Automatic Human Body Feature Extraction in Serious Games applied to rehabilitation Robotics
Eva Mogena, Pedro Núñez and José Luis González

Abstract—Current modern society is characterized by an increasing level of elderly population. This population group is usually suffers important physical and cognitive impairments, which implies that older people need care, attention and supervision by health professionals. In this paper, a new system for supervising rehabilitation therapies using autonomous robots for elderly people is presented. The therapy explained in this work is a modified version of the classical ‘Simon Says’ game, where a robot executes a list of motions and gestures that the patient has to repeat. The success of this therapy from the point of view of the software is to provide from an algorithm that detect and classified the gestures that the human is imitating. The algorithm proposed in this paper is based on the analysis of sequences of images acquired by a low cost RGB-D sensor. A set of human body features is detected and characterized during the motion, allowing the robot to classify the different gestures. In addition, this paper describes the human-robot interaction performed by the ‘Simon Says’ game implementation. Experimental results demonstrate the robustness and accuracy of the detection and classification method, which is crucial for the development of the therapy.

Index Terms—Robotics in rehabilitation, RGBD image analysis, SVM, Decision tree, KNN

I. INTRODUCTION

There is a huge demographic change underway in the current modern society, which is characterized by a significant and continuous increase of the elderly population. In Spain, for instance, the last report of the Spanish Statistics Institute (INE) shows how elderly population has been on an upward trend over years (see Fig. 1). Similar reports have been published in most industrialized countries. These reports also conclude that the regular physical and cognitive exercise and adapted for this population group are usually associated with a lower risk of mortality and morbidity, and with a better quality of life. In fact, to improve their quality of life, Information and Communications Technologies (ICT) have successfully been used to assist people with different needs either in hospital or in the patient’s home. The development of physical and/or cognitive therapies supported by these new technologies and supervised by professionals have increased in the last decade. This has several interesting advantages, such as the ability to record the exercise using sensors for a later analysis, e.g., cameras or microphone, or the possibility to automatically adapt the exercises to changes in the environment. Also, one other advantage is the fact that they can provide usable information to occupational therapists at real-time.

Current therapies for elderly population are focused on physical exercises adapted for them and supervised by human professionals. Most of these exercises are based on imitating motions that comprise different joints (e.g., moving an arm making an arc or walking a distance in a straight line). Other therapies focus on cognitive exercises that require different skills (e.g., memory, judgment, abstract reasoning, concentration, attention and praxis) where the elderly interacts with the environment and the practitioner [3]. Therapies that combine both physical and cognitive skills are also common in the literature. The use of a specific robot in these emerging therapies could provide professionals with alternative to conventional therapy sessions.

The main goal of this paper is the development of a system that allows to play a gesture-based variant of the classic ‘Simon Says’ game with senior patients. The system is used to perform a physical and cognitive therapy guided by an autonomous robot. In the proposed therapy, the robot executes a list of gestures that the human has to repeat with an increasing level of difficulty. This therapy helps to improve the physical condition of patients (i.e., by imitating the movements) and their memory (i.e., by remembering the sequence of movements). The robot is equipped with a low-cost RGBD camera. The sequence of images is analyzed for extracting a set of human body features. These features are the input to a second stage, whose output is the human pose estimate by the robot.

In addition, the therapist robot includes a simple Human-Robot Interaction system. This HRI looks for improving patient empathy and motivation through well-defined dialogs.

The remainder of this article is structured as follows:

Fig. 1. Demographic evolution of Spain. Source: INE report

Pedro Núñez and Eva Mogena are with University of Extremadura.
E-mail: emogenac@alumnos.unex.es
José Luis González is with Fundación CénitS-ComputaEX.

1World Health Organization (WHO) defines “Elderly people” as those individuals who are aged 60 years or older [2].
Section II provides a brief summary about similar works in this field of research. In Section III, the 'Simon says' game proposed in the therapy is described. A detailed description of the human-robot interaction is made in Section V. Section IV provides the description of the proposed approach. Experimental results are detailed in Section VI. Finally, the main conclusions of this paper is presented in Section VII.

II. RELATED WORK

In the last decade robotics has rendered outstanding services in the field of medicine. Several therapy robots have been developed by the scientific community for being used in therapies. On the one hand, a group of robots that help patients during physical exercises can be distinguished. Armeo Spring [4] is an exoskeleton used for upper limb rehabilitation in patients with multiple sclerosis. Its main feature is that it can cancel the weight of the arms to an adjustable extent. ReoGo robot [5] helps patients with stroke to regain the upper motor functions. It consists of two connected platforms: a seat with a motorized armrest and an on-board computer that controls the armrest. The MIT’s robot Manus [6] is another robotic device designed to help people that have suffered a stroke, which consists on a robotic control lever that is held by the patient during the therapy.

On the other hand, another group of autonomous robots that supervise and interact with patients can be distinguished. They can generally record different variables of the therapies and adapt the therapy to the patient. Ursus is an autonomous robot developed by RoboLab [7] which has been designed to propose and supervise games for children with cerebral palsy in order to accelerate their recovery and to increase therapy adherence. The last version is equipped with RGB-D cameras to estimate the posture of the patient and acquire information from the environment, a speaker, microphones and two robotic arms for interaction. Ursus has been used in real therapies in the Hospital Virgen del Rocío (Seville, Spain), with favorable results [7]. In similar therapies, also with children with some kind of disability in their upper limbs, the robot Nao has been successfully used [8]. Both robots, Ursus and Nao, perform the gestures that children should imitate, and they are able to perceive the reactions of the patients and interact with them, modifying the exercises when they are not properly conducted. In [9], the authors also use a similar therapy with children with autism spectrum disorders (ASD). The main advantage of all these works is that making a game out of the therapy where the robots look like toys increases the therapy adherence. In order to determine the posture of the patient during the game, in [8] the authors use the normalized Euclidean distance to compare the angles between upper joints from the robot Nao (desired posture) and the patient. A similar metric, also normalized, between the pose performed by patients and the robot is used in [9]. These two methods, [8] and [9], are dependent of the robot, and the number of exercises is then limited, which is a disadvantage with respect to the algorithm described in this paper.

III. ‘SIMON SAYS’ GAME

'Simon Says' is a well-known memory game played with an electronic device with four buttons of different colors (yellow, blue, green and red). The device generates a pattern of sounds and lights and expects the user to repeat the pattern. If the user succeeds the patterns become progressively longer and more complex. Because the patterns that Simon creates are of increasing difficulty and it has to be remembered and imitated by the user, this game is considered to help develop the memory. In this paper, a modified version of the 'Simon says' game is described for being used in therapies with elderly people. Instead of a series of lights and tones, the patient has to reproduce a series of physical exercises during the therapy. Henceforth the proposed game helps to develop the psychomotricity and the memory of the patient.

A. Exercises

In this section, the set of exercises proposed in the therapy is described. It is composed of four human poses with low difficulty level. These exercises have been defined by a team of occupational therapists, and they are especially designed for elderly people. Fig. 2 illustrates the set of exercises used in the proposal, which are described below.

- Exercise 1: Cross posture. The hands must be at the height of the shoulder.
- Exercise 2: Put the hands in the hips, with the arms akimbo.
- Exercise 3: Hold the arms in front of their corresponding shoulder, with the arms straight.
- Exercise 4: Raise the hands over the head, with the arms outstretched.

IV. RGBD DATA ANALYSIS FOR HUMAN POSE RECOGNITION AND CLASSIFICATION

The proposed approach is based on the analysis of RGBD image sequences. From each image, the human skeleton is detected and a set of human body features is extracted and characterized. First, the system has to determine when the senior is performing a movement and when he has finished it and has changed his pose. Then, the robot has to extract body features and classify the human pose into the set of postures associated to each exercise.
A. RGBD image sequence acquisition and human skeleton detection

In this work, a low level RGB-D camera is used for acquiring video and distance data (Kinect sensor for Windows). This sensor offers RGB images of 640x480 pixels. The depth data has a spatial x/y resolution of 3 mm at 2 m distance from the sensor [10]. This camera produces a real-time 30fps stream of VGA frames, which is enough for a regular therapies. In order to quickly and accurately predict 3D positions of body joints from a single depth image, using no temporal information, the method described in [11] is used, which uses the Kinect SDK for Windows. The proposed approach takes into account eight body joints: head \( H(x,y,z) \), left and right shoulders, \( LS(x,y,z) \) and \( RS(x,y,z) \), left and right elbows, \( LE(x,y,z) \) and \( RE(x,y,z) \), left and right hands, \( LH(x,y,z) \) and \( RH(x,y,z) \), and chest \( CH(x,y,z) \).

B. Body gesture detection and characterization

Once the skeleton of the patient is detected, the system extracts a set of human body gesture features. These features are described below:

- **Quantity of motion** (QoM) is a measure of the amount of detected skeleton motion. This value is used to determine when the patient is moving. Let \( N \) being the total number of consecutive frames taken, and \( x_A^i \) being the 3D position of an articulation at an instant of time \( i \). Then, QoM\(^A\) is defined as:

\[
QoM_i^A = \frac{1}{N} \sum_{k=0}^{N} x_A^i - x_A^{i-1}
\]

where \( A \in \) (left hand, right hand, left elbow, right elbow, chest, head). Finally, the total QoM is evaluated as the average value of QoM\(^A\).

- **Contraction Index** (CI). It measures the amount of contraction and expansion of the body. It takes range of values between 0 and 1. CI is bigger when the patient’s posture keeps limbs near the barycenter. To calculate this value, only the relationship between the human chest and the hand poses has been taken into account. The algorithm doesn’t consider the elbows due to their limited dynamic range. This relationship has been carried out by the area of the triangle defined by these three points (i.e., chest and two hands). First, the semi perimeter \( s \) of the triangle is calculated:

\[
s = \frac{u + v + w}{2}
\]

where \( u, v \) and \( w \) are the sides of the triangle. The CI, using Heron’s Method, remains as follows:

\[
CI = \sqrt{s \cdot (s-u) \cdot (s-v) \cdot (s-w)}
\]

- **Angle Arm** (\( \alpha \)). This feature allows the system to get information on whether the arms are stretched or bent. In order to calculate \( \alpha \) value, a triangle is built as:
  - Side A: length from the shoulder to the elbow
  - Side B: length from the elbow to the hand
  - Side C: length from the shoulder to the hand

As hand, elbow and shoulder positions are given by the algorithm, these lengths can be easily calculated. Hence, the desired \( \alpha \) angle can be calculated by the law of cosines, which reads as follows:

\[
A^2 = B^2 + C^2 - 2BC \cos \alpha
\]

\[
\alpha = \arccos \left( \frac{B^2 + C^2 - A^2}{2BC} \right)
\]

- **Height of the hands** (\( Y \)). It corresponds to the average value of the height at which the hands are located. First, it has been assumed that the ground of the frame of reference is placed in the subject’s foot. The maximum possible value of the hand is estimated from the value of the arm’s length plus the shoulder’s height. By defining \( H, E \) and \( S \) as elbow, shoulder and length, respectively, and being \( x, y \) and \( z \) the 3D-coordinate associated to a joint, then

\[
L_{EH} = \sqrt{(E.x - H.x)^2 + (E.y - H.y)^2 + (E.z - H.z)^2}
\]

\[
L_{SE} = \sqrt{(S.x - E.x)^2 + (S.y - E.y)^2 + (S.z - E.z)^2}
\]

\[
L_{arm} = L_{EH} + L_{SE}
\]

\[
y_{max} = L_{arm} + S.y
\]

Finally, the hand height (\( Y \)) can be normalized as is shown in (7).

\[
Y = \frac{H.y}{y_{max}}
\]

- **Depth of the hands** (\( D \)). This feature is used to indicate to what extent the hands are in the \( z \) plane of the patient’s torso or not. To calculate this value, the difference between the depth value of hands (\( H \)) and the depth value of the chest (\( Ch \)) is estimated, as shown in equation 8.

\[
D = |H.z - Ch.z|
\]

Finally, expected values for the features described in this section for the different exercises of the therapy ‘Simon Says’ are summarized in Table I.

C. Human body pose classification

Once the features have been detected by the robot during the therapy, the next step is to decide what kind of exercise has been performed by patient. Classification models are algorithms that are able to learn after performing a training process with known data. In the proposed work three well-known models are studied: Decision Trees (DT) [12], [13], K-Nearest Neighbours (KNN) [14], [15], and Support Vector Machines (SVM) [16], [17].

V. HUMAN-ROBOT INTERACTION

One of the main goals of the ‘Simon Says’ game is to motivate patients, making therapies more attractive, and to prevent patients from falling into a state of discouragement, which dramatically decreases the therapy adherence.

In order to prevent patients from associating negative feelings to the therapy (e.g., boredom, de-motivation) the robot must interact with the patient, guiding the therapy session in the same way that a therapist would guide their patients. The robot must communicate with the patient in an appropriate
way. Besides, the therapist robot has to encourage the final user to continue with the session, telling whether the exercises are being performed properly or not, so the patient can have a useful feedback.

The robot should also be able to show the patient which exercises has to be performed during the therapy, and advise him when the patient is not able to remember the exercises. It should also be able to encourage the patient by providing information of the progress in the exercises, for example, telling the patient when they are getting close to the next difficulty level. All these feedback messages help increasing the motivation of the patient and to improve the outcome of the therapy session.

Figure 3 illustrates the state machine implemented in the 'Simon Says' game. This state machine shows the process followed by the robot in the interaction with the patient. The robot begins telling the patient the game level. Depending on the game level in which the patient is, he will have to remember more or fewer exercises. For example, in the level 1 the patient has to perform a single exercise. In the level 2, he has to perform the exercises of the level 1 and the level 2, respectively. Then, in the level 3 he will have to perform the exercises of the level 1, the level 2 and, finally, the level 3. And so on.

Once the patient knows his level within the game, the robot tells the patient what exercise has to perform. At the same time, the robot is acquiring the information from the human motion and evaluate its success. If the exercise is not properly performed by the patient, the robot explains to the patient about the mistake and then, it repeats the instructions. If the patient performs the required exercise properly, the robot informs the patient and then, it requests him to keep doing the exercise.

If the series is not over, the robot will ask the patient to perform the following exercise. Once the series ends, the robot increases the level of difficulty of the game, which means that the patient should remember a new physical exercise, and finally the robot starts with the new level.

### VI. Experimental Results

In order to demonstrate the accuracy and robustness of the proposal, a set of experiments were conducted. Figure 4 shows the robot used in the experiments with the RGB-D sensors marked on the picture. Currently, this robot has two different RGB-D cameras, but only the one marked as ‘1’ in the figure is used for the therapy. This section describes the database used for the experiments (including the main interface of the application), and analyses a set of representative tests. Both tests and application are programmed using the RoboComp framework, developed by RoboLab [18].

#### A. Database

A database with recorded data for the training and recognition process has been created. This database consists of data of 20 subjects that have been recorded with the Kinect. All files that compose the database are stored in plain text format. Data from 5 of them has been used to train the component, whereas the recognition tests were carried out with the remaining 15 subjects whose data have not been used for training. The patients are older than 65 and do not have previous experience with the described system. There are 12 female patients and 8 male patients.

This database has been made publicly through the download service of the Universidad de Extremadura, using the link http://goo.gl/IxovJ4, so that any developer can use it with their systems.

#### B. Interface

As shown in Fig. 5, the interface of the application is composed by several parts that are detailed below:

- **A:** This block shows the different calculated features.
- **B:** It allows to choose the level of difficulty in the “Simon Mode”.
- **C:** Enables to chose the mode in which the component is run: Training, Detecting, Accurate\(^2\).
- **D:** It allows to choose the decision method.
- **E:** Buttons to allow training the component.
- **F:** Buttons used to simulate the performance of one of the defined exercises.
- **G:** Label showing different messages depending on the mode selected.
- **H:** Buttons used to store data in a Matlab file to generate graphs.
- **I:** Button used to stop the program.
- **J:** Three-dimensional view of the patient’s model.

#### C. Evaluation of the features extraction algorithm

An analysis of the features extracted for each exercise has been conducted. The tests check whether if these features are among the expected values, and if they discriminate between

\(^2\)This mode allows to calculate the accuracy of the different algorithms used, and the success rate of each exercise for each method.
the different exercises. To do this, some graphics have been generated using Matlab software.

Fig. 6 shows the Contraction Index that has been calculated for the different exercises. As is shown in the figure, the values of the Contraction Index are as expected: a high value for the Exercise 1, in which the hands are widely separated from the body, a median value for the Exercise 4, since this exercise also separates the arms from the body, but the arms are close together, and a low value for the Exercises 2 and 3.

In Fig. 7 and Fig. 8 the data collected with regards to the angles formed by the right and left arms are illustrated, respectively. As the graphics show, the values of the angles are as expected, being approximately $180^\circ$ for the Exercises 1, 3 and 4, due to in these exercises the arms are stretched, and about $90^\circ$ for the Exercise 2, because in this exercise the arms are flexed. It can be appreciated that for this patient there are several oscillations in the value of the calculated angle for the right arm, due to the Kinect resolution.

Fig. 9 and Fig. 10 show the heights of the hands, normalized to 1, associated to each exercise. The results are again as expected. In the Exercise 4, in which the hands are above the head, these features find their maximum with an average value of 1. For the exercises 1 and 3, in which hands are at the shoulder height, the average value is 0.7. Finally, for the exercise 2, in which the hands are on his hips, the average values are close to 0.4.

Finally, data associated to the depth of both hands are shown in Fig. 11 and Fig. 12. Again, these values are as expected. In the exercises 1, 2 and 4, in which the hands are kept in the plane of the body, some very low values, close to 0, are obtained. However, for the exercise 3, a higher value, between 60 and 70 cm, is obtained.

D. Recognition Accuracy

This section evaluates the accuracy of the three decision methods described in this paper. Table II shows the results obtained. As is illustrated in the table, the decision method that gets the best results is the Decision Tree. This algorithm has an average of probability of success of 99.61%. KNN method also obtains good results, although lower than the DT algorithm (79.38%). SVM method presents some results in which the success rate is very satisfactory and others where it is not as good, but in any of the cases the accuracy obtained is higher than the accuracy obtained by using DT or KNN.

The Decision Tree algorithm gets better results because it asks questions about the variables to decide whether an exercise or another. On the contrary, the KNN and SVM algorithms associate the exercise to points, and represent these points spatially. When a new point has to be classified, KNN algorithm compares its distance to the closest points, and the SVM algorithm evaluates the area where the point is. As these exercises are differentiated by a few characteristics, it is possible that the points are together in the spatial representation and this creates confusion when deciding.

VII. Conclusions

This paper presented a system for supervising rehabilitation therapies using autonomous robots for elderly people. In the therapy, it has been implemented a modified version of the classic 'Simon says' game, in which a robot executes a list of motions and gestures that the human has to repeat each
time. Four different exercises (i.e., human body poses) have been described in the game, thus, the therapist robot has to be able to detect, recognize and classify human body poses. To do that, in this paper has been described a set of human body features in order to characterize these postures, and also three different classification algorithms have been evaluated. The results of this work demonstrates the accuracy of the described algorithm. Finally, a RoboComp component has been included in the repository for later development.

Future challenges should focus on the capability of the robot to analyze not only static exercises, but also dynamic ones. Also, similar to other works like [8], try to describe a plan during the game session to improve the empathy and evaluate the improvement of the results. Moreover, a set of questionnaires could be used in order to collect the impressions of elderly people after interacting with the robot.

ACKNOWLEDGMENTS
This work has been partially supported by the MICINN Project TIN2015-65686-C5-5-R, by the Extremadura Government project GR15120, and by COMPUTAEX Fundation.

REFERENCES
Fig. 8. Angle of the Left Arm for the four exercises of the therapy.

Fig. 9. Height of the Right Hand for the four exercises of the therapy.

Fig. 10. Height of the Left Hand for the four exercises of the therapy.

Fig. 11. Depth of the Right Hand for the four exercises of the therapy.


Fig. 12. Depth of the Left Hand for the four exercises of the therapy.

<table>
<thead>
<tr>
<th>Subject</th>
<th>DT</th>
<th>KNN</th>
<th>SVM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>100.00%</td>
<td>91.67%</td>
<td>80.00%</td>
</tr>
<tr>
<td>2</td>
<td>100.00%</td>
<td>95.83%</td>
<td>87.50%</td>
</tr>
<tr>
<td>3</td>
<td>100.00%</td>
<td>81.67%</td>
<td>62.50%</td>
</tr>
<tr>
<td>4</td>
<td>100.00%</td>
<td>72.50%</td>
<td>61.67%</td>
</tr>
<tr>
<td>5</td>
<td>99.17%</td>
<td>70.00%</td>
<td>39.17%</td>
</tr>
<tr>
<td>6</td>
<td>100.00%</td>
<td>77.50%</td>
<td>42.50%</td>
</tr>
<tr>
<td>7</td>
<td>100.00%</td>
<td>80.00%</td>
<td>75.83%</td>
</tr>
<tr>
<td>8</td>
<td>95.00%</td>
<td>68.33%</td>
<td>56.67%</td>
</tr>
<tr>
<td>9</td>
<td>100.00%</td>
<td>71.67%</td>
<td>45.00%</td>
</tr>
<tr>
<td>10</td>
<td>100.00%</td>
<td>95.00%</td>
<td>87.50%</td>
</tr>
<tr>
<td>11</td>
<td>100.00%</td>
<td>62.50%</td>
<td>38.33%</td>
</tr>
<tr>
<td>12</td>
<td>100.00%</td>
<td>68.33%</td>
<td>50.00%</td>
</tr>
<tr>
<td>13</td>
<td>100.00%</td>
<td>97.50%</td>
<td>65.83%</td>
</tr>
<tr>
<td>14</td>
<td>100.00%</td>
<td>76.67%</td>
<td>67.50%</td>
</tr>
<tr>
<td>15</td>
<td>100.00%</td>
<td>81.67%</td>
<td>65.83%</td>
</tr>
<tr>
<td>Mean</td>
<td>99.61%</td>
<td>79.38%</td>
<td>61.72%</td>
</tr>
</tbody>
</table>

TABLE II
ACCURACY OBTAINED FROM THE DIFFERENT DECISION METHODS