COMPUTER METHODS AND PROGRAMS IN BIOMEDICINE 110 (2013) 490–502

journal homepage: www.intl.elsevierhealth.com/journals/cmpb

# Design of an efficient framework for fast prototyping of customized human–computer interfaces and virtual environments for rehabilitation

*Danilo Avola, Matteo Spezialetti, Giuseppe Placidi**

*Department of Life, Health and Environmental Sciences, University of L'Aquila, Via Vetoio, 67100 L'Aquila, Italy*

ARTICLE INFO

*Article history:*
Received 25 October 2012
Received in revised form
10 January 2013
Accepted 13 January 2013

*Keywords:*
Rehabilitation
Gesture recognition
Human–computer interfaces
Time-of-Flight camera
Depth map
Virtual reality

ABSTRACT

Rehabilitation is often required after stroke, surgery, or degenerative diseases. It has to be specific for each patient and can be easily calibrated if assisted by human–computer interfaces and virtual reality. Recognition and tracking of different human body landmarks represent the basic features for the design of the next generation of human–computer interfaces. The most advanced systems for capturing human gestures are focused on vision-based techniques which, on the one hand, may require compromises from real-time and spatial precision and, on the other hand, ensure natural interaction experience. The integration of vision-based interfaces with thematic virtual environments encourages the development of novel applications and services regarding rehabilitation activities. The algorithmic processes involved during gesture recognition activity, as well as the characteristics of the virtual environments, can be developed with different levels of accuracy. This paper describes the architectural aspects of a framework supporting real-time vision-based gesture recognition and virtual environments for fast prototyping of customized exercises for rehabilitation purposes. The goal is to provide the therapist with a tool for fast implementation and modification of specific rehabilitation exercises for specific patients, during functional recovery. Pilot examples of designed applications and preliminary system evaluation are reported and discussed.



## 1. Introduction

Residual impairments are often present after stroke, head traumas, surgery, or in people suffering from orthopedic or degenerative diseases. In these cases, rehabilitation is necessary. The rehabilitation program has to be accurately studied and prepared by a therapist and the program is strongly dependent on several parameters (e.g., degree of residual impairments), whose values are always different for different patients and change with time; the residual impairment should decrease with time, if the rehabilitation is effective. This implies that both the rehabilitation protocol and its temporal modification have to be calibrated for each patient, needing continuous monitoring. For these reasons, traditional rehabilitation is done one-to-one, namely, one therapist (or sometimes several) working with one patient. Thus costs are high, especially for demanding patients, such as those with brain or post surgery injuries. There is currently no monitoring for the parts of the therapy that the patient does at home. Recent studies suggested that repetitive and long duration training using human–computer interfaces and virtual reality

* *Corresponding author*. Tel.: +39 0862433493; fax: +39 0862433633.
E-mail addresses: danilo.avola@univaq.it (D. Avola), matteo.spezialetti@univaq.it (M. Spezialetti), giuseppe.placidi@univaq.it (G. Placidi).

http://dx.doi.org/10.1016/j.cmpb.2013.01.009

is helpful both for physical and functional recovery [1,2], as is the continuous control, in an objective way, of the rehabilitation activity performed at home. Moreover, by using computer assisted rehabilitation, it can be easier to implement specific rehabilitation exercises, to modify their difficulty, or to consider other parameters.

The new tendency in movement-based human–computer interfaces is to enable users to manage devices or interact with artificial or virtual environments through free movement of the hands and/or parts of the body (e.g., legs, arms). The implementation of these interfaces requires considerable technical effort. The first task is the recognition of specific anatomic landmarks (e.g., feet, hands, head) that the system has to consider as an input for interaction. Another task concerns the movement interpretation analysis performed on the chosen body landmarks aimed at identifying patterns or specific poses.

The most advanced systems for capturing, managing, and interpreting human gestures adopt vision-based techniques to allow a natural interaction experience without using physical controllers or body suits. Fortunately, current improvements in Time-of-Flight (ToF) cameras [3] allow the gathering of a suitable set of information (i.e., Depth and RGB data maps) that facilitates video processing activities. Despite this, vision-based techniques require a careful implementation of the algorithms used to process data maps to minimize errors in recognition and tracking of the body landmarks. Often, spatial accuracy can be obtained by compromises on the real-time interaction. This last aspect is tightly tied to the applicative context: in the rehabilitation field it assumes a great importance while in others (e.g., the entertainment field) it is not crucial. Another factor influencing the real-time interaction is the level of detail used during gesture recognition, where the identification of a simple target (e.g., a hand pose) requires less complicated models and computations than a complex one (e.g., hand articulations and their accurate spatial position measurements).

The use of vision-based gesture interfaces to support 3D reconstruction of avatars representing the whole body (or part of it) is the first step toward the new human-oriented rehabilitation applications and services, where a patient can interact with real-based or imaginary environments simply using himself, thus increasing personal motivation. Also in this case, the specific applicative contexts define complexity, representativeness and characteristics both of the avatar and of the virtual environments.

This paper describes the architectural aspects of a framework for fast prototyping and implementation of vision-based gesture recognition exercises to be used in rehabilitation. The goal is to provide the therapist with a tool for fast creation, adaptation and modification of rehabilitation exercises to deal with specific impairments affecting specific patients. The framework is composed of two main modules. The first deals with acquiring, filtering and managing spatial information of the human targets that will be passed to the second module. The second module serves to create and lead the pilot virtual environments. We start from our previous experience in the rehabilitation field [4–7] to extend and generalize the architecture by including the rehabilitative exercises of others (hereinafter ‘tasks’). The framework is designed to support different interactive scenarios in a smart and easy way. The experimental sessions performed according to ad hoc virtual environments and qualitative observations of the related interactive behavior led to the conclusion that the system is robust and reliable.

The paper is structured as follows. Section 2 discusses related work on different aspects of the gesture recognition systems and the related integration of virtual environments. Section 3 presents the framework architecture, highlighting the management of data maps and the creation of virtual environments modules. Section 4 shows three basic pilot examples and a preliminary experimental evaluation of the system. Section 5 concludes the paper and proposes some improvements to the system.

## 2. Related work

Gesture recognition can be classified according to the use or not of physical controllers. In the first case, advanced haptic interfaces (e.g., mechanical gloves, sensors-based body suits) are used to catch user movements [8,9]. These kinds of interfaces produce high spatial precision and do not require a segmentation step, since position and pose are derived from sensors installed on the device. Moreover, haptic interfaces can easily support the tracking step using ad hoc sensors, thus efficiently solving occlusion problems [10]. On the other hand, their limited usability, portability and adaptability make these systems expensive and obsolete. In addition, they are not recommended in rehabilitation because their encumbrance greatly limits user movements [11,12]. In other contexts (e.g., simulator based games) the use of controllers, to provide feedback to the user, is strongly recommended. For these reasons, vision-based systems are assuming always greater importance in each application area. Among these, those based on Sony Move Controller technology [13] still require haptic interfaces to enable users to manage gestures. Conversely, other systems like those based on Microsoft Kinect technology [14], are wholly driven by the visual information coming from ToF cameras. Our intention was to base the proposed framework on the last mentioned technology to provide an alternative, customizable and open system to design a general set of rehabilitative tasks. With this aim, we adopted guidelines which differed from those published elsewhere; only visual information was exploited to derive poses and gestures of users. The first set of publications, [15,16], presented skin color based methods to locate and track human body landmarks within a video stream. They faced problems with the variation of the colors due to different lighting conditions. In [17] the authors exploited depth maps to implement a model-driven approach for estimating human poses from the position of some anatomical landmarks, detected and tracked through probabilistic inference. A related approach is described in [18] where a novel method for facilitating the development of gesture recognition interfaces was defined. The author’s framework used both depth maps and greyscale image sequences to generate a set of gesture based controls. Another interesting approach to deriving gesture based interfaces is reported in [19], where a method for human full body pose estimation from depth data was presented. This