ARTICLE IN PRESS
JID: YCVIU [m5G;March 25, 2016;7:34]

Computer Vision and Image Understanding 000 (2016) 1–13

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

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

A real-time Human-Robot Interaction system based on gestures for assistive scenarios

Gerard Canal [a,b,d,*], Sergio Escalera [c,a], Cecilio Angulo [b]

[a] Computer Vision Center, Campus UAB, Edifici O, 08193 Bellaterra (Cerdanyola), Barcelona, Spain
[b] Dept. Automatic Control, UPC - BarcelonaTech, FME Building, Pau Gargallo 5, 08028 Barcelona, Spain
[c] Dept. Matemàtica Aplicada i Anàlisi, UB, Gran Via de les Corts Catalanes 585, 08007 Barcelona, Spain
[d] Institut de Robòtica i Informàtica Industrial, CSIC-UPC, Llorens i Artigas 4-6, 08028 Barcelona, Spain

ARTICLE INFO

Article history:
Received 15 April 2015
Accepted 2 March 2016
Available online xxx

Keywords:
Gesture recognition
Human Robot Interaction
Dynamic Time Warping
Pointing location estimation

ABSTRACT

Natural and intuitive human interaction with robotic systems is a key point to develop robots assisting people in an easy and effective way. In this paper, a Human Robot Interaction (HRI) system able to recognize gestures usually employed in human non-verbal communication is introduced, and an in-depth study of its usability is performed. The system deals with dynamic gestures such as waving or nodding which are recognized using a Dynamic Time Warping approach based on gesture specific features computed from depth maps. A static gesture consisting in pointing at an object is also recognized. The pointed location is then estimated in order to detect candidate objects the user may refer to. When the pointed object is unclear for the robot, a disambiguation procedure by means of either a verbal or gestural dialogue is performed. This skill would lead to the robot picking an object in behalf of the user, which could present difficulties to do it by itself. The overall system — which is composed by a NAO and Wifibot robots, a Kinect™ v2 sensor and two laptops — is firstly evaluated in a structured lab setup. Then, a broad set of user tests has been completed, which allows to assess correct performance in terms of recognition rates, easiness of use and response times.

1. Introduction

Autonomous robots are making their way into human inhabited environments such as homes and workplaces: for entertainment, helping users in their domestic activities of daily living, or helping disabled people in personal care or basic activities, which would improve their autonomy and quality of life.

In order to deploy such robotic systems inhabiting unstructured social spaces, robots should be endowed with some communication skills so that users can interact with them just as they would intuitively do, eventually considering a minimal training. Besides, given that a great part of the human communication is carried out by means of non-verbal channels [1,2], skills like gesture recognition and human behavior analysis reveal to be very useful for this kind of robotic systems, which would include viewing and understanding their surroundings and the humans that inhabit them.

Gesture recognition is an active field of research in Computer Vision that benefits from many machine learning algorithms, such as temporal warping [3–5], Hidden Markov Models (HMMs), Support Vector Machines (SVMs) [6], random forest classifiers [7] and deep learning [8], just to mention a few of them. Moreover, gesture recognition personalization techniques have also been proposed in [9] to adapt the system to a given user. Studies in Human Computer Interaction (HCI) and more specifically Human Robot Interaction (HRI) take advantage of this field. Hence, many recent contributions [10–14] consider Kinect™-like sensors to recognize gestures given the discriminative information provided by multimodal RGB-Depth data. A Kinect™ based application is introduced in [15] for taking order service of an elderly care robot. Static body posture is analyzed by an assistive robot in [16] to detect whether the user is open towards the robot interaction or not. Communicative gestures are contrasted from daily living activities in [17] for an intuitive human robot interaction. A novice user can generate his/her gesture library in a semi-supervised way in [18], which are then recognized using a non-parametric stochastic segmentation algorithm. In [19], the user can define specific gestures that mean some message in a human-robot dialogue, and in [20] a framework to define user gestures to control a robot is presented. Deep neural networks are used in [21] to recognize gestures in real time by considering only RGB information. Pointing gestures, similar to the one we propose in this paper, have been studied mostly focusing

* Corresponding author.
E-mail addresses: gerard.canal@cvc.uab.cat (G. Canal), sergio@maia.ub.es (S. Escalera), cecilio.angulo@upc.edu (C. Angulo).

http://dx.doi.org/10.1016/j.cviu.2016.03.004
1077-3142/

in hand gestures [22], using the hand orientation and face pose [23]. The pointing direction is estimated in [24,25] using gaze and finger orientation, and deictic gesture interactions that people use to refer to objects in the environment are studied in [26]. Related pointing interactions have also been used for robot guidance [27].

In this work we introduce a real time Human Robot Interaction (HRI) system whose objective is to allow user communication with the robot in an easy, natural and intuitive gesture-based fashion. The experimental setup is composed by a humanoid robot (Aldebaran's NAO) and a wheeled platform (Wifibot) that carries the NAO humanoid and a Kinect™ sensor. In this set-up, the multi-robot system is able to recognize static and dynamic gestures from humans based on geometric features extracted from biometric information and dynamic programming techniques. From the gesture understanding of a deictic visual indication of the user, robots can assist him/her in tasks such as picking up an object from the floor and bringing it to the user. In order to validate the system and extract robust conclusions of the interactive behavior, the proposed system has been tested in offline experiments, reporting high recognition rates, as well as with an extensive set of user tests in which 67 people assessed its performance.

The remainder of the paper is organized as follows: Section 2 introduces the methods used for gesture recognition and Human Robot Interaction. Section 3 presents the experimental results including the offline and user tests and, finally, Section 4 concludes the paper.

2. Gesture based Human Robot Interaction

With the aim to study gestural communication for HRI, a robotic system has been developed able to understand four different gestures so a human user can interact with it: wave (hand is raised and moved left and right), pointing at (with an outstretched arm), head shake (for expressing disagreement) and nod (head gesture for agreement).

The overall robotic system involves several elements: an Aldebaran's NAO robot, a small size humanoid robot which is very suitable to interact with human users; a Microsoft's Kinect™ v2 sensor to get RGB-Depth visual data from the environment and track the user; and, given that the vision sensor exceeds NAO's robot capabilities (in size and computing performance), a Nexter Robotics' Wifibot wheeled platform is used to carry the sensor as well as the NAO, easing its navigation and precision at long ranges.

In fact, the proposed robotic system takes inspiration from the DARPA Robotics Challenge 2015[1] in which a humanoid robot should drive a car towards an interest place and exit the car in order to finish its work by foot. In a similar way, the wheeled robot was added to the system in order to carry the sensor along with the little humanoid, which should also exit it to complete its task by walking. This multi-robot setup allows the NAO to use the information from the Kinect's™ v2 sensor and eases its navigation. And for its side, the NAO is the one in charge of directly interacting with the user, also being able to act on the environment, for instance, by grasping objects. The overall setup is shown in Fig. 1, with the NAO seated on the Wifibot. The setup also includes a laptop with an Intel i5 processor to deal with Kinect™'s data and another Intel Core 2 Duo laptop, which sends commands to the robots using the Robot Operating System (ROS)[2] [28]. The depth maps are processed using the Point Clouds Library (PCL)[3] [29], and body tracking information is obtained using the Kinect™ v2 SDK.

[1] theroboticschallenge.org
[2] ros.org
[3] pointclouds.org

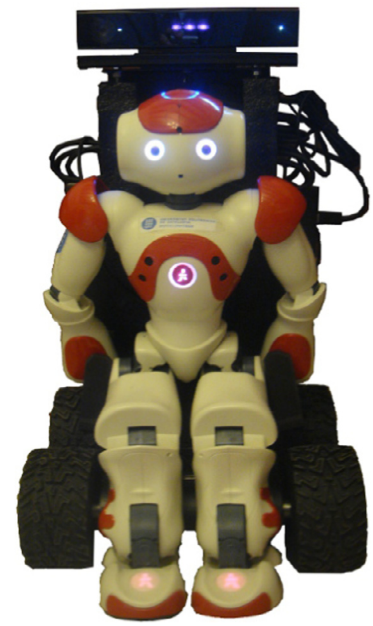

Fig. 1. The robotic system designed for this work.

The system has been programmed as an interactive application, and tested with several users of different ages and not related with the robotics world (see Section 3.2).

2.1. Real time gesture recognition: Interaction with a robot

This section explains the methods used to perform the gesture recognition and image understanding. Given that the application of the system is to enhance the interaction between a human user and a robot, the defined gestures should be as natural for the user as possible, avoiding user training or learning of a specific set of gestures. Instead, the robot should understand gestures as a human would understand another human's gestures, and should reply to that visual stimulus in real time.

The considered set of human gestures has been divided into two categories, depending on the amount of movement involved in their execution:

- Static gestures are those in which the user places his/her limbs in a specific position and stands for a while, without any dynamics or movement involved. In this case, the transmitted information is obtained through the static pose configuration. Pointing at an object is an example of static gesture.
- Dynamic gestures are, in contrast, those in which the movement is the main gesture's feature. The transmitted information comes from the type of movement as well as its execution velocity. It may also contain a particular pose for a limb during the movement. Examples of dynamic gestures are a wave to salute someone or a gesture with the hand to ask someone to approach to the user's location.

Four different gestures have been included in the designed system to interact with the robot, being three of them dynamic and the remaining one static. The dynamic gestures are the wave, the nod and a facial negation gesture. The static one is the pointing at an object. Both categories are tackled using different approaches. Next we describe the extracted features, the gesture recognition methods and how the gesture's semantic information is extracted.

2.1.1. Definition of gesture specific features

Gesture recognition is performed based on some features extracted from the user body information obtained from depth maps.