

**Virtual Reality manuscript No.**  
(will be inserted by the editor)

---

## **EVA: EVAluating at-home rehabilitation exercises using augmented reality and low cost sensors**

**Felix Escalona · Ester Martinez-Martin ·  
Edmanuel Cruz · Miguel Cazorla ·  
Francisco Gomez-Donoso**

Received: date / Accepted: date

**Abstract** Over one billion people in the world live with some form of disability. This is incessantly increasing due to aging population and chronic diseases. Among the emerging social needs, rehabilitation services are the most required. However, they are scarce and expensive what considerably limits access to them. In this paper, we propose EVA, an augmented reality platform to engage and supervise rehabilitation sessions at home using low-cost sensors. It also stores the user's statistics and allows therapists to tailor the exercise programs according to their performance. This system has been evaluated in both qualitative and quantitative ways obtaining very promising results.

**Keywords** Rehabilitation Exercises · Deep Learning · Augmented Reality · Human-Computer Interaction · 3D Visualization · Low-Cost Sensors

### **1 Introduction**

According to the World Health Organization (WHO), about 15% of the world's population suffers some form of disability. In addition, this rate is continuously increasing as a result of the society aging and the growth in the prevalence of chronic diseases such as cancer or mental health disorders. This fact has led to a social concern about their health care, especially those people with significant difficulties in functioning.

In this regard, one of the most noteworthy shortcomings is the rehabilitation services since they play a main role in the person's autonomy reinforcement, the decrease in their vulnerability and the improvement in their physical

---

All the authors are with  
University Institute for Computer Research  
Universidad de Alicante, Spain.  
Tel.: +34-965903400  
Fax: +34-965903902  
E-mail: fgomez@dccia.ua.es

condition. Furthermore, a proper recovery prevents an early retirement from work, a considerable decrease in accumulated wealth and/or a reduction of social functions. However, the deficiencies in rehabilitation services, their cost and their long duration demand the development of technology supporting this process at home.

In this work, we propose an augmented reality system to perform and evaluate rehabilitation exercises at home. This system is aimed at two kind of people: patients who require rehabilitation at home after an injury, and the elderly people. Therefore, our purpose is to help them to recover from their affections and, consequently, to improve their quality of life. Our proposal consists of spawning a personal trainer on the patient's home by taking advantage of the augmented reality methods. The user is able to watch the personal trainer and carry out the exercises by imitating him in real time, just like if they were in an actual gym. Upon the end of an exercise, the system automatically grades the patient's performance taking into account the similarity between the trainer's movements and the patient's ones. The exercise sessions are recorded such that the patient and the therapist could review them anytime in order to improve their performance and know the patient's health status at any time. In addition, low-cost sensors like regular color cameras are used, making the system easily affordable.

We conceived the approach as a cloud-based service. The heavy computation part of the system is carried out in remote servers that are maintained by the entity that offers the service. Namely, the government, hospitals, clinics or retirement homes for instance. The final user only needs a low-end terminal like an embedded device or a tablet/smartphone. Despite the computation power requirements are high in the server side, a single machine could render service to several clients.

Summarizing, the main contributions of this work are:

- A low-cost AR rehabilitation app which successfully integrates different deep learning methods
- Our system enables a remote rehabilitation, making therapy accessible to users who would otherwise not have access. This feature is desirable as stated by the WHO [World Health Organization, 2015]
- In contrast with other solutions, ours is non-intrusive. It relies only on vision algorithms
- A system usability scale study that validates the benefits of the app
- Our system integrates a *reference mat*, who shows how to do any required movement. This is crucial to guide the patient when at-home rehabilitation takes place and the therapist is not present

The rest of the paper is organized as follows: First, the state-of the art in the field is presented in Section 2. Next, Section 3 describes EVA workflow. Then, Section 4 explains the augmented reality application as well as the user interface. The considered rehabilitation exercises are detailed in Section 5. The procedures for testing the proposed approach and their results are

presented in Section 6. Finally, Section 7 includes the discussion, conclusions and limitations of the work.

## 2 Related Work

The increase of disabled people together with a growing demand for rehabilitation worldwide require new ways to reduce the cost of the rehabilitation process while maintaining its quality. In this context, assistive technologies play an important role in functioning and increasing independence and participation [World Health Organization, 2010]. However, literature has mainly focused on the physical therapy.

In this line, Robotics has been a very active research area, going from companion robots for therapy to social assistive robots [Martinez-Martin and del Pobil, 2017]. So, rehabilitation robots are designed to assist people recovery in two different scenarios. On the one hand, robot systems can be used as a support tool in the rehabilitation process. This is the case of the Lokomat [Hocoma, 2018], an exoskeletal robot for physiological gait rehabilitation; NeReBot (NEuro REhabilitation roBOT) [Stefano et al., 2014], a cable-suspended device for upper limb rehabilitation of post-stroke patients; or, AMADEO [Tyromotion, 2018], a neurological rehabilitation device designed for the rehabilitation of the hand, fingers and thumb.

On the other hand, social autonomous robots could supervise the rehabilitation at home, a treatment appropriate for both people who cannot travel easily and those who require less care. From this starting point, for instance, Gomez-Donoso et al. [Gomez-Donoso et al., 2017] developed a multisensor system for rehabilitation and interaction with persons with motor and cognitive disabilities. More recently, Costa et al. [Costa et al.] proposed PHAROS, an interactive robot system that recommends and monitors physical exercises at home designed for staying active, an important part of rehabilitation for chronic diseases.

Nonetheless, robotic solutions are far from being affordable, particularly in some low- and middle- income countries. In this regard, a virtual environment may be a reasonable substitute as shown in [Levy-Tzedek et al., 2017]. Actually, Virtual Reality (VR) has been used for evaluating and treating a number of pathologies. We have to note that the term VR nowadays implies the use of immersive technologies (like

So, Jack et al. presented a virtual reality system for rehabilitating hand function in stroke patients [Jack et al., 2001]. Four rehabilitation routines for hand recovery could be carried out: range, speed, fractionation or strength. For that, two different input devices must be worn: a Cyber-Glove and a Rutgers Master II-ND (RMII) force feedback glove.

Steffen et al. proposed a home-based platform for physical activity supervision and motivation [Steffen et al., 2013]. In particular, its system PAMAP (Physical Activity Monitoring for Aging People) uses the television as an interface for two applications: to set the exercises to be done (defined by a

healthcare professional) and to provide feedback to the person. It is noteworthy that a group of sensors capturing person's motion is required to be able to provide any information about their performance. Given that those sensors must be worn in strategic body positions, the person's performance could not be properly measured. In addition, a training phase is required to adjust the exercise performance evaluation to the person's physical limitations.

A more complex system is proposed by Toyra [Indra, 2019], a rehabilitation platform integrating healthcare information technology, virtual reality and motion capture to develop tailored interactive therapy exercises for upper limb recovery. As previously, several sensors should be worn to properly measure the patient's motion.

Virtual games can also be used in this context. Among them, Wii Fit [Nintendo, 2008] is highlighted since it has been applied in several scenarios such as Parkinson's disease [dos Santos Mendesa et al., 2012], knee replacement [Fung et al., 2012], stroke rehabilitation [Barcala et al., 2013] or balance recovery [Meldrum et al., 2012].

Another technology widely studied in the rehabilitation field is Augmented Reality (AR). In fact, AR provides the user with a better sense of presence and reality judgements of the environment as the interaction elements are real [Al-Issa et al., 2013]. This is the underlying idea of NeuroR [de Assis et al., 2014], an AR system for motor rehabilitation of chronic stroke patients that replaces the paralyzed arm in a virtual avatar. **However, this kind of systems rely on privative technologies, such as the devices Apple Iphone or Apple Ipad that come with the Apple ARKit (Augmented Reality Kit), specialized hardware like Simblee, bio-signal sensors such as electrocardiography (ECG) or electromyography (EMG) or positional sensors such as inertial measurement unit (IMU).** This is the case of the AR systems for upper-limb rehabilitation presented by Aung et al., i.e. ARIS (Augmented Reality based Illusion System) [Aung et al., 2014] and RehaBio [Aung and Al-Jumaily, 2014]. These systems combine a visual illusory environment with EMG signal (electromyography signal) to monitor the user's performance. Similarly, [Gazzoni and Cerone, 2018] used AR together with surface EMG (sEMG) detection/acquisition systems for physical rehabilitation. More recently, Monge & Postolache [Monge and Postolache, 2018] combined AR serious games with a wearable sensor network based on *Simblee*, **an Arduino-based programmable board with a wearable design**, for physical rehabilitation of lower limb. Although the experimental results are very promising, the system requires an iPhone, what is not affordable for all the people.

Alternatively, SleeveAR [Sousa et al., 2016] is an AR system to perform rehabilitation exercises at home to complement in-clinic physical therapy of an injured arm. For that, three different stages take place: *the Recording Stage*, involves the demonstration of the exercise being recorded by the therapist; *the Movement Guidance Stage*, focuses on guiding the patient to recreate the prescribed exercise as previously recorded; and, *the Performance Review Stage*, provides the patient with an overview of their performance, by comparing

with the original prescribed exercise. Note that this is a stationary system that requires a special covering floor and a custom sleeve, what considerably restricts its use. On its behalf, Desai et al. [Desai et al., 2016] used low-cost RGB-D cameras (i.e. Microsoft Kinect) to place the user into an AR scene for exercising by using Mirror therapy. Their pilot experiments result in a positive reinforcement. However, the set of exercises is very reduced (only 4 exercises) what considerably restricts its use. In a similar way, SilverFit [SilverFit, 2019] employs a 3D camera to register the user’s movements and links them to a game. At the end, the game progress and the final results are displayed on the screen. It could be said that this system is the most similar to our approach since visual data is used to analyze the user’s performance that can be recorded and reviewed afterwards. Despite the user progress and score is shown, there is no way to know the errors made except with a session with the therapist.

Note that all the proposed approaches have three main handicaps. The first one is the used of worn devices, what may considerably affect the patient’s evaluation apart from the inconvenience caused. The second one is their cost what makes them unreachable for most of the disabled people. Thus, with the aim of overcoming these issues, we present a low-cost **augmented** reality system to monitor and evaluate physical rehabilitation at home. Finally, they do not include any *reference person* showing them how to do a specific exercise or a precise pose.

### 3 Proposal

In this work we propose an augmented reality system to perform and evaluate at-home rehabilitation exercises. This system is aimed at two kinds of people: patients requiring at-home rehabilitation after an injury, and elderly citizens. Thus, our purpose is aiding them to recover the mobility of the affected limbs and, consequently, to improve their quality of life.

With that aim, our system comprehends several rehabilitation exercises that are displayed in a television, computer screen or projector. These exercises are performed by a virtual personal trainer that appears in the patient’s room through augmented reality. So, the display acts as the mirror in a gym: the user is watching himself aside the virtual personal trainer. Then, the user must choose an exercise and mimic the virtual trainer movements. After each exercise, a score is given so that the patient has immediate feedback about their performance.

In rough lines, the system’s workflow can be summarized as follows (see Fig. 1): when an exercise is running, the system first takes an image of the scene using a regular camera. Then, it looks for *“the trainer’s mat”*. This is an object that is used to estimate the floor plane and to set a common 3D reference frame. Once the floor plane is detected, a person detector is then used to extract the person’s position within the image. The area of interest corresponding to the person is sent to the human 3D pose estimator that returns the 3D pose of the person. After that, the virtual trainer and the user

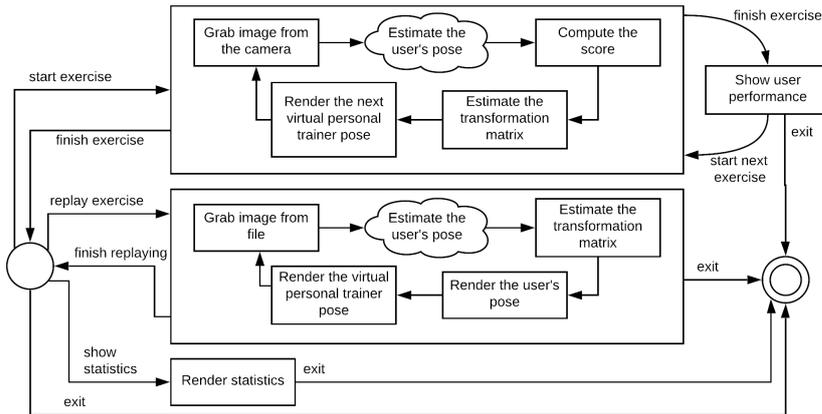


Fig. 1: Flowchart of EVA system. Note that the cloud-based parts are shown in a cloud shape.

interface are rendered over the image to be properly displayed on the screen. When the exercise is completed, the stored 3D human poses and the virtual trainer's poses are used to measure the user's performance. This measure is displayed to the user together with their feedback, and is also stored for further analysis.

It is worth noting that the virtual trainer performing the exercises has been recorded beforehand and the corresponding 3D poses were also estimated offline. The estimation of the trainer's poses is computed following the same method we used for the user's ones. In this way, the system is only rendering the trainer's poses like a video.

As above mentioned, each exercise execution is stored allowing the user or the therapists to replay their exercises to spot mistakes and further improve their performance. By taking advantage of augmented reality, the user can show the replays in the room's floor or even in their desk. Furthermore, our system EVA features a web service that accepts an image stream and renders on it the replay of an exercise session. The web server follows the client-server methodology to provide remote access to the capabilities of the system. This option could be used to build a smartphone and/or tablet application to show the replays anywhere and anytime by connecting with the system via WiFi or even 4G. Note that the scores are also stored with the purpose to provide user's statistics to both users and therapists.

Additionally, EVA has an easy-to-use, friendly user interface based on virtual touch buttons leveraging augmented reality.

### 3.1 Common 3D Coordinate Frame Estimation

As above mentioned, the common 3D coordinate frame is estimated by detecting "the trainer's mat". For that, a well-known chessboard pattern has been used (see Fig. 2). It must be placed on the floor away from the user. In this way, the common 3D coordinate frame allows EVA to accurately detect the floor plane and properly render the avatar of the virtual personal trainer in the scene. As a result, the user senses the virtual trainer next to them in the room.



Fig. 2: A chessboard pattern is used as "the trainer's mat". It is a well-known pattern that allows the system to estimate a common 3D coordinate frame. The leftmost image depicts the pattern with the 3D axis superimposed. The rightmost image depicts the personal trainer (skeleton in green) in augmented reality in front of the patient who is imitating them. The personal trainer is rendered *in situ* thanks to the pattern's detection.

Although some state-of-the-art approaches are able to detect the floor plane without placing any pattern within a scene (e.g. [Yang et al., 2016, Mur-Artal and Tardos, 2017]), their lack of accuracy, the high complexity of their calibration procedure or the need of non-static cameras make them unsuitable for the problem at hand. In addition, the presence of a physical element in the room prevents the user to inadvertently trespass the trainer *personal* space, what would negatively affect the perception of the augmented reality.

With the aim to detect the floor plane and to set a common coordinate frame, EVA firstly looks for the intersections between the pattern squares in the image. As a result, a set of 2D points on the image plane is obtained. Then, as the pattern is known, the corresponding 3D points can be easily stated. The points lay in a plane so the Z-value is 0. Then, the length of each square side is also known. Therefore, we can easily compute a set of corresponding 3D points for each 2D point detected in the color image. Finally, we solve the Perspective-n-Point problem. As a result, a transformation matrix is obtained. This matrix transforms a point from the 3D world coordinate frame to the 2D coordinate frame of the image.

Finally, it is worth noting that the intrinsic camera parameters are required to solve the Perspective-n-Point problem and, as a consequence, the camera must be calibrated. This calibration process is carried out offline.

As the personal trainer poses are stated in the 3D space, they can be easily translated into the image coordinate frame from the transformation matrix computed in this stage. This helps to achieve the illusion of the personal trainer being actually aside the user in their own living room. Furthermore, the trainer is scaled according to the user’s size. This virtual scaled view is a useful feature since it adjusts the taken video recordings to the chessboard pattern size. So, for instance, the therapist could reproduce any exercise session by means of a reduced chessboard pattern on its desk.

### 3.2 Human 3D Pose Estimation from Monocular Frames

With the aim to compare the user’s and the trainer’s movements in a robust way, a 3D pose is mandatory. A solution could be the estimation of the 2D human poses since low processing load techniques can be found in the literature [Cao et al., 2017]. However, the 2D poses may cause ambiguities and singularities between two poses. This fact could lead to a bad similarity estimation. This is the case depicted in Fig. 3, where it is impossible to state whether the user is rising his arm forward or backward.

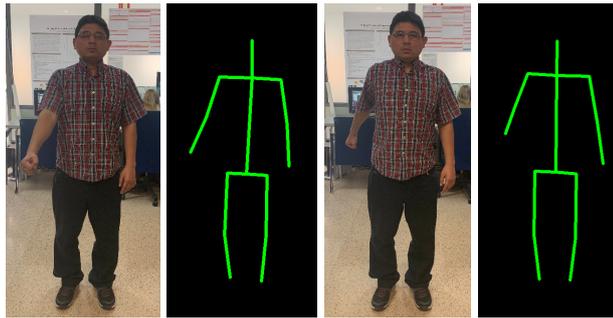


Fig. 3: The 2D poses cannot be used for a robust pose comparison between the user and the trainer because it may eventually fall into a singularity or ambiguous poses such as the depicted in this figure. According to the 2D pose, it is impossible to state whether the patient is rising his arm forward or backward.

Due to the required accuracy in human pose comparison, a 3D pose estimation approach has been used. Our 3D human pose estimation system is based on the Human Mesh Recovery (HMR) approach [Kanazawa et al., 2018]. This method consists of an end-to-end framework for reconstructing a full 3D mesh of a human body from a monocular *RGB* image. So, a deep learning-based

encoder is able to predict the camera pose, the person’s shape and the person’s pose for each taken image. These predictions are used to render a model which is then validated by a discriminator. The discriminator is able to state if the predictions correspond to a real person or not. Given that this approach does not use a 2D intermediate representation being able to make final predictions in one forward step, it is very fast. Note that EVA only works on the inferred 3D human pose, discarding the camera pose and the person’s shape. This 3D human pose is expressed as a list of 3D points in camera coordinates, corresponding to 19 joints of the human body.



Fig. 4: . The 3D human pose estimation system is able to accurately estimate the human pose even in difficult cases such as persons in wheelchair or scenarios with high self-occlusion level caused by the exercises.

This approach for 3D human pose estimation is able to accurately estimate the human pose even under different scenarios as illustrated in Fig. 4. For instance, the system works well for persons sitting in wheelchairs. It can also deal with high levels of self-occlusion. These cases play a main role when working with physical injured people and elderly.

Nonetheless, this approach has an important issue to be pointed out. The best human pose estimations are obtained when there is only one person in the image. However, our system captures images that includes the trainer’s mat and the room elements apart from the patient. For that reason, a person detection stage is needed. To carry out this process, we leveraged YOLOv3 [Redmon and Farhadi, 2018]. It is a well-known method for object detection and recognition, providing a decent accuracy while keeping the computation cost at bay. This region convolutional neural network architecture is able to detect the position of the objects in the image plane, estimate the label of those objects as well as their corresponding confidence score. This architecture achieved a 0.51 mAP (measured over the intersection over union) on the test set of the

COCO MS dataset [Lin et al., 2014]. From this pretrained model, a wide range of objects including persons, can be detected. Fig. 5 shows the performance of this architecture for person detection.



Fig. 5: The person detector YOLOv3 [Redmon and Farhadi, 2018] is able to accurately estimate the area of interest enclosing the patient.

So, the complete human 3D pose estimation pipeline can be described as follows: first, an image is grabbed from the camera. Then, this image feeds the YOLOv3 detector, which provides the area of interest of several objects. From them, the patient is identified as the area of interest with the best confidence score for person. This area of interest is cropped from the original image and forwarded to the HMR network. As a result, the 3D pose of the patient is obtained. Fig. 6 shows the performance of the HMR network feeding it on the whole image and on the person area detected by YOLOv3. As it can be observed, the predictions are more precise when the network is fed with just the person area. It is worth mentioning that we use the models released by the original authors from both HMR and YOLOv3 methods.

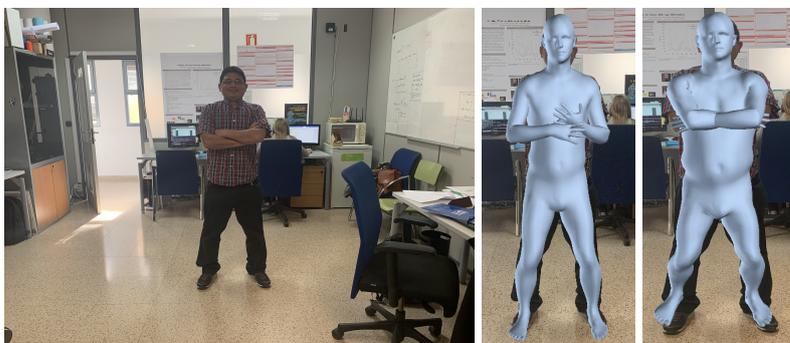


Fig. 6: Performance of the Human Mesh Recovery network feeding it the whole image (center) and the person area of interest detected by YOLOv3 (right-most).

### 3.3 Scoring the User's Performance

With the purpose to evaluate the user's performance during rehabilitation exercises, we propose to compare the user's joints trajectory with the trainer's ones by using a Dynamic Time Warping (DTW) approach [Berndt and Clifford, 1994]. This state-of-the-art method is aimed to find patterns in time series. So, it finds a warping path to align the elements of two time sequences such that the distance between them is minimized. This distance between two elements must be defined, being the euclidean distance the most common. We can see the dynamic time warping problem as a minimization of the cumulative distance over the whole possible paths of two time series elements.

This method provides a score quantifying the degree of adjustment of two times series when stretching or compressing their elements along the time. Note that this score stays in the range  $[0, 1]$  when comparing different series. For that, the distance measurement is often modified to be relative to a baseline distance.

Given that the time dimension is not considered, this method is appropriate to compare two 3D human poses separated by a short period of time as it is the case (i.e. the user may take time to imitate the trainer's movements due to their disability).

Assuming that different patients use different cameras at different places, it is necessary to define a common reference frame to avoid the influence of these factors in the comparison of the joint positions. We have to transform from the camera frame, defined by 3 camera axis ( $c$ ) and the origin of coordinates ( $Q_0$ ) in Equation 1, to a new invariant reference frame. This reference frame is defined using the user's body in a preset position. The reference point ( $P_0$ ) of this frame is the neck (joint 1 in the Fig. 7). Then, we build two perpendicular vectors ( $v_0$  and  $v_1$ ) from shoulder center to the head (joint 0) and to the right shoulder (joint 2) respectively. The last axis vector ( $v_2$ ) is calculated applying the cross product of the previous vectors, as shown in Equation 2.

$$\begin{aligned}
 \text{Cameraframe} &: (c_0, c_1, c_2, Q_0) \\
 c_0 &= (0, 0, 1) \\
 c_1 &= (0, 1, 0) \\
 c_2 &= (1, 0, 0) \\
 Q_0 &= (0, 0, 0)
 \end{aligned} \tag{1}$$

$$\begin{aligned}
 \text{Bodyframe} &: (v_0, v_1, v_2, P_0) \\
 v_0 &= (j0.x - j1.x, j0.y - j1.y, j0.z - j1.z) \\
 v_1 &= (j2.x - j1.x, j2.y - j1.y, j2.z - j1.z) \\
 v_2 &= v_0 \times v_1 \\
 P_0 &= (j1.x, j1.y, j1.z)
 \end{aligned} \tag{2}$$

Once we have defined the reference frame centered in the body, we have to estimate the transformation matrix from the camera's frame to this reference



Fig. 7: Joint positions considered by the system as returned by the Human 3D Pose Estimation system and used to compare the human poses of the user and the virtual trainer.

frame. The transformation matrix is estimated from the Equation 3, that calculates the coefficients necessary to put the new basis vectors  $v_0$ ,  $v_1$  and  $v_2$  as a linear combination of the camera axis  $c_0$ ,  $c_1$  and  $c_2$  and the new origin of coordinates in terms of these vectors and the camera's origin of coordinates.

$$\begin{aligned}
 v_0 &= a_{11}c_0 + a_{12}c_1 + a_{13}c_2 \\
 v_1 &= a_{21}c_0 + a_{22}c_1 + a_{23}c_2 \\
 v_2 &= a_{31}c_0 + a_{32}c_1 + a_{33}c_2 \\
 P_0 &= a_{41}c_0 + a_{42}c_1 + a_{43}c_2 + Q_0
 \end{aligned} \tag{3}$$

$$M = \begin{bmatrix} a_{11} & a_{12} & a_{13} & 0 \\ a_{21} & a_{22} & a_{23} & 0 \\ a_{31} & a_{32} & a_{33} & 0 \\ a_{41} & a_{42} & a_{43} & 1 \end{bmatrix}$$

With the transformation matrix between the camera and the body reference frames, we can simply transform the points multiplying this matrix with

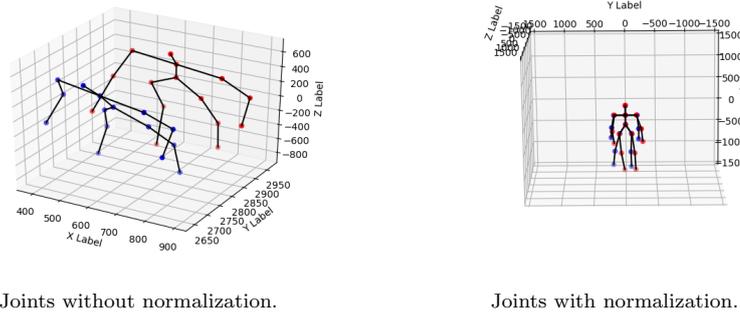
the column vector of homogeneous coordinates of each point. Suppose that  $a$  and  $b$  are the homogeneous representation of the same point on the camera and the body reference frames respectively, so they satisfy the Equation 4.

$$\begin{aligned} a &= M^T b \\ b &= (M^T)^{-1} a \end{aligned} \quad (4)$$

Then, we transform the whole points into the body reference frame and calculate the distance between two sequences of movements for every pair of joints using FastDTW, an accurate and efficient DTW implementation presented in [Salvador and Chan, 2007]. So, the final distance is the average of the accumulated distance of the joints.

Due to the fact that the final distance is estimated for every joint individually, we can report to the user which are the joints they are performing worse. In addition, the DTW is computed every 120 frames with an offset of 60 frames. In this way, we can even know which moment of the rehabilitation session has been performed accurately and which has not.

Fig. 8 depicts a representation of joints in body coordinates for two different bodies. The left image shows the body joints without coordinate normalization, while the right image illustrates both body joints after coordinate normalization.



Joints without normalization.

Joints with normalization.

Fig. 8: Representation of joints in body coordinates for two different bodies.

#### 4 Augmented Reality Application

The browsing within the menus of the developed application is made through body motion. The user must lay the hand on a button to trigger the corresponding action. For that, the Human 3D Pose Estimation pipeline described in Section 3.2 has been used. So, the estimated 3D position of the right hand

is translated into the image plane. If this point lays on a button, this button is selected (see Fig. 9, where the user is selecting the *Replay* button in the EVA's main menu). Note that the user must lay the hand on a button for 5 seconds to trigger the corresponding action. This prevents from inadvertently activating a button. In addition, the user interface distinguishes the two actions (i.e. selection and trigger) such that an arrow marks the selected button, whilst a progress circle in the bottom left shows the time left to trigger the corresponding action.

Additionally, a keyboard could be also used for browsing the menus. This feature is useful when the user does not want to perform a rehabilitation session but only replay it on their desktop, or review their statistics. In this way, it is not mandatory to set up the big "trainer's mat" nor to use the body for browsing the application.

It is also worth noting that the user interface was created considering accessibility and usability. It features big and descriptive buttons with high contrast colors so that the use of the application is intuitive and straightforward.



Fig. 9: Screenshot of the main menu of the developed application. The user can select the different options from the menus with its own body by placing the hand behind the buttons during 5 seconds.

#### 4.1 User Interface

The first button of the main menu is "Start". This option triggers a rehabilitation exercise session as depicted in Fig. 10. So, after pushing the button, a list

of exercises are shown where the user should choose the ones to be performed in the same way as before. Once an exercise session is selected, a virtual personal trainer is rendered over the "trainer's mat". The feed of the camera is being displayed so that the user is watching the trainer besides him in an augmented reality fashion. The user must imitate the movements of the virtual trainer upon the completion of the session. A plot in the left corner of the display is continuously showing the user's update score. Finally, a summarized score representing the patient's performance over the whole exercise is shown. In addition, a timeline of the session is also displayed so that the user can review which part of the session they did better and which one needs more work on.

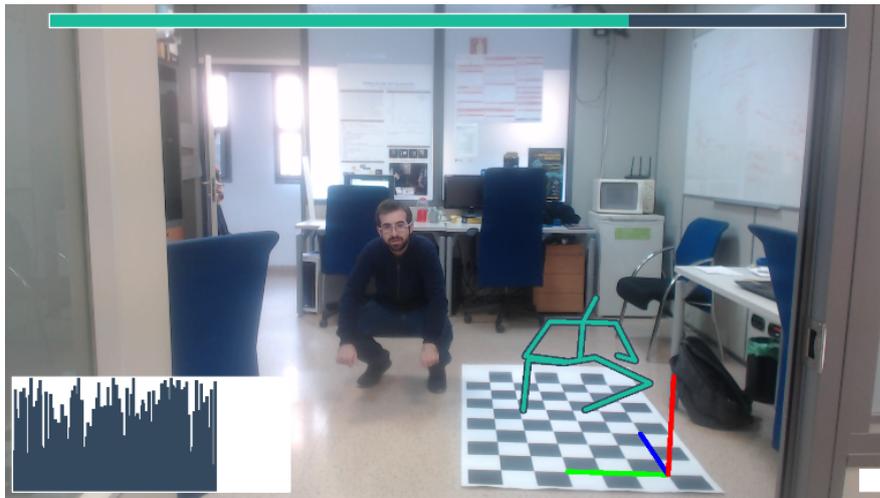


Fig. 10: A sample of a patient performing a rehabilitation session. The virtual trainer is being rendered using augmented reality methods.

The next button of the main menu is "Replay". As above mentioned, this option allows both patient and therapist to replay an exercise session. During the replay, the app shows the positions where the patient stood rendered in an augmented reality video-like way. The replay can be paused such that the therapist or the patient could rotate or zoom in/out a certain pose. This can be done by moving the "trainer's mat". The avatar is customizable so that it can render a person or a skeleton-like character. A score histogram is also displayed in the bottom left corner. This histogram shows the DTW score for each instant of the rehabilitation session. Finally, a progress bar in the top of the window shows how much of the replay is left. Fig. 11 depicts a still of a replay being displayed. Note that, in this case, the therapist is using a reduced size "trainer's mat" allowing the app to render the augmented reality

avatar right on their desk. The replay feature allows an easy analysis of the rehabilitation sessions with the goal of evaluating the patients' performance and modifying the rehabilitation program if needed.

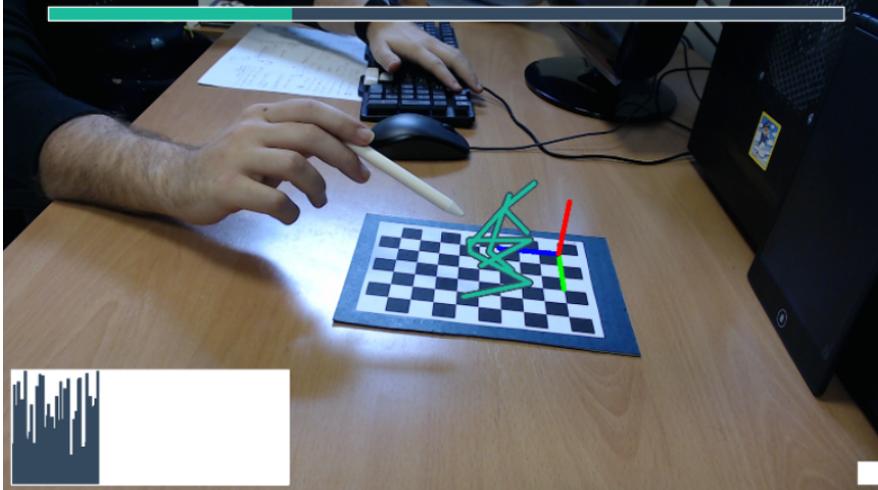


Fig. 11: The image depicts a replay being displayed. Note that, in this case, the therapist is using a reduced size "trainer's mat" that allowed the application to render the augmented reality avatar right on his desk.

The last entry of the main menu is "*Statistics*". This menu shows a new window with the history of the past sessions performed by the user. This window displays a summary of the scores and plots the patient's adherence and performance. These metrics could be used for the patient and the therapist for review and evaluation purposes.

Finally, a very early version of a web service was implemented. This web service allows EVA to be offered as an external service. Therefore, patients can benefit from EVA wherever through low computation powered machines such as tablets, embedded computers and smartphones. In this way, the taken images from the user's terminal are sent to the remote server to be processed. So, the user can get their performance score in real-time independent of the computing power.

## 5 Rehabilitation Exercises

An important step in any rehabilitation treatment is the home exercise program. This program consists in the patient's performance of prescribed physical exercises at home. In this sense, there is a wide variety of exercises de-

Table 1: Activities included in the KARD Gaglio et al. [2015].

ID	Activity	ID	Activity	ID	Activity
1	Horizontal arm wave	7	Draw tick	13	Hand clap
2	High arm wave	8	Toss paper	14	Walk
3	Two hand wave	9	Forward kick	15	Phone call
4	Catch cap	10	Side kick	16	Drink
5	High throw	11	Take umbrella	17	Sit down
6	Draw X	12	Bend	18	Stand up

pending on the body part to be recovered and/or their goal. In particular, a public exercise program suggested by the British National Health Security (NHS) to improve elderly’s fitness and well-being has been considered in this paper [British National Health Security (NHS), 2018]. This guide is composed of twenty-three exercises divided into four groups: flexibility, strength, balance and sitting. However, the therapist is who will decide the exercise program to be followed by each patient. So, the virtual trainer reproduces the tailored set of exercises while analyzes the patient’s evolution in their performance.

## 6 Experimentation and Results

With the aim to validate EVA, two kinds of experiments were carried out. On the one hand, each subsystem has been individually evaluated in a qualitative and quantitative way. On the other hand, the whole system was qualitatively analyzed.

Firstly, the evaluation of the Human 3D Pose Estimation and the Dynamic Time Warping Scoring subsystems were performed by using the Kinect Activity Recognition Dataset (KARD) [Gaglio et al., 2015]. This dataset contains 18 different short-time activities, summarized in Table 1, which overall match the features of some rehabilitation exercises. It involved 10 individuals repeating 3 times each activity. The dataset provides the video of each person performing an activity and the position of the body joints in both 2D and 3D coordinates as captured by a Kinect device. In addition, it provides the corresponding depth maps (discarded in our experiments).

For these experiments the following hardware was used: an Intel Core i5-3570 with 8 GiB of Kingston HyperX 1600 MHz and CL10 DDR3 RAM on an Asus P8H77-M PRO motherboard (Intel H77 chipset). The system also included an Nvidia GTX1080Ti. The framework of choice was Keras 1.2.0 with Tensor Flow 1.8 as the backbone, running on Ubuntu 16.04. CUDA 9.0 and cuDNN v7.1 were used to accelerate the computations.

### 6.1 Human 3D Pose Estimation Experiments

The accuracy of the Human 3D Pose Estimation was evaluated through the KARD dataset. Thus, it was required to divide the activity videos in still frames. As described in Section 3.2, this process involves two different stages. Firstly, each video frame is forwarded to the YOLO architecture which provides the area of interest of the patient within the scene. Then, this area is cropped and used to perform the 3D pose estimation with the HMR approach. As a result, the 3D position of the patient body joints is estimated. Then, the 3D points are projected to the image plane. Some results randomly chosen are shown in Fig. 12. Note that the images were resized to  $224 \times 224$  px since it is the architecture’s input size.



Fig. 12: The top row depicts random samples extracted from the KARD and the bottom row shows the estimated human 3D poses.

Fig. 13 illustrates the amount of samples per distance error threshold. Note that the amount of joints are expressed in percentages. As it can be seen, 92% of the joints yielded an error below 20 px, and 80% below 12 px. The mean error averaged across the entire dataset is 9.58 px for an image of  $224 \times 224$  px resolution. As the results show, this approach is accurate enough to be used for a successful human pose detection from a single RGB frame.

In addition, we also benchmarked the Human 3D Pose Estimation system in the 3D space using the KARD. In this case, we also use YOLO to detect the person in the scene and then forward only the area of interest to

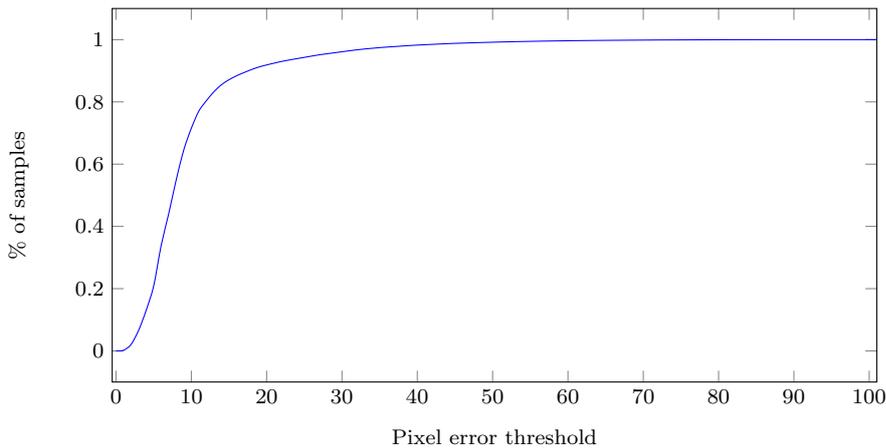


Fig. 13: The accumulated percentage of samples per error (px on a  $224 \times 224$  image) threshold between the 2D points provided by KARD and those estimated by our Human 3D Pose Estimation approach.

the pose estimator. As the 3D pose returned by the system yields no scale, the tridimensional positions of the joints are arbitrary located in the space, yet depicting the correct pose. To enable a fair comparison, we followed the method described in Section 3.3. Thus, we express all the poses in a coordinate frame local to the body pose. Then, as the scale is also different, we computed a scale factor using the norm of the vector that goes from the neck to the right shoulder. This process is applied to both ground truth and estimated poses. As a result, both poses are in the same coordinate frame and yield the same scale. Finally, we measure the mean euclidean distance between each joint averaged across all the frames.

As shown in Fig. 14, the 70.66% of the samples are under the  $150mm$  threshold. Note that this threshold is the same as that used by the authors of the HMR system to report the percentage of correct keypoints. In their proposal, they achieved a 72.9% on the MMPI-INF-3DHP dataset. In addition, we have obtained a mean error of  $129.52mm$ .

These experiments demonstrate the accuracy of the system. Regarding the 2D pose estimation, it can be concluded that it is accurate enough to provide a good representation on the screen, what is important to encourage both users and therapists to use the application. Nonetheless, the 3D poses computed by the system are approximated. Therefore, the obtained evaluation is a reference, but not a medical diagnosis.

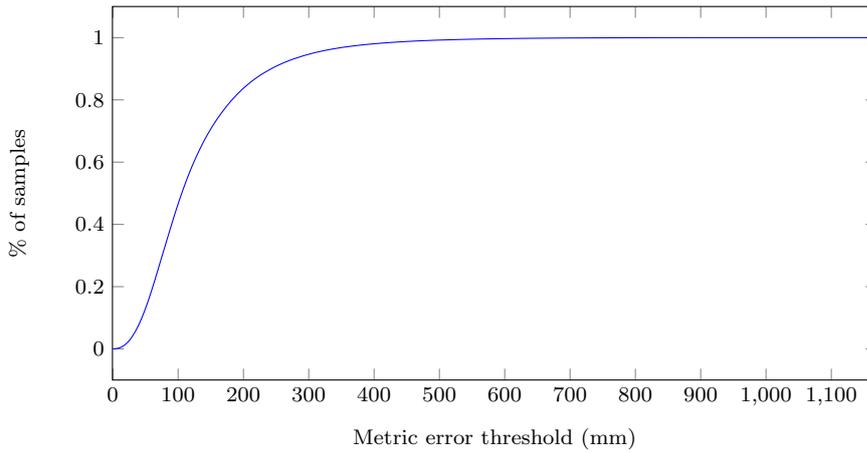


Fig. 14: The accumulated percentage of samples per error (mm) threshold between the 3D points provided by KARD and those estimated by our Human 3D Pose Estimation approach.

## 6.2 Dynamic Time Warping Scoring Experiments

The KARD has been also used to evaluate the effectiveness of the scoring system. As above mentioned, this dataset contains 18 activities and each activity is performed by 10 different subjects. We compare them to each other. The effectiveness of the DTW is determined in the following way: the lower the value obtained, the closer or similar the activity will be. Moreover, the different score will be then compared against another activity. The obtained results are shown in Table 2 where each value indicates the similarity between the activity in the row with respect to that in the column. So, the value in bold represents the closest similarity, while a higher value indicates a higher difference between two activities. As it can be observed, the system is clearly confused between the activity pairs 7 – 5, 11 – 8 and 16 – 5. This is due to two main factors: (1) an overlap of movements (7 – 5, 16 – 5), and (2) the similarity in the activity movements (an object is released in activity 8 whilst an object is grabbed in activity 11).

## 6.3 Integrated System Experiments

The integrated system was qualitatively tested in a controlled environment. Ten different individuals were involved. The experiment consisted in using EVA to perform one exercise per individual and to provide feedback.

For the evaluation of the feedback, the System Usability Scale (SUS) questionnaire proposed by Brooke was used [Brooke, 1996]. This questionnaire is composed of 10 items such that the odd items were written positively, while the even ones were written negatively. So, the SUS questions are as follows:

Table 2: Confusion matrix for the DTW Scoring experiments with the KARD.

	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11	G12	G13	G14	G15	G16	G17	G18
G1	0.19	0.29	0.43	0.28	0.22	0.36	0.32	0.31	0.24	0.23	0.47	0.35	0.26	0.81	0.66	0.32	0.42	0.34
G2	0.29	<b>0.15</b>	0.31	0.25	0.19	0.36	0.29	0.29	0.22	0.22	0.45	0.32	0.25	0.78	0.61	0.33	0.39	0.3
G3	0.43	0.31	<b>0.22</b>	0.34	0.32	0.46	0.41	0.36	0.32	0.35	0.52	0.42	0.36	0.9	0.73	0.46	0.53	0.44
G4	0.28	0.25	0.34	<b>0.12</b>	0.17	0.34	0.29	0.23	0.15	0.19	0.41	0.28	0.19	0.78	0.65	0.34	0.34	0.22
G5	0.22	0.19	0.32	0.17	<b>0.01</b>	0.29	0.21	0.18	0.08	0.09	0.4	0.21	0.11	0.79	0.6	0.25	0.3	0.17
G6	0.36	0.36	0.46	0.34	0.29	<b>0.26</b>	0.35	0.35	0.29	0.32	0.48	0.41	0.32	0.79	0.65	0.41	0.49	0.4
G7	0.32	0.29	0.41	0.29	<b>0.21</b>	0.35	0.23	0.31	0.23	0.24	0.48	0.35	0.26	0.8	0.65	0.36	0.42	0.32
G8	0.31	0.29	0.36	0.23	0.18	0.35	0.31	<b>0.07</b>	0.14	0.19	0.29	0.2	0.19	0.76	0.72	0.39	0.28	0.18
G9	0.24	0.22	0.32	0.15	0.08	0.29	0.23	0.14	<b>0</b>	0.1	0.36	0.19	0.13	0.75	0.65	0.3	0.27	0.13
G10	0.23	0.22	0.35	0.19	0.09	0.32	0.24	0.19	0.1	<b>0.02</b>	0.4	0.22	0.14	0.77	0.64	0.3	0.29	0.17
G11	0.47	0.45	0.52	0.41	0.4	0.48	0.48	<b>0.29</b>	0.36	0.4	0.31	0.4	0.42	0.82	0.79	0.54	0.43	0.38
G12	0.35	0.32	0.42	0.28	0.21	0.41	0.35	0.2	0.19	0.22	0.4	<b>0.13</b>	0.21	0.77	0.74	0.42	0.29	0.2
G13	0.26	<b>1.09</b>	0.36	0.19	0.11	0.32	0.26	0.19	0.13	0.14	0.42	0.21	<b>0.05</b>	0.77	0.65	0.32	0.31	0.19
G14	0.81	0.78	0.9	0.78	0.79	0.79	0.8	0.76	0.75	0.77	0.82	0.77	0.77	<b>0.75</b>	<b>1</b>	0.83	0.83	0.78
G15	0.66	0.61	0.73	0.65	0.6	0.65	0.65	0.72	0.65	0.64	0.79	0.74	0.65	<b>1</b>	0.6	0.65	0.76	0.69
G16	0.32	0.33	0.46	0.34	<b>0.25</b>	0.41	0.36	0.39	0.3	0.3	0.54	0.42	0.32	0.83	0.65	0.26	0.46	0.37
G17	0.42	0.39	0.53	0.34	0.3	0.49	0.42	0.28	0.27	0.29	0.43	0.29	0.31	0.83	<b>0.76</b>	0.46	<b>0.2</b>	0.23
G18	0.34	0.3	0.44	0.22	0.17	0.4	0.32	0.18	0.13	0.17	0.38	0.2	0.19	0.77	0.69	0.37	0.23	<b>0.02</b>

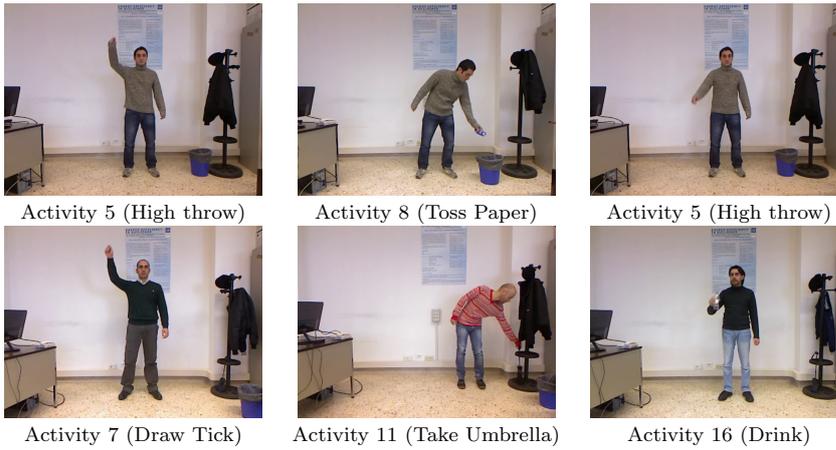


Fig. 15: The first row shows some sample activities from a certain class and the second row depicts different samples from different classes that causes confusion in the system. As the images depict, the confusion is expectable as the classes are very similar despite being labeled as different.

- I think that I would like to use this system frequently.
- I found the system unnecessarily complex.
- I thought the system was easy to use.
- I think that I would need the support of a technical person to be able to use this system.
- I found the various functions in this system were well integrated.
- I thought there was too much inconsistency in this system.
- I would imagine that most people would learn to use this system very quickly.
- I found the system very cumbersome to use.
- I felt very confident using the system.
- I needed to learn a lot of things before I could get going with this system.

To evaluate the results of the SUS test, we used the Adjectives method proposed by Bangor et al. [Bangor et al., 2009]. This method is based on the association of the SUS score with an adjective scale (i.e. worst, poor, ok, good, excellent, best) as illustrated in Fig. 16.

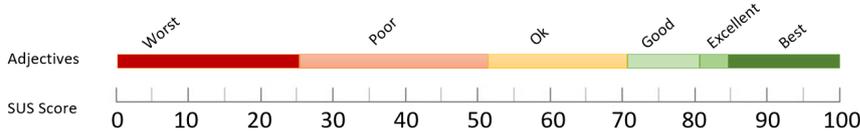


Fig. 16: Adjectives scale method to measure SUS.

As shown in Fig. 17, the total score obtained for each user was positive, obtaining an average of 78.0 (good according to the adjective scale method). This result is significant providing a level of confidence in the use of the system.

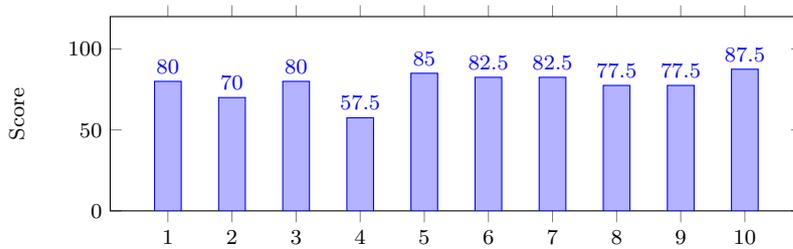


Fig. 17: Score per user obtained by SUS Evaluation.

In addition, some conclusions could be extracted from this study. Participants felt comfortable using the system. Many users commented that the application effectively encouraged them to complete the rehabilitation session and thought that EVA would help improve adherence to these home rehabilitation sessions. Some of the users were reluctant with the presence of the trainer's carpet at first, but then got used to it.

A therapist from ADACEA (A Spanish foundation for the acquired brain damaged) supervised the experiment. He praised EVA by highlighting its potential to improve the adherence to the long-term rehabilitation programs, which he admitted to be one of the weakest points in the patient's rehabilitation. He also found interesting the idea of replaying the exercising sessions in his office or home with the use of augmented reality. Finally, the therapist felt indifferent about the scoring system. He admitted to be an interesting feature to encourage the patients, but he hesitated about the effectiveness to measure

the actual patient’s performance. He also remarked that a professional review is mandatory for the sake of ameliorating the patient’s health.

Table 3: Runtime for each piece of the proposed pipeline. The whole system takes 45.7 ms for computing a frame, which means about 22 fps. The DTW scoring could be performed offline and is independent of the rest of the pipeline.

Subsystem	Time (ms)
Common 3D Coordinate Frame Estimation	10.6
YOLO (person detection)	22.7
HMR (pose estimation)	10.2
Rendering Overhead	2.2
Global Runtime	45.7
DTW Scoring	354.79

Regarding the computation time, EVA is currently able to run at about 22 fps, which is enough to provide real time output. More details about the individual computation time of each piece of the pipeline could be seen in Fig. 3.

## 7 Conclusions and Future Work

In the face of the relentless demand for rehabilitation services and the lack of resources to access them, we propose EVA, an augmented reality application for evaluating rehabilitation programs at home. This low-cost application only requires a regular camera to capture and evaluate the 3D patient’s pose. So, EVA stores the rehabilitation sessions such that patients and therapists could review them and adjust the exercises accordingly. EVA was qualitative and quantitative evaluated to show their suitability. For that, two different types of experiments were carried out. Firstly, each integrating module was individually tested and then a complete evaluation was performed.

Despite the promising results, the experiments brought to light an EVA limitation in terms of human pose estimation. The human pose estimation sometimes fails when people has an amputated limb since it assumes that the images depict a canonical person and it tries to estimate the position of all their joints. This fact will be pointed out in the future since a considerable amount of EVA’s users may present this handicap.

Apart from that, we plan to further explore the web service potential and the idea of a smartphone application. It would not be only able to show replays, but also to provide statistics to the therapists and patients and act as a client as well. We also plan to extend the amount of exercises and to add more customization features to our proposal. For instance, more trainer avatars or new user interface themes to help the visually impaired, will be added. The possibility to add wearable sensors like smartbands will be also explored in order to gather more information about the user’s health status (e.g. ECG or

heart rate). In this way, the therapist could have more information to properly adapt the rehabilitation process. Finally, a native VR/AR environment like Vuforia [PTC, 2019] would be used to leverage and enhance the visualization part.

**Acknowledgements** This work has been supported by the Spanish Government TIN2016-76515R Grant, supported with Feder funds. Edmanuel Cruz is funded by a Panamenian grant for PhD studies IFARHU & SENACYT 270-2016-207. This work has also been supported by a Spanish grant for PhD studies ACIF/2017/243 and FPU16/00887. Thanks also to Nvidia for the generous donation of a Titan Xp and a Quadro P6000.

## References

- World Health Organization. Who global disability action plan 2014-2021. better health for all people with disability. Technical report, 2015.
- World Health Organization. Community-based rehabilitation:cbr guidelines. Technical report, Geneva, 2010.
- Ester Martinez-Martin and Angel P. del Pobil. Personal robot assistants for elderly care: An overview. In A. Costa, V. Julian, and P. Novais, editors, *Personal Assistants: Emerging Computational Technologies*, pages 77–91. Springer, 2017. doi: 10.1007/978-3-319-62530-0\\_5.
- Hocoma. Lokomat. <https://esa.un.org/unpd/wpp/>, 2018.
- Masiero Stefano, Poli Patrizia, Armani Mario, Gregorio Ferlini, Roberto Rizzello, and Giulio Rosati. Robotic upper limb rehabilitation after acute stroke by nerebot: Evaluation of treatment costs. *BioMed Research International*, 2014, 2014. doi: 10.1155/2014/265634.
- Tyromotion. Amadeo. <https://tyromotion.com/en/produkte/amadeo/>, 2018.
- Francisco Gomez-Donoso, Sergio Orts-Escolano, Alberto Garcia-Garcia, Jose Garcia-Rodriguez, John Alejandro Castro-Vargas, Sergiu Ovidiu-Oprea, and Miguel Cazorla. A robotic platform for customized and interactive rehabilitation of persons with disabilities. *Pattern Recognition Letters*, 99:105 – 113, 2017. ISSN 0167-8655. doi: 10.1016/j.patrec.2017.05.027.
- Angelo Costa, Ester Martinez-Martin, Miguel Cazorla, and Vicente Julian. Pharos-physical assistant robot system. *Sensors*, 18:2633. doi: 10.3390/s18082633.
- Shelly Levy-Tzedek, Sigal Berman, Yehuda Stiefel, Ehud Sharlin, James Young, and Daniel Rea. Robotic mirror game for movement rehabilitation. In *2017 International Conference on Virtual Rehabilitation (ICVR)*, pages 1–2, June 2017. doi: 10.1109/ICVR.2017.8007494.
- D. Jack, R. Boian, A. S. Merians, M. Tremaine, G. C. Burdea, S. V. Adamovich, M. Recce, and H. Poizner. Virtual reality-enhanced stroke rehabilitation. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 9(3):308–318, September 2001. ISSN 1534-4320. doi: 10.1109/7333.948460.

- Daniel Steffen, Gabriele Bleser, Markus Weber, Didier Stricker, Laetitia Fradet, and Frédéric Marin. A personalized exercise trainer for the elderly. *Journal of Ambient Intelligence and Smart Environments*, 5:547–562, 2013.
- Indra. Toyra. <https://www.tecnologiasaccesibles.com/es/proyectos/toyra>, 2019.
- Nintendo. Wii fit. <https://www.nintendo.es/Juegos/Wii/Wii-Fit-283894.html>, 2008.
- Felipe Augusto dos Santos Mendesa, Jos Eduardo Pompeua, Alexandra Modenesi Lobo, Keyte Guedes da Silva, Tatiana de Paula Oliveira, Andrea Peterson Zomignani, and Maria Elisa Pimentel Piemonte. Motor learning, retention and transfer after virtual-reality-based training in parkinsons disease effect of motor and cognitive demands of games: a longitudinal, controlled clinical study. *Physiotherapy*, 98:217–223, 2012.
- Vera Fung, Aileen Ho, Jennifer Shaffer, Esther Chung, and Manuel Gomez. Use of nintendo wii fit in the rehabilitation of outpatients following total knee replacement: a preliminary randomised controlled trial. *Physiotherapy*, 98:183–188, 2012.
- Luciana Barcala, Luanda Andr Collange Grecco, Fernanda Colella, Paulo Roberto Garcia Lucareli, Afonso Shiguemi Inoue Salgado, and Claudia Santos Oliveira. Visual biofeedback balance training using wii fit after stroke: A randomized controlled trial. *Journal of Physical Therapy Science*, 25(8):1027–1032, 2013. doi: 10.1589/jpts.25.1027.
- Dara Meldrum, Aine Glennon, Susan Herdman, Deirdre Murray, and Rory McConn-Walsh. Virtual reality rehabilitation of balance: assessment of the usability of the nintendo wii fit plus. *Disability and Rehabilitation: Assistive Technology*, 7(3):205–210, 2012. doi: 10.3109/17483107.2011.616922.
- Hussain Al-Issa, Holger Regenbrecht, and Leigh Hale. Augmented reality applications in rehabilitation to improve physical outcomes. *Physical Therapy Reviews*, 17(1):16–28, 2013. doi: 10.1179/1743288X11Y.0000000051.
- Gilda Aparecida de Assis, Ana Grasielle Dionisio Correa, Maria Bernardete Rodrigues Martins, Wendel Goes Pedrozo, and Roseli de Deus Lopes. An augmented reality system for upper-limb post-stroke motor rehabilitation: a feasibility study. *Disability and Rehabilitation: Assistive Technology*, 11(16):521–526, 2014.
- Yee Mon Aung, Adel Al-Jumaily, and K. Anam. A novel upper limb rehabilitation system with self-driven virtual arm illusion. In *36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 3614–3617, Chicago, IL, USA, August 2014.
- Yee Mon Aung and Adel Al-Jumaily. Augmented reality-based rehabio system for shoulder rehabilitation. *International Journal on Mechatronics and Automation*, 4(1):52–62, 2014.
- Marco Gazzoni and Giacinto Liugi Cerone. Augmented reality system for muscle activity biofeedback. *Annals of Physical and Rehabilitation Medicine*, 61:e483–e484, July 2018.
- Joao Monge and Octavian Postolache. Augmented reality and smart sensors for physical rehabilitation. In *10th International Conference and Exposition*

- on *Electrical and Power Engineering (EPE2018)*, pages 1010–1014, 2018.
- Mauricio Sousa, Joao Vieira, Daniel Medeiros, Artur Arsenio, and Joaquim Jorge. SleeveAR: Augmented reality for rehabilitation using realtime feedback. In *Proceedings of the 21st International Conference on Intelligent User Interfaces*, pages 175–185, Sonoma, California, USA, March 2016. doi: 10.1145/2856767.2856773.
- Kevin Desai, Kanchan Bahirat, Sudhir Ramalingam, Balakrishnan Prabhakaran, Thiru Annaswamy, and Una E. Makris. Augmented reality-based exergames for rehabilitation. In *7th International Conference on Multimedia Systems*, Klagenfurt, Austria, May 2016.
- SilverFit. Silverfit 3d. <https://silverfit.com/>, 2019.
- Shichao Yang, Daniel Maturana, and Sebastian Scherer. Real-time 3d scene layout from a single image using convolutional neural networks. In *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2183–2189, May 2016. doi: 10.1109/ICRA.2016.7487368.
- Raul Mur-Artal and Juan D. Tardos. ORB-SLAM2: an Open-Source SLAM System for Monocular, Stereo and RGB-D Cameras. *IEEE Transactions on Robotics*, 33:1255–1262, October 2017. doi: 10.1109/TRO.2017.2705103.
- Zhe Cao, Tomas Simon, Shih-En Wei, and Yaser Sheikh. Realtime multi-person 2d pose estimation using part affinity fields. In *Computer Vision and Pattern Recognition (CVPR)*, 2017.
- Angjoo Kanazawa, Michael J. Black, David W. Jacobs, and Jitendra Malik. End-to-end recovery of human shape and pose. In *Computer Vision and Pattern Recognition (CVPR)*, 2018.
- Joseph Redmon and Ali Farhadi. Yolov3: An incremental improvement. *ArXiv e-prints*, April 2018.
- Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. Microsoft coco: Common objects in context. In David Fleet, Tomas Pajdla, Bernt Schiele, and Tinne Tuytelaars, editors, *European Conference on Computer Vision (ECCV)*, pages 740–755, Cham, 2014. Springer International Publishing. ISBN 978-3-319-10602-1.
- Donald J. Berndt and James Clifford. Using dynamic time warping to find patterns in time series. In *KDD workshop*, volume 10, pages 359–370. Seattle, WA, 1994.
- Stan Salvador and Philip Chan. Toward accurate dynamic time warping in linear time and space. *Intelligent Data Analysis*, 11(5):561–580, 2007.
- British National Health Security (NHS). [https://www.nhs.uk/Tools/Documents/NHS\\\_ExercisesForOlderPeople.pdf](https://www.nhs.uk/Tools/Documents/NHS\_ExercisesForOlderPeople.pdf), 2018.
- Salvatore Gaglio, Giuseppe Lo Re, and Marco Morana. Human activity recognition process using 3-d posture data. *IEEE Transactions on Human-Machine Systems*, 45(5):586–597, Oct 2015. ISSN 2168-2291. doi: 10.1109/THMS.2014.2377111.
- John Brooke. SUS-a quick and dirty usability scale. *Usability evaluation in industry*, 189(194):4–7, 1996.

---

Aaron Bangor, Philip Kortum, and James Miller. Determining what individual SUS scores mean: Adding an adjective rating scale. *Journal of usability studies*, 4(3):114–123, 2009.

Inc. PTC. Vuforia. url<https://developer.vuforia.com/>, 2019.