

# Low Cost RGB-D Vision Based System to Support Motor Disabilities Rehabilitation at Home

**Flavia Benettazzo, Sabrina Iarlori, Francesco Ferracuti, Andrea Giantomassi, Davide Ortenzi, Alessandro Freddi, Andrea Monteriù, Silvia Innocenzi, Marianna Capecci, Maria Gabriella Ceravolo and Sauro Longhi**

**Abstract** Physical rehabilitation is an important medical activity sector for the recovery of physical functions and clinical treatment of people affected by different pathologies, as neurodegenerative diseases (i.e. multiple sclerosis, Parkinson and Alzheimer diseases, amyotrophic lateral sclerosis), neuromuscular disorders (i.e. dystrophies, myopathies, amyotrophies and neuropathies), neurovascular disorders/trauma (i.e. stroke and traumatic brain injuries), and mobility for the elderly. During the rehabilitation, the patient has to perform different exercises specific for the own disease: while some exercises have to be performed with specific equipment and under the supervision of professional staff, others can be

---

F. Benettazzo · S. Iarlori · F. Ferracuti · A. Giantomassi (✉) · D. Ortenzi · A. Monteriù · S. Innocenzi · S. Longhi  
Dipartimento di Ingegneria dell'Informazione, Università Politecnica delle Marche,  
Via Brezze Bianche, 60131 Ancona, Italy  
e-mail: a.giantomassi@univpm.it

F. Benettazzo  
e-mail: f.benettazzo@univpm.it

S. Iarlori  
e-mail: s.iarlori@univpm.it

F. Ferracuti  
e-mail: f.ferracuti@univpm.it

D. Ortenzi  
e-mail: d.ortenzi@univpm.it

A. Monteriù  
e-mail: a.monteriu@univpm.it

S. Innocenzi  
e-mail: s.innocenzi@univpm.it

performed by patients without the supervision of physiotherapists. In this last case, it is possible to reduce the costs of health and care national system and to accomplish the treatment at home. In this work, a computer vision system for physical rehabilitation at home is proposed. The vision system exploits a low cost RGB-D camera and open source libraries for the image processing, in order to monitor the exercises performed by the patients, and returns a video feedback to improve the treatment effectiveness and to increase the user's motivation, interest, and perseverance. Moreover, the vision system evaluates an exercise score in order to monitor the rehabilitation progress, an helpful information both for the clinician staff and patients, and allow physiotherapists to monitor the patients at home and correct their posture if the exercises are not well performed. This approach has been implemented and experimentally tested using the Microsoft Kinect camera, demonstrating good and reliable performances.

## 1 Introduction

In recent years, researches in the field of computer vision have given rise to different applications with significant impact on the life of people with disabilities or temporarily affected by motor difficulties: to support them, many approaches have been investigated. Physical and cognitive rehabilitation is a complex and long-term process that requires clinician experts and appropriate tools [1]. Conventional rehabilitation training programs typically involve extensive repetitive range-of-motion and coordination exercises, and require professional therapists to supervise the patients' movements and assess the progress. This approach provides limited objective performance measurement and typically lacks engaging content to motivate individuals during the program. To address these problems, new rehabilitation methodologies are being developed and have recently gained significant interest in the physical therapy area. The main idea of this rehabilitation approach is to use sensing devices to capture and quantitatively assess the movements of patients under treatment, in order to track their progress more accurately [2]. In addition, the patients need to be more motivated and engaged in these physical activities [3,

---

S. Longhi

e-mail: sauro.longhi@univpm.it

M. Capecci · M.G. Ceravolo

Dipartimento di Medicina Sperimentale e Clinica, Università Politecnica delle Marche,

Via Brezze Bianche, 60131 Ancona, Italy

e-mail: m.capecci@univpm.it

M.G. Ceravolo

e-mail: m.g.ceravolo@univpm.it

A. Freddi

Università degli Studi eCampus, Via Isimbardi 10, 22060 Novedrate, CO, Italy

e-mail: alessandro.freddi@unicampus.it

4]. People often cite a lack of motivation as an impediment to them performing the exercise regularly [5]: this study indicates that only 31 % of people with motor disabilities perform the exercises as recommended which can result in negative consequences such as chronic health conditions. Results show that video capture virtual reality technology can be used as complement to conventional physiotherapy. This technology is cheap, easy to use by users and therapists as a home based rehabilitation tool. Further it helps stroke patients to use this technology as part of their home based rehabilitation program encouraging self-dependability in performing activities of daily living [6]. The aim of this paper is to develop a Kinect-based assistive system, who can help people who have movement disorders to perform rehabilitative exercises program at home [7]. Kinect is a cheap and easy to set up sensor and can be used in both home and clinical environments; his accessibility could significantly facilitate rehabilitation, allowing more frequent repetition of exercises outside standard therapy sessions [3].

In literature it is possible to find different approaches to these rehabilitation systems. In [7] is presented a Kinect-based rehabilitation system to assist patients with movement disorders by performing the “seated Tai Chi” exercises at home. By this system authors can evaluate if patients achieve rehabilitation results correctly or not. By using skeletal tracking of Microsoft Kinect sensor, each gesture of the patient, performing Tai Chi is recognized and validated. In [8] a system for physical rehabilitation is proposed, called the Kinerehab and based on Microsoft Kinect, which is webcam-style add-on peripheral intended for the Xbox 360 game console. Kinerehab uses image processing technology of Kinect to detect patient’s movement, detecting automatically patient’s joint position, and uses the data to determine whether the patient’s movement have reached the rehabilitation standard and if the number of exercises in a therapy session is sufficient. Using this system, users can evaluate the accuracy of their movements during rehabilitation. The system also includes an interactive interface with audio and video feedback to enhance patient’s motivation, interest, and perseverance to engage in physical rehabilitation. Details of users’ rehabilitation conditions are also automatically recorded in the system, allowing therapists to review the rehabilitation progress quickly. In [9] is presented the analysis of the use of Kinect sensor as an interaction support tool for rehabilitation systems: using a scoring mechanism, the patient performance is measured, as well as his improvement by displaying a positive feedback.

Authors propose a RGB-D camera vision system to monitor patients affected by motor disabilities and evaluate the precision by which they repeat the rehabilitation exercise at home. Particularly, the vision system supports the physiotherapists’ work of controlling the validity and efficacy of each exercise and aims to increase the user motivation. Microsoft Kinect camera is used, and after the calibration, a set of functional body joints and a body skeleton are mapped and extracted. These features allow an Artificial Neural Network (ANN) to recognise, from the start position, which exercise is realised. Then, the Online Supporting Algorithm evaluates user performance during the exercise execution. It also provide a video feedback to the user if position and/or speed of execution are not

correct respect to the right exercise. Moreover the vision system is able to compute a score about the exercise helpful both for the clinician staff and the patient.

This paper is organized as follows: the Sect. 2 proposes the system details with the hardware and software for the RGB-D camera. In Sect. 3 is described the exercise recognition method by means of the Calibration Algorithm and neural network. In Sect. 4 authors present the online support algorithm, which allows to obtain the score feedback on how the patient is performing the exercise. In Sect. 5 the experimental results are presented, with the two kind of exercises tested and in Sect. 6 conclusions and future works are given.

## 2 System Configuration

In this section the computer vision system details are provided. The proposed algorithms exploit a low cost RGB-D camera and the open source software Open Natural Interaction (OpenNI) [10]. More information about *OpenNI* applications are given in [11]. The vision algorithm has been developed in Robot Operating System (ROS) [12]. This choice is motivated by the fact that ROS is spreading exponentially [13].

### 2.1 Hardware

A Microsoft Kinect camera is adopted as RGB-D vision sensor. A RGB-D camera allows to reach better performance in people and objects identification, even if the background and the person or the object have the same color, and recognize overlapped objects by calculating the distance for each of them. More in detail, the Kinect camera falls within the category of Structured Light (SL) cameras. These images are captured by a normal 2D camera and analyzed. Then, depth information is extracted. Using a SL camera, the depth can be recovered from simple triangulation by giving a specific angle between emitter and sensor. An SL camera is composed by an IR projector, a diffraction grating and a standard Complementary Metal-Oxide Semiconductor (CMOS) detector with a band-pass filter centered at the IR light wavelength. The diffraction grating is a Computer-Generated Hologram (CGH) that produces a specific periodic structure of IR light when the laser shines through it. The projected image does not change in time. The IR CMOS sensor detects this pattern projected onto the room and scene, and generates the corresponding depth image. Compared with cameras based on Time of Flight (TOF) technology, SL cameras have a shorter range and images appear to be noisier and less accurate, but thanks to post processing capability, it is possible to address these issues. Moreover SL cameras are cheaper than TOF cameras. More information about SL cameras can be found in [14].

## 2.2 Software

OpenNI is used to implement more functionality of the vision sensor. It is a multi-language and multi-platform framework that defines the API (Application Programming Interface) for writing applications that use natural interaction, i.e. interfaces that do not require remote controls but allow people to interact with a machine through gestures and words typical of human-human interactions. This API has been chosen because it incorporates algorithms for background suppression and identification of people motion, without causing a slowdown in the video. The proposed algorithm has been developed in ROS, an open source framework for robotic applications. ROS is a message-based, tool-based system designed for mobile manipulators: the system is composed of libraries that are designed to work independently. ROS is based on the Unix-like philosophy to building many small tools that are designed to work together.

## 3 Exercises Recognition

In order to track the exercises, the subject has to be identified. This operation is called *Calibration*. The *Calibration Algorithm* recognizes different points of interest of the person's body, associating a joint to each of them, as shown in Fig. 1. The calibration operation is required by the vision sensor to find the user in its field of view and is performed by using the functionality already available within the OpenNI library.

After the user calibration, the system starts the exercises recognition procedure. This procedure allows to identify the performed exercise by the user in order to

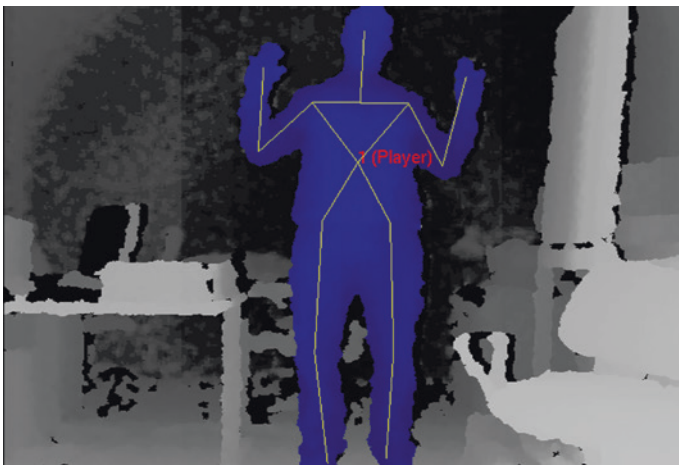


Fig. 1 Person calibration

set-up the online algorithm for the user support. In this work, the motion sequence of the considered exercises are shown in Figs. 3 and 4. After the action recognition, the online algorithm chooses the joints of interest for evaluating the action in order to monitor the movement, evaluates if the exercise is well performed and returns a feedback to the user. The exercises recognition procedure is based on the Artificial Neural Networks (ANNs) [15], which are computational models inspired to the human brain. ANNs allow to detect patterns and data relationships through a training process that is based on learning and generalization procedures. ANNs are found in almost all fields of contemporary science, and are associated with a wide variety of artificial intelligence applications which are concerned with problems of pattern recognition and classification.

In the present work a feed-forward neural network, with three weighted layers (i.e. input, hidden and output layer) is considered [16]; the network is used to classify two exercises performed by users. The ANN is used to classify the initial position assumed by a user, in other words users can perform an exercise as they like, without setting the system before. Once the patient is correctly calibrated, then the system waits until the user is standing still in the initial position. After the classification is performed, the ANN updates the online supporting algorithm with the actual exercise that the patient starts to perform.

## 4 Online Supporting Algorithm

The online algorithm allows users to obtain immediate feedback on how they are performing the exercises. The proposed *Online Supporting Algorithm* gives a feedback during the exercise execution. The *Online Supporting Algorithm* is defined by the following steps:

1. **Identification of the joints of interest:** after the exercise identification, as described in Sect. 3, the joints of interest for evaluating the specific exercise are selected, in order to monitor the movement.
2. **Waiting for the starting position:** if the user does not take the correct starting position, he receives a feedback with the explanation about what he is doing wrong. When the user has taken the correct starting position, he receives a video feedback, which invites him to start the exercise.
3. **Tracking the movement:** during the performance, the joints of interest are tracked and two features are computed and compared with the same features stored from the physiotherapist performance, which are taken as reference. The first feature is defined by the position error computed at each time instant of all joints of interest. The error is given by the euclidean distance between the reference position (i.e. the physiotherapist stored features) and the user actual position, and it is defined as:

$$Perr_{\alpha}(k) = \sqrt[2]{\sum_i (c_{\alpha,i} - \tilde{c}_{\alpha,i})^2} \quad i = 1, \dots, N_{\alpha}, \quad (1)$$

where  $Perr_\alpha(k)$  is the position error computed at time  $k$  for the exercise  $\alpha$ ,  $c_{\alpha,i}$  is the  $i$ th body coordinate considered for the exercise  $\alpha$  and  $\tilde{c}_{\alpha,i}$  is the reference  $i$ th body coordinate (i.e. form the physiotherapist stored features), of the same exercise;  $N_\alpha$  is the number of coordinates taken into account for the exercise  $\alpha$ .

**Assumption 1** Note that Eq. 1 holds in the Euclidean space, even if an arbitrary set of coordinates from body joints is taken, due to the hypotheses that the set of  $N_\alpha$  coordinates taken into account, for the exercise  $\alpha$ , defines a new Euclidean space of  $N_\alpha$  dimensions.

This feature is used to evaluate if the position is correct at each time step. The position error integral is computed over each sample and used for the score computation at the end of exercise execution.

At the same way the second feature is defined by the velocity error as:

$$Verr_\alpha(k) = \sqrt{\sum_i (\dot{c}_{\alpha,i} - \tilde{c}_{\alpha,i})^2} \quad i = 1, \dots, N_\alpha, \quad (2)$$

where  $Verr_\alpha(k)$  is the velocity error computed at time  $k$  for the exercise  $\alpha$ ,  $\dot{c}_{\alpha,i}$  is the  $i$ th body coordinate velocity considered for the exercise  $\alpha$  and  $\tilde{c}_{\alpha,i}$  is the reference  $i$ th body coordinate velocity (i.e. form the physiotherapist stored features), of the same exercise;  $N_\alpha$  is the number of coordinates taken into account for the exercise  $\alpha$ . The following Assumption 2 remarks the validity of the defined velocity features.

**Assumption 2** Note that Eq. 2 holds in the Euclidean space, even if an arbitrary set of coordinates from body joints is taken, due to the hypotheses that the set of  $N_\alpha$  coordinates taken into account, for the exercise  $\alpha$ , defines a new Euclidean space of  $N_\alpha$  dimensions.

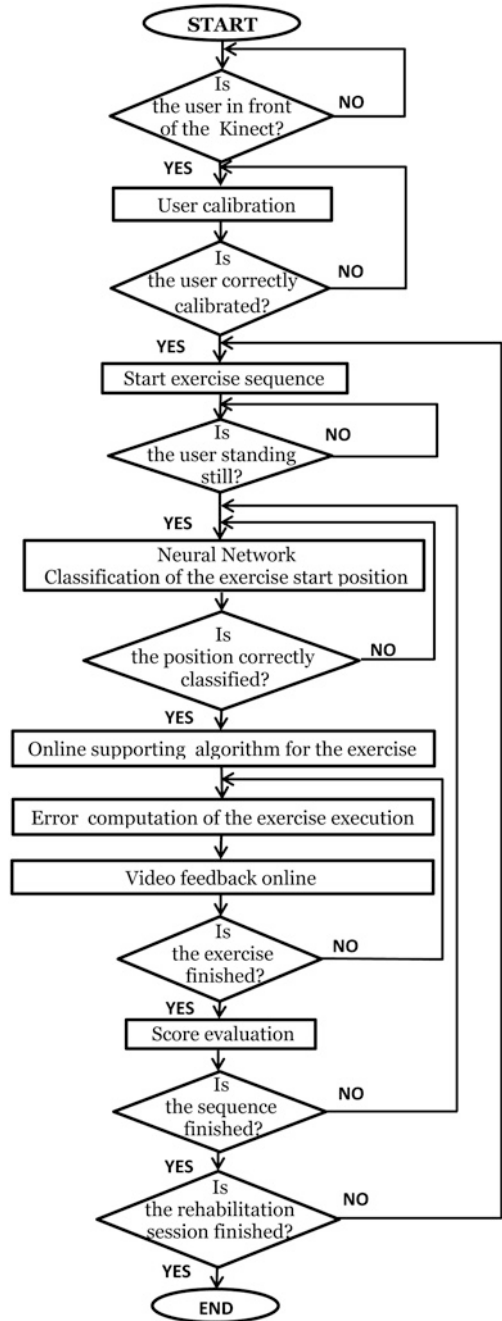
This feature is defined to assess if the user performs the exercise at the right time. The velocity error integral is computed over each sample and used for the score computation at the end of exercise execution.

During the performance, the user receives feedback on how he is performing the exercise. The feedback can relate both if the speed is right and/or if he is taking an incorrect position. The exercises should be performed at a constant speed and without abrupt movements.

4. **Detection of the final position and evaluation:** once the user reaches the final position and the algorithm recognizes it, a score is computed as an inverse proportionality factor of the sum of the final position and velocity errors. If the user does not reach the right final position he receives a feedback and the exercise receives a negative evaluation.

A flowchart describing the proposed algorithms is provided in Fig. 2.

Fig. 2 Flow chart of the proposed computer vision system procedure





### 5 Experiment and Results Discussion

In this section the proposed methodology is tested considering two postural exercises for the back performed by 10 actors. Users should stand up in front of the RGB-D camera with legs slightly apart. The execution of the first exercise consists of lifting the upper limbs extending arms up over the head. The elbow should stay extended throughout the exercise execution. For the second exercise patient starts holding a stick with both hands with arms in elevation (above the head), with his elbows extended. The exercise consists in tilting the torso slowly to the left, returning to a neutral standing position and then proceeds with the same exercise on the right side.

The chosen postural exercises are shown in Figs. 3 and 4 respectively, the photo sequences shown start from the initial position, the left photo, to the end position, the right photo. For each exercise a set of coordinates of interest are selected as reported in Table 1, these sets of coordinates are considered crucial for the evaluation of the movements during the exercises execution and avoid the use of all joints information: 20-3 coordinates.

The ANN for the exercise classification uses 20-6 input signals and has 2 outputs, one for each considered exercise. In this work, 20 joints are tracked and 6

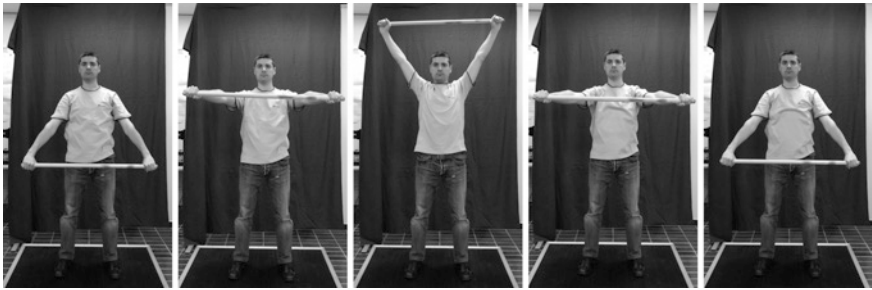


Fig. 3 Motion sequence of first exercise



Fig. 4 Motion sequence of second exercise

**Table 1** Skeleton joints coordinates selected as features for each exercise

Joints	Exercise 1			Exercise 2		
	x	y	z	x	y	z
Hand right	◆	◆	–	◆	–	◆
Hand left	◆	◆	–	◆	–	◆
Elbow right	◆	–	–	–	–	–
Elbow left	◆	–	–	–	–	–
Foot left	◆	◆	◆	◆	◆	◆
Foot right	◆	◆	◆	◆	◆	◆
Head	–	–	–	◆	–	–

features are considered for each joint: the  $x$ ,  $y$  and  $z$  positions and the  $\dot{x}$ ,  $\dot{y}$  and  $\dot{z}$  velocities. So the input and output layers are composed by 20.6 and 2 neurons respectively. The neurons number of the hidden layer is set to 10.

### 5.1 Exercises Execution Results

Figure 5 shows the exercises classification results, which are described through the confusion matrices. In the test, the training data are the 40 % of all data, while the validation and testing data are each the 30 % of all data. The neural network is trained using the scaled conjugate gradient (SCG) back-propagation algorithm both with exercises well performed and not [17]. The Fig. 5 shows as the neural network is able to detect the initial positions assumed by the users.

When the user assumes the right starting position, the exercise starts and the *Online Supporting Algorithm* computes the errors defined in Eqs. 1 and 2 between the coordinates selected of the joints of interests, at each time instant.

For the Exercise 1 the errors on position and velocity are shown in Fig. 6. The black straight lines denote the error limits. If the user error exceeds the limit on position or velocity at a certain time instant then the *Online Supporting Algorithm* gives a feedback. The feedback messages that the algorithm can send to the user are: “Too fast”, “Too slow”, “Do not spread your arms”, “Stretch out your elbows”, “Do not move your feet”. If the user assumes the right position at the right time, he will take the right final position. At the end of the exercise, the performance is evaluated calculating the score, as described in Sect. 4. In Fig. 6a, b, the patient denoted by blue lines performs the exercise correctly, while the red lines show a patient who makes position and velocity errors and is unable to retrieve the error.

For the exercise 2 the errors on position and velocity are depicted in Fig. 7. If the user error exceeds the limit about position or velocity at a certain time instant then the algorithm gives a feedback to the patient. In this case the feedback messages that the algorithm can send to the user are: “Align arms with the body”, “Fold more the torso”, “Do not move your feet”. If the user assumes the right position at the right time, he will take the right final position. At the end of the exercise, the performance is evaluated calculating the score, as described in Sect. 4. In Fig. 7a, b, the patient

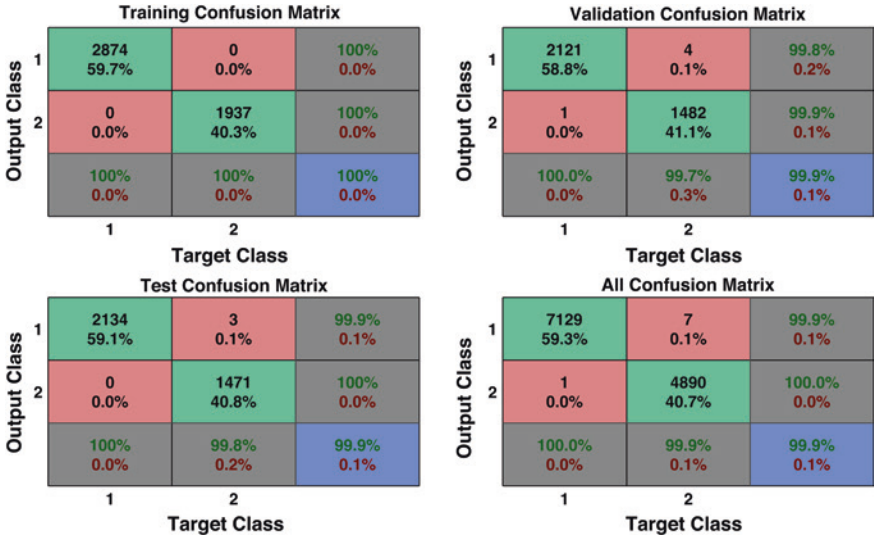


Fig. 5 Exercises recognition confusion matrices

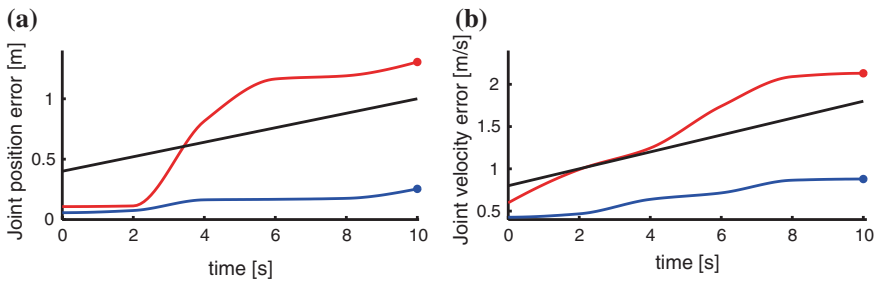


Fig. 6 Exercise 1, the black straight line denotes the error threshold, the blue line denotes the error in case of correct execution. **a** Position; **b** Velocity

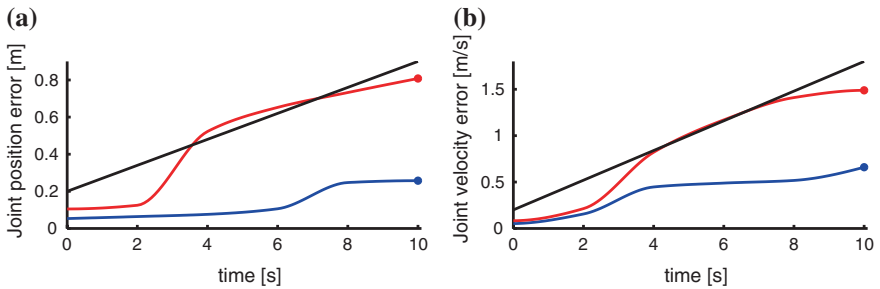


Fig. 7 Exercise 2, the black straight line denotes the error threshold, the blue line denotes the error in case of correct execution. **a** Position; **b** Velocity

denoted by blue lines performs the exercise correctly, while the red lines show a patient that makes position and velocity errors and later retrieves the error.

## 6 Conclusions and Future Works

Physical rehabilitation aims to enhance and restore functional ability and quality of life of people with physical impairments. Since some exercises can be performed by patients at home without the supervision of physiotherapists, a smart solution is needed to monitor the patients. The authors main contribution is the development of a computer vision system able to evaluate the correctness of the exercises performed by the patients, to return a video feedback in order to improve the treatment effectiveness and increase the user's motivation, interest, and perseverance. The vision system exploits a low cost RGB-D camera and the open source libraries OpenNI for the image processing and it is developed in ROS framework. Preliminary results show that the proposed procedure is able to detect the exercises and evaluate reliably their correctness. The authors are currently considering the employment of patients, instead of actors, for the validation of the proposed system to support them in the rehabilitation care.

## References

1. González-Ortega, D., Díaz-Pernas, F., Martínez-Zarzuela, M., Antón-Rodríguez, M.: A Kinect-based system for cognitive rehabilitation exercises monitoring. *Comput. Methods Programs Biomed.* **113**(2), 620–631 (2014)
2. Iarlori, S., Ferracuti, F., Giantomassi, A., Longhi, S.: RGB-D video monitoring system to assess the dementia disease state based on recurrent neural networks with parametric bias action recognition and DAFs index evaluation. In: Miesenberger, K., Fels, D., Archambault, D., Peñáz, P., Zagler, W. (eds.) *Computers Helping People with Special Needs. Lecture Notes in Computer Science*, vol. 8548, pp. 156–163. Springer International Publishing (2014)
3. Chang, C.Y., Lange, B., Zhang, M., Koenig, S., Requejo, P., Somboon, N., Sawchuk, A., Rizzo, A.: Towards pervasive physical rehabilitation using microsoft kinect. In: 6th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth), May 2012, pp. 159–162
4. Iarlori, S., Ferracuti, F., Giantomassi, A., Longhi, S.: RGBD camera monitoring system for Alzheimer's disease assessment using recurrent neural networks with parametric bias action recognition. In: 19th World Congress The International Federation of Automatic Control, 24–29 Aug 2014, pp. 3863–3868
5. Shaughnessy, M., Resnick, B., Macko, R.: Testing a model of post-stroke exercise behavior. *Rehabil. Nurs.* **31**, 15–21 (2006)
6. Hanif, M., Niaz, H., Khan, M.: Investigating the possible role and usefulness of video capture virtual reality in motor impairment rehabilitation. In: The 2nd International Conference on Next Generation Information Technology (ICNIT), June 2011, pp. 23–30
7. Lin, T.Y., Hsieh, C.H., Lee, J.D.: A kinect-based system for physical rehabilitation: utilizing tai chi exercises to improve movement disorders in patients with balance ability. In: 7th Asia Modelling Symposium (AMS), July 2013, pp. 149–153

8. Chang, Y.J., Chen, S.F., Huang, J.D.: A kinect-based system for physical rehabilitation: a pilot study for young adults with motor disabilities. *Res. Dev. Disabil.* **32**(6), 2566–2570 (2011)
9. Da Gama, A., Chaves, T., Figueiredo, L., Teichrieb, V.: Poster: improving motor rehabilitation process through a natural interaction based system using kinect sensor. In: *IEEE Symposium on 3D User Interfaces (3DUI)*, March 2012, pp. 145–146
10. OpenNI: Open Natural Interaction. <http://structure.io/openni>
11. Villaroman, N., Rowe, D., Swan, B.: Teaching natural user interaction using OpenNI and the microsoft kinect sensor. In: *Proceedings of the 2011 Conference on Information Technology Education*, 2011, pp. 227–232
12. Cousins, S.: Exponential growth of ROS [ROS topics]. *Robot. Autom. Mag. IEEE* **18**(1), 19–20 (2011)
13. ROS: Robot Operating System. <http://www.ros.org>
14. Mrazovac, B., Bjelica, M., Papp, I., Teslic, N.: Smart audio/video playback control based on presence detection and user localization in home environment. In: *2nd Eastern European Regional Conference on the Engineering of Computer Based Systems (ECBS-EERC)*, Sept 2011, pp. 44–53
15. Rumelhart, D.E., Hinton, G.E., Williams, R.J.: Learning internal representations by error propagation. In: Rumelhart, D.E., McClelland, J.L. (eds.) *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, vol. 1, pp. 318–362. CORPORATE PDP Research Group. MIT Press, Cambridge, MA, USA (1986)
16. Haykin, S.: *Neural Networks: A Comprehensive Foundation*, 2nd edn. Prentice Hall, Englewood Cliffs (1999)
17. Møller, M.: A scaled conjugate gradient algorithm for fast supervised learning. *Neural Netw.* **6**(14), 525–533 (1993)