Engineering Applications of Artificial Intelligence 66 (2017) 33–48

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai

Gesture recognition system for real-time mobile robot control based on inertial sensors and motion strings

Ivo Stančić, Josip Musić *, Tamara Grujić

Faculty of Electrical Engineering, Mechanical Engineering and Naval Architecture, University of Split, Ruđera Boškovića 32, 21000 Split, Croatia

A R T I C L E I N F O

Keywords:
Human–robot interaction
Mobile robot control
Inertial sensors
Machine learning
Real-time classification
Hand gestures

A B S T R A C T

Navigating and controlling a mobile robot in an indoor or outdoor environment by using a range of body-worn sensors is becoming an increasingly interesting research area in the robotics community. In such scenarios, hand gestures offer some unique capabilities for human–robot interaction inherent to nonverbal communication with features and application scenarios not possible with the currently predominant vision-based systems. Therefore, in this paper, we propose and develop an effective inertial-sensor-based system, worn by the user, along with a microprocessor and wireless module for communication with the robot at distances of up to 250 m. Possible features describing hand-gesture dynamics are introduced and their feasibility is demonstrated in an off-line scenario by using several classification methods (e.g., random forests and artificial neural networks). Refined motion features are then used in K-means unsupervised clustering for motion primitive extraction, which forms the motion strings used for real-time classification. The system demonstrated an $F1$ score of 90.05% with the possibility of gesture spotting and null class classification (e.g., undefined gestures were discarded from the analysis). Finally, to demonstrate the feasibility of the proposed algorithm, it was implemented in an Arduino-based 8-bit ATmega2560 microcontroller for control of a mobile, tracked robot platform.

1. Introduction

Autonomous mobile robots have the capability of moving, that is, not being restrained to one physical location. Despite their autonomy, in applications where mobile robots are used, the ability to control or direct a mobile robot by a professional operator is sometimes essential. Controllers that are commonly used to control mobile robots are cumbersome, heavy, and, owing to their design, occupy the operator's hands, precluding the performance of other tasks. Gestural interfaces enable a fast and accurate interaction between users and computers through the motion of the body, face, and/or hands. Hand gestures are the natural way of nonverbal communication between humans in many cultures, and usually a small set of gestures is used.

Gestures can be characterised by trajectories of the limbs or body key points in space, which may be recorded by a variety of devices. Two groups of devices are commonly used: vision-based and inertial-sensors-based. Vision-based gestural systems use computer vision to track, identify, and interpret the gestures, while not requiring the users to wear sensors on their hands or body. The major drawbacks of most vision-based systems is their inability to track people beyond the camera's field of view and their sensitivity to changes in lightning conditions. Moreover, they require a high computational power, which is usually not possible with simple and affordable wearable systems. This makes them inappropriate for the majority of in-field applications.

To work in unconstrained environments, a body-worn sensor system would be more acceptable than a vision-based system. However, using specialised instrumented glove-like devices requires wearing cumbersome equipment and carrying cables that connect the device to a microcontroller or computer. Although body-worn sensors embedded inside instrumented gloves may obstruct the natural operator's hand movement, miniaturisation of sensors makes them well suited for demanding in-field use. The development of small sensor platforms from relatively inexpensive components has created new opportunities for a novel human–robotic interface design.

2. Literature review

Vision-based systems have previously been used for human body tracking, and a number of advanced algorithms have been developed for tracking of body segment motions (Bandera et al., 2009; Stancic et al., 2012). Numerous researchers have used visual identification of

* Corresponding author.
E-mail addresses: istancic@fesb.hr (I. Stančić), jmusic@fesb.hr (J. Musić), tada@fesb.hr (T. Grujić).

http://dx.doi.org/10.1016/j.engappai.2017.08.013
Received 2 September 2015; Received in revised form 1 March 2017; Accepted 21 August 2017
0952-1976/

gestures for robot control (Faudzi et al., 2012; Koceski and Koceska, 2010; Hasanuzzaman et al., 2007; Eliav et al., 2011; Elakkiyaa et al., 2012; Park and S-Lee, 2011; Nickel and Stiefelhagen, 2007; Waldherr et al., 2000). The benefit of this approach is in avoiding any equipment placed on the user's body, but, as a consequence, they have all the drawbacks that are associated with vision-based systems. Recently, body tracking research has been complemented by advanced RGB-D sensors such as Microsoft's Kinect (Palacios et al., 2013), which allows easier identification and tracking of the human body (Fardana et al., 2013; Liu et al., 2014; Libardi et al., 2014; Xu et al., 2015). Sign language has always been a popular topic in gesture recognition, as it offers a predefined set of gestures that could be identified by a vision system (McGuire et al., 2004).

As alternative to vision-based systems, current technology uses devices, such as data gloves or body-worn sensors (usually inertial sensors), which in some scenarios may limit user hand movements (Amma et al., 2013; Iba et al., 1999; Xu et al., 2012; Guenterberg et al., 2009). As an example, human body motion capture systems based on inertial sensors were presented in Benbasat and Paradiso (2002), Stiefmeier et al. (2007), Babu (2014) and Kumar and Onesim (2014). Simultaneous utilisation of both a body-worn inertial sensor and a depth sensor has the benefit of improved usability (Liu et al., 2014). In Liu's paper (Liu et al., 2014), an average recognition rate of 88% was obtained on a sample of 10 subjects by using a system based solely on inertial sensors, while inclusion of a Kinect device improved the accuracy to 93%.

For in-home use, some authors have successfully implemented inertial sensors inside handheld controllers (Wu et al., 2013) or smartphones (Wang et al., 2012). It is worth noting that not only inertial sensors can be utilised as appropriate body-worn sensors for gesture recognition. As an example, forearm electromyography (EMG) has been shown to provide accurate representations of hand movements for robot control (Wolf et al., 2013).

In research related to human–robot interaction, successful application of hand gestures to control robotic service tasks (Fujii et al., 2014; Yin and Xie, 2007), robot manipulators (Khassanov et al., 2014; Wang et al., 2013), or humanoid robots (Riek et al., 2010) has been reported. A cooperative surgical robot system, guided by hand gestures and supported by an augmented reality, has been successfully implemented in Wen et al. (2014).

Body-worn inertial sensors have been used for human activity recognition for years (Junker et al., 2008; Stiefmeier et al., 2007; Zhang and Sawchuk, 2012; Xu et al., 2012; Huỳnh et al., 2007; Krause et al., 2013). Several methods have been utilised for motion identification and classification. The hidden Markov model (HMM) is one of the oldest methods and appears in several research studies (Kao and Fahn, 2011; Park and S-Lee, 2011; Nickel and Stiefelhagen, 2007). Recently, a new method has been proposed in which human motion is modelled with motion primitives, as opposed to the more frequent approach that relies on the whole motion model (Stiefmeier et al., 2007; Zhang and Sawchuk, 2011b). In a motion-primitives-based model (Ijspeert et al., 2013), each activity is represented as a sequence of motion primitives, which acts as the smallest unit that describes motion. If an adequate motion or activity model (which describes it) is present, the motion can be detected and recognised from sensor data by analysing strings of motion primitives and comparing them with the "motion template". The detection and recognition parts are straightforward to implement, as they are usually combinations of existing string-matching and histogram-matching techniques. An additional challenge is the selection of the sensor data or the features used for building a motion primitive model (Zhang and Sawchuk, 2012). This work is in some ways similar to ours, but, because mostly cyclic activities are considered and not (discrete) gestures, it uses different features sets that require the use of more complex algorithmic operations. The approach was tested in an off-line manner (with emphasis on accuracy and not computational cost) with nine human activities to be classified (walk forwards, walk left, walk right, go upstairs, go downstairs, run forwards, jump up, sit on a chair, and stand). The obtained results demonstrated a 92.7% classification accuracy, with motion vocabulary consisting of 125 motion primitives.

A similar approach, but one with on-line implementation (on a desktop computer), was given in Stiefmeier et al. (2007). Here the authors represented activities (23 of them during a bicycle maintenance task) as strings and used trajectory calculation and aggregation with string matching for classification. String matching was done to calculate how many symbols needed to be substituted to obtain the required string. The system was tested with three subjects and had an accuracy of 82.7% for spotting and classification of gestures.

2.1. Manuscript contributions

Our aim is to design and develop a system that would be accurate enough and provide the correct spotting and classification of hand gestures in real time and could be implemented on a low-power microcontroller. To achieve that goal, we propose a method for real-time detection and classification of features that describe each performed gesture by a string of motion primitives that encode hand dynamics. The motion primitives are inspired by their similarity with human speech signals (Ghasemzadeh et al., 2008). In human speech recognition, sentences are first divided into words, which are then divided into a sequence of phonemes. A similar approach could be used to divide complex hand movements into a sequence of motion primitives (Ijspeert et al., 2013). The most important component of the proposed framework is based on the Bag-of-Features (BoF) (Zhang and Sawchuk, 2012), which builds hand motion models by using strings of symbols. We have also explored several features sets and proposed the one that offers the best possibility for efficient real-time implementation.

Thus, the paper's contributions can be summarised as follows:

- the development and application of a motion primitives vocabulary for discrete hand motion modelling based on data from inertial sensors;
- the development and analysis of features set(s) used in building a motion primitive vocabulary;
- the development and implementation of a novel string- and histogram-matching method for gesture spotting and classification in real time; and
- application of the proposed method on an 8-bit microcontroller for mobile robot control by using hand gestures.

3. Materials and methods

To achieve the defined goals, several intermediate aims were defined and implemented: 1. the construction of an experimental measurement unit, 2. the recording of hand gestures by using the developed measurement unit on a large number of subjects, 3. off-line analysis of the recorded gestures and identification of appropriate signal features, 4. off-line classification of the gestures across all subjects to demonstrate the feasibility of the used gestures and features, 5. development of a real-time classification algorithm based on conclusions from the previous step, and 6. implementation of the real-time algorithm to mobile robot control for final verification of the proposed approach. In the remainder of the section, details for each intermediate goal and how these goals were achieved will be presented and explained.

3.1. Experimental setup

The experimental setup consisted of three segments: a microcontroller module (Fig. 1a), a hand-worn sensor module (Fig. 1b), and a mobile robot (Fig. 1c).

The mobile robot used in the experiments (Fig. 1c) was a DFRobotShop rover, Arduino-compatible tracked robot with a differential drive. It is able to move in all terrains, as it is driven by a pair of continuous