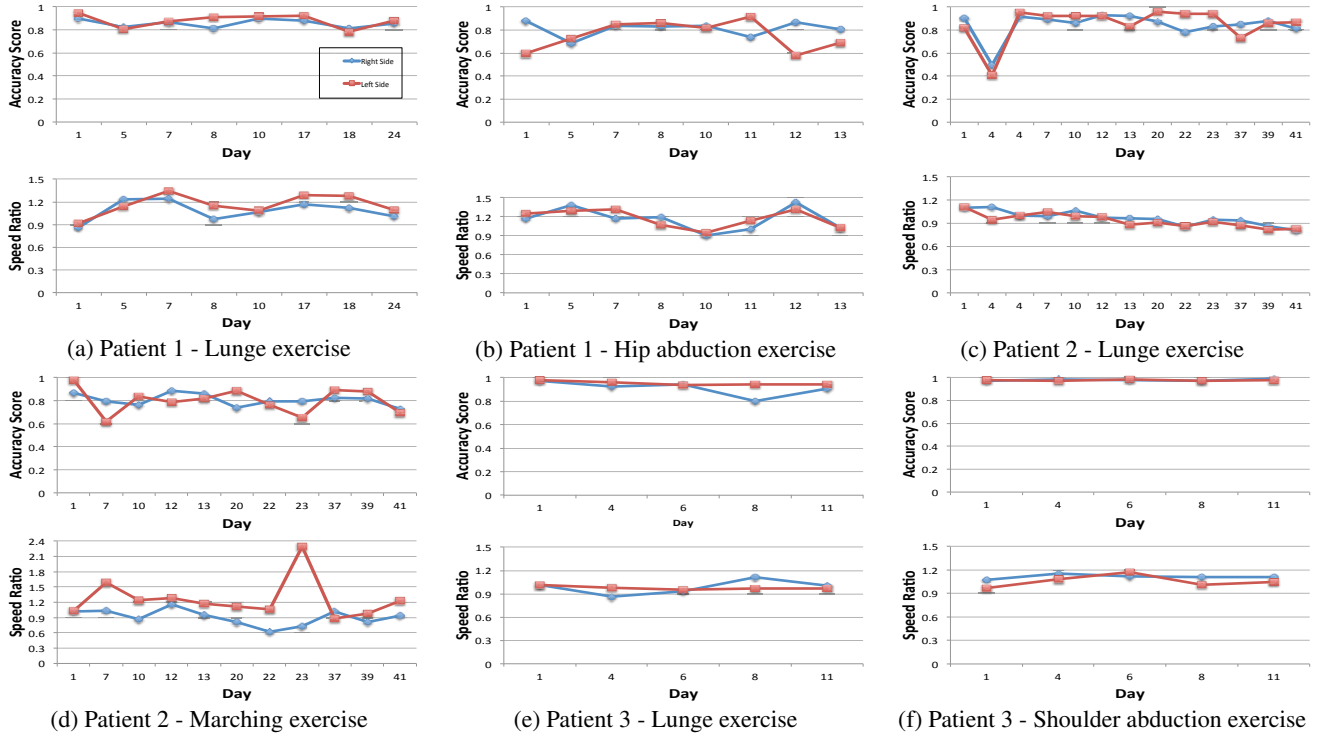
Figure 4: Performance scores for accuracy and speed of the three patients during the home-based experiments obtained by our system. All three patients performed reasonably well according to the therapist's qualitative evaluation. Note that the performance of each patient should be evaluated based on his or her physical capabilities and should not be compared against each other. The blue and red represent the right and left limb respectively

specified angle is decreasing over time for the practice exercise. This analysis informs the patient that the knee was not raised adequately and thus the exercise was not performed accurately. The patient rested for a longer time between each repetition, leading to a longer time over all. However, this rest time did not impact the performance considerably. The third patient performed the shoulder abduction exercise accurately (Fig. 5c) and managed to reach the desired angle. The last graph represents the lunge exercise for the right leg, performed by the third patient. This exercise can be explained in two parts. First the patient took a large step to move the right leg forward, then he moved the hip toward the ground by bending the right knee. Figure 5d highlights the two parts. The small changes in the angle represents stepping forward and the larger peaks represent when the knee is bent. Comparing the reference exercise (blue line) and the second exercise (green line), one can see that the patient did not bend the knee forward enough to reach the expected angle. Our system can notify the patient to go lower for subsequent trials.

5.2.2 Qualitative Results

After the home-based session, each patient was given a set of follow-up questions asking about their experience using ExerciseCheck. We learned the following points: Visual feedback was helpful for the patients to adjust their exercise speed and also to remember what to do and not to do. Two of the patients found it easier to let the reference exercise start playing before they started to move. They followed after approximately two seconds. They stated that, this way, they could easily follow the reference and make sure that they were performing the exercise correctly. One of them reported that the side-by-side view with the reference set was her favorite. The third patient, on the other hand, found it overwhelming. The patient reported that, in his first session, he could not use all the information provided to him while doing the practice exercise. To be more specific, he mentioned that, since he was trying to match his speed to that of the reference exercise, he could not check the repetition counter at the same time. However, he reported that, after multiple times working with the system, he was more familiar with the interface, and he felt he was getting benefits from the presented information on the screen during the practice exercise.

In addition, we learned that, as opposed to the clinic where the patients begin the exercise at the right time, when they are doing an exercise at home, they may begin the exercise a couple of seconds after the recording has started.

(a) Patient 1 - Hip Abduction exercise

(b) Patient 2 - Marching exercise

(c) Patient 3 - Shoulder Abduction exercise
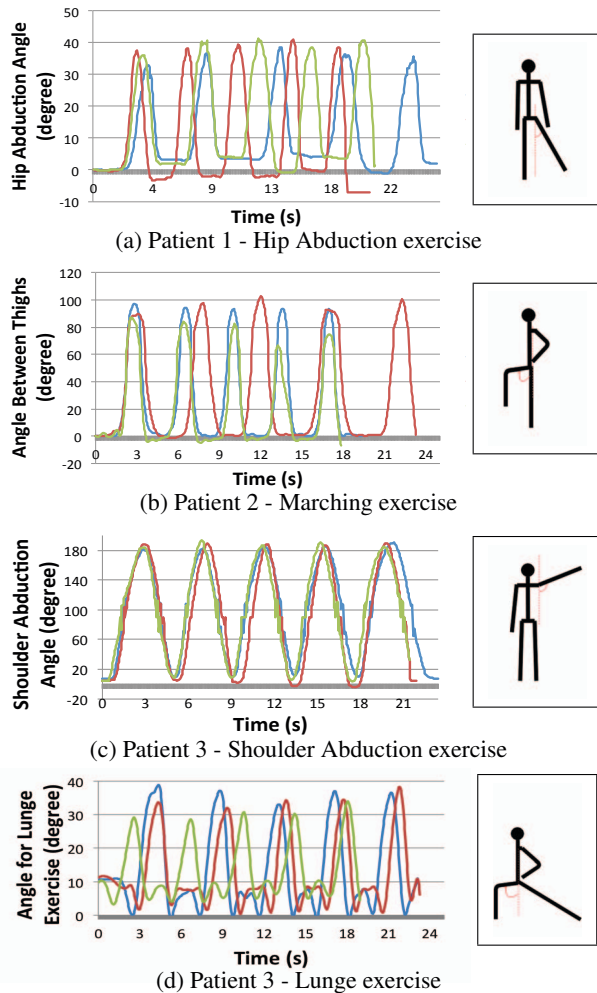

(d) Patient 3 - Lunge exercise

Figure 5: Joint angle most engaged in a particular exercise as a function of time during reference (blue) and two practice (red and green) movements of four exercises.

This has no effect on the speed calculation due the adjusted speed formula implemented in our previous version. However, the accuracy measure may be compromised if a patient begins the exercise late or the length of the rest time is more than expected by the system. Thus, in future work, we need to update the accuracy analysis to be responsive to this type of unexpected noise.

Furthermore, we learned that our normalization mechanism enables our system to handle errors in setting up the exercise and recording space. It works even if the patient does not stand exactly at the same distance from the camera during the practice and reference exercises. However, the height of the table, onto which the camera is placed, is important. Our current system requires that we ensure that the camera is placed at the same height for recording both reference and practice exercises. Future work will investigate how to relax this requirement.

Overall, all three patients mentioned that the Exer-

ciseCheck was a user friendly system, and they were happy with their experience. We also noticed that the quantitative performance measure was a positive factor in motivating the participants to do their exercises and try to improve their performance.

6. Conclusions and Future Work

In this paper, we evaluated the feasibility of using RGB data and deep learning models for human pose estimation for rehabilitation therapy. Our evaluation suggests that deep models are not ready to replace depth cameras for clinical based experiments and analysis. One possible future approach is to employ transfer learning, where we re-train the last few layers of models previously trained on large pose datasets, and then use those for the pose estimates. We will continue experimenting with more models and hope that with current advances in deep learning, we will soon be able to use deep models in real time to provide accurate 3D pose estimates. Moreover, we hope that our paper motivates computer vision researchers to consider the use-case scenario of physical therapy in order to guide their work on improving 3D human pose estimation.

We presented a full-fledged system and experiments at patients' homes. Patients who participated in our experiments found ExerciseCheck easy to use and were able to interact with it well. Our results show the great potential ExerciseCheck has for improving the experience of patients when they are engaged in home-based physical therapy.

Future improvements will be made to ExerciseCheck based on the patients' feedback. One idea is to overlay the practice skeleton on the reference, which appears as a "ghost." This may make it easier for the user to compare speed and motion. In addition, we want to provide users a side-view option. Exercises like squat and lunge can hence have a better visualization. We also plan to develop an algorithm that gives more detailed feedback on how the patient is doing. Instead of giving feedback based on the overall performance, our system will provide instructions like "your third repetition is too fast or not high enough." In such a way, we can direct the patients' focus to where it is needed most and help them improve.

Last but not least, we plan to test the new cloud-based Kinect [2] with our system. Backed by Azure, the computation capability will no longer be limited by our server, and so we will be able to afford computation-intense algorithms and provide users with even more accurate feedback.

---

[2]https://azure.microsoft.com/en-us/services/kinect-dk/

References

[1] G. Alankus, R. Proffitt, C. Kelleher, and J. Engsberg. Stroke therapy through motion-based games: a case study. *ACM Transactions on Accessible Computing (TACCESS)*, 4(1):3, 2011. 2

[2] A. Allam, Z. Kostova, K. Nakamoto, and P. J. Schulz. The effect of social support features and gamification on a web-based intervention for rheumatoid arthritis patients: randomized controlled trial. *Journal of medical Internet research*, 17(1), 2015. 2

[3] M. Andriluka, L. Pishchulin, P. Gehler, and B. Schiele. 2d human pose estimation: New benchmark and state of the art analysis. In *Proceedings of the IEEE Conference on computer Vision and Pattern Recognition*, pages 3686–3693, 2014. 3

[4] A. Baldominos, Y. Saez, and C. G. del Pozo. An approach to physical rehabilitation using state-of-the-art virtual reality and motion tracking technologies. *Procedia Computer Science*, 64:10–16, 2015. 2

[5] R. W. Bohannon. Physical rehabilitation in neurologic diseases. *Current Opinion in Neurology*, 6(5):765–772, 1993. 2

[6] J. W. Burke, M. McNeill, D. K. Charles, P. J. Morrow, J. H. Crosbie, and S. M. McDonough. Optimising engagement for stroke rehabilitation using serious games. *The Visual Computer*, 25(12):1085, 2009. 2

[7] H. Chen, Q. Wang, and L. Cao. Design of the workstation for hand rehabilitation based on data glove. In *Bioinformatics and Biomedicine Workshops (BIBMW), 2010 IEEE International Conference on*, pages 769–771. IEEE, 2010. 2

[8] J. Chhor, Y. Gong, and P.-L. P. Rau. Breakout: Design and evaluation of a serious game for health employing intel realsense. In *International Conference on Cross-Cultural Design*, pages 531–545. Springer, 2017. 2

[9] A. Da Gama, P. Fallavollita, V. Teichrieb, and N. Navab. Motor rehabilitation using Kinect: a systematic review. *Games for health journal*, 4(2):123–135, 2015. 2

[10] J. E. Deutsch, D. Robbins, J. Morrison, and P. G. Bowlby. Wii-based compared to standard of care balance and mobility rehabilitation for two individuals post-stroke. In *Virtual Rehabilitation International Conference, 2009*, pages 117–120. Ieee, 2009. 2

[11] O. Ferche, A. Moldoveanu, and F. Moldoveanu. Evaluating lightweight optical hand tracking for virtual reality rehabilitation. *Romanian Journal of Human-Computer Interaction*, 9(2):85, 2016. 2

[12] O. Gargin, O. Pizzi, et al. Wii-hab: Using the wii video game system as an occupational therapy intervention with patients in the hospital setting. *Population Health Matters (Formerly Health Policy Newsletter)*, 23(1):4, 2010. 2

[13] S. N. Gieser, V. Metsis, and F. Makedon. Quantitative evaluation of the Kinect skeleton tracker for physical rehabilitation exercises. In *Proceedings of the 7th International Conference on Pervasive Technologies Related to Assistive Environments*, pages 48:1–48:4, 2014. 2

[14] S. Hellweg and S. Johannes. Physiotherapy after traumatic brain injury: a systematic review of the literature. *Brain Injury*, 22(5):365–373, 2008. 2

[15] C. Ionescu, D. Papava, V. Olaru, and C. Sminchisescu. Human3. 6m: Large scale datasets and predictive methods for 3d human sensing in natural environments. *IEEE transactions on pattern analysis and machine intelligence*, 36(7):1325–1339, 2013. 3

[16] C. Kaminer, K. LeBras, J. McCall, T. Phan, P. Naud, M. Teodorescu, and S. Kurniawan. An immersive physical therapy game for stroke survivors. In *Proceedings of the 16th international ACM SIGACCESS conference on Computers & accessibility*, pages 299–300. ACM, 2014. 2

[17] R. Komatireddy, A. Chokshi, J. Basnett, M. Casale, D. Goble, and T. Shubert. Quality and quantity of rehabilitation exercises delivered by a 3-D motion controlled camera: A pilot study. *International Journal of Physical Medicine & Rehabilitation*, 2(4), 2014. 2

[18] B. Lange, C.-Y. Chang, E. Suma, B. Newman, A. S. Rizzo, and M. Bolas. Development and evaluation of low cost game-based balance rehabilitation tool using the Microsoft Kinect sensor. In *33rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC*, pages 1831–1834, 2011. 2

[19] B. Lange, S. Flynn, and A. Rizzo. Initial usability assessment of off-the-shelf video game consoles for clinical game-based motor rehabilitation. *Physical Therapy Reviews*, 14(5):355–363, 2009. 2

[20] B. Lange, S. Koenig, E. McConnell, C.-Y. Chang, R. Juang, E. Suma, M. Bolas, and A. Rizzo. Interactive game-based rehabilitation using the microsoft kinect. In *2012 IEEE Virtual Reality Workshops (VRW)*, pages 171–172. IEEE, 2012. 2

[21] M. Leo, A. Furnari, G. G. Medioni, M. Trivedi, and G. M. Farinella. Deep learning for assistive computer vision. In *The European Conference on Computer Vision (ECCV) Workshops*, September 2018. 2

[22] M. Leo, G. Medioni, M. Trivedi, T. Kanade, and G. M. Farinella. Computer vision for assistive technologies. *Computer Vision and Image Understanding*, 154:1–15, 2017. 2

[23] W.-w. Liao, S. McCombe Waller, and J. Whitall. Kinect-based individualized upper extremity rehabilitation is effective and feasible for individuals with stroke using a transition from clinic to home protocol. *Cogent Medicine*, 5(1), 2018. 12 pages. 2

[24] J. Liu, H. Ding, A. Shahroudy, L.-Y. Duan, X. Jiang, G. Wang, and A. C. Kot. Feature boosting network for 3D pose estimation. *arXiv preprint arXiv:1901.04877*, 2019. 2

[25] A. Mathis, P. Mamidanna, K. M. Cury, T. Abe, V. N. Murthy, M. W. Mathis, and M. Bethge. DeepLabCut: Markerless pose estimation of user-defined body parts with deep learning. *Nature Neuroscience*, 21:1281–1289, 2018. 2

[26] R. Mehrizi, X. Peng, Z. Tang, X. Xu, D. Metaxas, and K. Li. Toward marker-free 3D pose estimation in lifting: A deep multi-view solution. In *The 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, pages 485–491, 2018. 2

[27] D. Mehta, H. Rhodin, D. Casas, P. Fua, O. Sotnychenko, W. Xu, and C. Theobalt. Monocular 3d human pose estimation in the wild using improved cnn supervision. In *2017 International Conference on 3D Vision (3DV)*, pages 506–516. IEEE, 2017. 2, 3, 4

[28] D. Mehta, S. Sridhar, O. Sotnychenko, H. Rhodin, M. Shafiei, H.-P. Seidel, W. Xu, D. Casas, and C. Theobalt. VNect: Real-time 3D human pose estimation with a single RGB camera. *ACM Transactions on Graphics (TOG)*, 36(4):44:1–44:14, 2017. 2

[29] 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*, 2014:846514, 2014. 2

[30] A. Newell, K. Yang, and J. Deng. Stacked hourglass networks for human pose estimation. In *European conference on computer vision*, pages 483–499. Springer, 2016. 2

[31] S. Pandit, S. Tran, Y. Gu, E. Saraee, F. Jansen, S. Singh, S. Cao, A. Sadeghi, E. Shandelman, T. Ellis, and M. Betke. Exercisecheck: A scalable platform for remote physical therapy deployed as a hybrid desktop and web application. In *Proceedings of the 12th ACM International Conference on PErvasive Technologies Related to Assistive Environments*, PETRA '19, pages 101–109, New York, NY, USA, 2019. ACM. 1, 4

[32] T. Pfister, J. Charles, and A. Zisserman. Flowing convnets for human pose estimation in videos. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1913–1921, 2015. 2

[33] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen. Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4510–4520, 2018. 3, 4

[34] E. Saraee, Y. Gu, S. Pandit, S. Tran, E. Shandelman, S. Singh, T. J. Nordahl, T. Ellis, and M. Betke. Exercisecheck: Data analytics for a remote monitoring and evaluation platform for home-based physical therapy. In *Proceedings of the 12th ACM International Conference on PErvasive Technologies Related to Assistive Environments*, PETRA '19, pages 110–118, New York, NY, USA, 2019. ACM. 1, 4

[35] E. Saraee, S. Singh, K. Hendron, M. Zheng, A. Joshi, T. Ellis, and M. Betke. Exercisecheck: remote monitoring and evaluation platform for home based physical therapy. In *Proceedings of the 10th International Conference on PErvasive Technologies Related to Assistive Environments*, pages 87–90. ACM, 2017. 4

[36] C. Schönauer, T. Pintaric, and H. Kaufmann. Full body interaction for serious games in motor rehabilitation. In *Proceedings of the 2nd Augmented Human International Conference*, page 4. ACM, 2011. 2

[37] L. E. Sucar, R. Luis, R. Leder, J. Hernández, and I. Sánchez. Gesture therapy: A vision-based system for upper extremity stroke rehabilitation. In *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*, pages 3690–3693. IEEE, 2010. 2

[38] D. Tome, C. Russell, and L. Agapito. Lifting from the deep: Convolutional 3d pose estimation from a single image. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2500–2509, 2017. 2, 3, 4

[39] A. Toshev and C. Szegedy. Deeppose: Human pose estimation via deep neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1653–1660, 2014. 2

[40] G. Ulm. The current significance of physiotherapeutic measures in the treatment of parkinson's disease. *Journal of neural transmission. Supplementum*, 46:455–460, 1995. 2

[41] A. Vakanski, J. Ferguson, and S. Lee. Mathematical modeling and evaluation of human motions in physical therapy using mixture density neural networks. *Journal of Physiotherapy & Physical Rehabilitation*, 1(4), 2016. 2

[42] M. Vasileiadis, C.-S. Bouganis, and D. Tzovaras. Multi-person 3d pose estimation from 3d cloud data using 3d convolutional neural networks. *Computer Vision and Image Understanding*, 2019. 2

[43] J. Venugopalan, C. Cheng, T. H. Stokes, and M. D. Wang. Kinect-based rehabilitation system for patients with traumatic brain injury. In *35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 4625–4628, 2013. 2

[44] D. Webster and O. Celik. Systematic review of Kinect applications in elderly care and stroke rehabilitation. *Journal of Neuroengineering and Rehabilitation*, 11(1):108, 2014. 2

[45] Y.-C. Wu, T.-Y. Wu, P. Taele, B. Wang, J.-Y. Liu, P.-s. Ku, P.-E. Lai, and M. Y. Chen. Activeergo: Automatic and personalized ergonomics using self-actuating furniture. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, page 558. ACM, 2018. 2

[46] B. Xiao, H. Wu, and Y. Wei. Simple baselines for human pose estimation and tracking. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 466–481, 2018. 2, 3, 4

[47] H. Yamaura, K. Matsushita, R. Kato, and H. Yokoi. Development of hand rehabilitation system for paralysis patient–universal design using wire-driven mechanism–. In *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*, pages 7122–7125. IEEE, 2009. 2

[48] W. Yang, S. Li, W. Ouyang, H. Li, and X. Wang. Learning feature pyramids for human pose estimation. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1281–1290, 2017. 2

[49] J. Zariffa and J. D. Steeves. Computer vision-based classification of hand grip variations in neurorehabilitation. In *Rehabilitation Robotics (ICORR), 2011 IEEE International Conference on*, pages 1–4. IEEE, 2011. 2

[50] X. Zhou, Q. Huang, X. Sun, X. Xue, and Y. Wei. Towards 3d human pose estimation in the wild: a weakly-supervised approach. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 398–407, 2017. 2