Available online at www.sciencedirect.com

ScienceDirect

Procedia Manufacturing 11 (2017) 91 – 98

Procedia MANUFACTURING

27th International Conference on Flexible Automation and Intelligent Manufacturing, FAIM2017, 27-30 June 2017, Modena, Italy

Human behavior and hand gesture classification for smart human-robot interaction

Nuno Mendes, João Ferrer, João Vitorino, Mohammad Safeea, Pedro Neto*

University of Coimbra, Department of Mechanical Engineering POLO II, 3030-788 Coimbra, Portugal

Abstract

This paper presents an intuitive human-robot interaction (HRI) framework for gesture and human behavior recognition. It relies on a vision-based system as interaction technology to classify gestures and a 3-axis accelerometer for behavior classification (stand, walking, etc.). An intelligent system integrates static gesture recognition recurring to artificial neural networks (ANNs) and dynamic gesture recognition using hidden Markov models (HMM). Results show a recognition rate of 95% for a library of 22 gestures and 97% for a library of 6 behaviors. Experiments show a robot controlled using gestures in a HRI process.

Peer-review under responsibility of the scientific committee of the 27th International Conference on Flexible Automation and Intelligent Manufacturing

Keywords: Human Robot Interaction; Human Behavior Recognition; Gestures; Segmentation; Accelerometer

1. Introduction

In a scenario in which robots and humans share the same environment and cooperate with each other, robotic systems need to know the human behavior to predict future actions and support decision making. Understanding human activities in the context in which they take place is challenging, since a person's activity is driven by goals,

* Corresponding author. Tel.: +351 239 790 700; fax: +351 239 790 701.
E-mail address: pedro.neto@dem.uc.pt

Peer-review under responsibility of the scientific committee of the 27th International Conference on Flexible Automation and Intelligent Manufacturing
doi:10.1016/j.promfg.2017.07.156

motives and needs that may be conflicting [1–4]. The HRI process is enriched with information of the human behavior, for example the robot velocity can be adjusted according to if the human is static or moving.

Intuitive HRI, for example using gestures, allows humans to focus on their own work and not in the robot programming. A robot can be instructed by a worker in the shop floor by natural means, using gestures and speech for example. In order to achieve this goal, robots should be prepared to be instructed with a high-level of abstraction from the robot language. Most of the approaches to HRI rely on a close mimicking of human-human communication using gestures, speech and the most diverse human actions. There are a number of different interaction technologies for HRI, namely data gloves [5,6], magnetic sensors [7,8], inertial sensors, Electromyography (EMG) [9], vision sensors [10–14] and hybrid solutions.

Gestures can be classified in three categories, the static gestures, the dynamic gestures and movement epenthesis. The static ones are easier to recognize because of its nature where just a frame of data (features) is required to identify them [15]. Thus, simple models of ANN have been proposed to perform an effective recognition [16]. The segmentation of static gestures is easier to be carried out than the segmentation of dynamic gestures. The dynamic gestures present some big challenges due to its spatial-temporal variability. A same gesture performed by different users can differ in shape, velocity, duration, and integrality [17]. In order to recognize dynamic gestures different approaches have been employed, for example using discrete HMM to recognize online dynamic human hand gestures [18] and using HMM for full body gesture recognition [19]. Recognitions rates above 84% were reported for a collection of seven dynamic gestures. Automatic recognition of facial emotion based on feed forward ANN and support vector regressors was presented by [20]. An interesting study in the field reports a neural-based classifier for gesture recognition with an accuracy of over 99% for a library of 6 gestures [21]. Other authors concluded that a hand contour-based neural network training is faster than complex moment-based neural network training but in the other hand the former proved to be less accurate (71%) than the latter (86%) [22]. User detection in activity and location patterns was studied by [23] who proposed a fusion-based HMM architecture.

2. Proposed approach

This paper proposes a HRI framework in which human behaviors and hand gestures are recognized and used to tele-operate a robot. The use of an accelerometer to collect representative information about human behavior is proposed. One of the goals of this study is to make use of a minimal number of sensors and at the same time being a low-cost solution to identify human behaviors that are used to adapt the way a robot performs its work. The robot can reduce its speed, stop, or move to a secure place depending on identified human behavior. Six behaviors were considered: walking, running, jumping, stand-to-sit, standing and seated.

A gesture spotting solution using an infrared stereo vision system as interaction technology, the Leap Motion Controller (LMC), is also proposed. Gesture patterns are recognized in continuous (not separately) recurring to ANNs and HMMs specifically adapted to the process of controlling an industrial robot (supervised learning). In continuous gesture recognition, communicative gestures (with an explicit meaning) appear intermittently with non-communicative gestures (transition gestures, emotional expressions, idling motion, etc.), with no specific order. In this way, it is proposed an ANN-based architecture to recognize static gestures and HMM-based architecture to recognize dynamic gestures. Experimental results demonstrated that the proposed solution presents high RRs, low training and learning time, a good capacity to generalize, it is intuitive to use and allows robot operation independently from the conditions of the surrounding environment. Fig. 1 shows a scheme of the proposed system.

2.1. Data collection

In order to collect communicative information from human to the robot, two kind of sensors are used: (1) an accelerometer to collect human behavior and an infrared vision system for collection of hand gestures. The acceleration data was collected using an Analog Devices ADXL330 tri-axial accelerometer. It measures over a range of at least +/- 3g and it yields a sensitivity accuracy of 10%. This sensor was connected to a computer over Bluetooth. Hand gestures are captured by an infrared vision system, which also converts images in hand shape features, the Leap Motion Controller (LMC).