Available online at www.sciencedirect.com

ScienceDirect

IFAC-PapersOnLine 49-19 (2016) 126–131

Gaze Trajectory Prediction in the Context of Social Robotics

François de La Bourdonnaye*, Rossitza Setchi,**
Cecilia Zanni-Merk*,***

**INSA de Strasbourg, 24 Boulevard de la Victoire, 67048 Strasbourg cedex, France*
***Cardiff University, Queen's buildings, 14-17 The Parade, Cardiff CF24 3AA, UK*
**** ICube Laboratory, 300 Boulevard Sébastien Brant, 6742 Ilkirch cedex, France*

Abstract: Social robotics is an emerging field of robotics that focuses on the interactions between robots and humans. It has attracted much interest due to concerns about an aging society and the need for assistive environments. Within this context, this paper focuses on gaze control and eye tracking as a means for robot control. It aims to improve the usability of human–machine interfaces based on gaze control by developing advanced algorithms for predicting the trajectory of the human gaze. The paper proposes two approaches to gaze-trajectory prediction: probabilistic and symbolic. Both approaches use machine learning. The probabilistic method mixes two state models representing gaze locations and directions. The symbolic method treats the gaze-trajectory prediction problem similar to how word-prediction problems are handled in web browsers. Comparative experiments prove the feasibility of both approaches and show that the probabilistic approach achieves better prediction results.

Keywords: Social robotics, human–robot interaction, eye tracking, gaze tracking, machine learning, trajectory prediction, time series, teleoperation

1. INTRODUCTION

Recent advances in domestic and humanoid robots have transformed robotics from an area primarily concerned with industrial automation to a field that supports social interactions between humans and robots. Social robotics is an emerging field that focuses on the interactions between robots and humans. It is the study of robots that interact and communicate among themselves, with humans, and with the environment, within the social and cultural structure attached to their roles (Ge and Mataric, 2009). The challenge is to overcome the existing human–robot barrier by constructing robots that behave more like humans and understand commands in an intuitive way.

The area of human–robot interaction benefits from a large range of sensors, including cameras, microphones, laser reading, and tactile sensors (Haibin, 2014). These sensors can be used for different perception tasks, including emotion recognition, object detection, face recognition, and human motion tracking. The scope of this paper is robot control using eye tracking. The use of gaze in the context of social robotics has been widely explored over the past decade. The use of gaze in teleoperation (Latif et al., 2008) and in the context of humanoid robots (Dickstein-Fischer et al., 2011) are some of the most well-known applications.

The aim of this paper is to improve existing teleoperation interfaces by developing advanced algorithms for gaze trajectory prediction. Teleoperation consists of controlling a robot where a human uses his/her gaze to draw a trajectory that is then executed by the robot. Such a scenario is based on a domestic environment where a disabled or elderly person controls a robot with his or her gaze. Gaze prediction involves communicating and understanding commands based on a portion of a trajectory. A person does not need to complete the whole trajectory but only to start drawing its beginning for the robot to understand the entire trajectory. The approach is based on comparing real-time eye-tracking data with pre-recorded classes of gaze trajectory.

The remainder of the paper is organized as follows. Section 2 reviews the use of eye tracking in social robotics and discusses relevant approaches to gaze trajectory prediction. Section 3 highlights the research objectives and introduces the teleoperation principle as well as the concept of gaze trajectory prediction. Section 4 outlines the algorithms developed to predict gaze trajectories. Section 5 presents the experiments used to validate the prediction algorithms. The results are discussed in Section 6. The final section concludes the paper.

2. LITERATURE REVIEW

2.1 Eye Tracking in the Context of Social Robotics

The use of eye-tracking systems originates in the 1990s with the first eye-wink interfaces designed to assist severely disabled persons with their everyday activities (Shaw et al., 1990; Crisman et al., 1991). More recent examples of using eye tracking include teleoperation in surgery applications (Staub et al., 2012), navigation, and exploration (Yu et al., 2014; Latif et al., 2008, 2009a, b). In addition, eye-tracking systems are also used as therapeutic tools. For instance, Dickstein-Fischer et al. (2011) developed a humanoid robot designed to diagnose autism and interact with autism spectrum disorder (ASD) persons so that the robot could improve their social behavior. Another example is the gaze-sensitive virtual social interactive system for ASD children designed by Lahiri et al. (2011). The system uses a computer screen (instead of a humanoid robot) to interact in real time with ASD children.

Peer review under responsibility of International Federation of Automatic Control.
10.1016/j.ifacol.2016.10.473

More sophisticated interfaces recently developed take into account gaze direction. For instance, the gaze communication system for amyotrophic lateral sclerosis (ALS) (Maehara et al., 2003) uses a simple charge-coupled device (CCD) camera that tracks the user's gaze on a screen. In another advanced solution, the user uses his/her gaze to generate a trajectory and move a wheelchair in the real world (Antonya et al., 2011).

2.2 Gaze Prediction

Gaze prediction is a term with different meanings depending on the context of its use. It may refer to techniques used to improve eye tracking (Han, 2013); given eye-tracking data, it reduces the lag between the acquisition of the measure and its display. In this context, improving the prediction leads to improving the quality of the eye tracking.

This term can also be used in the context of egocentric videos produced by wearable cameras, such as GoPro (Li et al., 2013). The gaze-fixation prediction is computed given the wearer's head motion, hand location, and a dynamic model of the gaze.

In addition, the term is used in interactive media applications, such as bit allocation in streaming video based on region-of-interest (Feng et al., 2013). The approach uses a gaze-prediction system based on the Hidden Markov Model (HMM) where the states correspond to two of a human's intrinsic gaze behavioral movements (saccades and fixations). The parameters of the model are derived off-line from the visual saliency maps of the video. The principle of bit allocation is to allocate more bits to the regions of interest and less to other spatial regions. This is achieved so that both video compression and visual quality are improved. In exactly the same context, another gaze-location prediction application has been developed for broadcast football video (Cheng et al., 2013). The method employs a Bayesian integration of bottom-up features (motion measurements and saliency) and top-down information (ball, players, shot-type label).

In contrast to previous approaches, the approach developed in this paper considers the gaze-prediction problem to be similar to time-series prediction. Instead of using image-processing techniques, it applies machine learning on eye-tracking data to predict the human gaze trajectory.

2.3 Trajectory Prediction

The trajectory-prediction problem has been investigated from two different perspectives. First, gaze-trajectory prediction can be seen as a time-series prediction problem. A survey of methods for long-term prediction by Hellbach et al. (2009) compares several time-series prediction methods applied to prediction problems in the field of mobile robots and human–robot interaction. These include autoregressive models, local modeling, cluster-weighted modeling, and echo state networks. Experiments show that echo state networks (Yao et al., 2013) and local modeling (Oh et al., 2003) produce better results for long-term motion prediction.

Second, probabilistic methods, such as Markov Chain Models (Ishikawa et al., 2004) and Hidden-Markov Models (Qiao et al., 2015) are widely used for trajectory prediction of moving objects and can be applied to this problem. This paper uses these methods in the development of the probabilistic algorithm.

3. CONCEPTUAL MODEL

Gaze control involves trajectory classification and trajectory prediction. This paper focuses on trajectory prediction. This section briefly describes teleoperation from the eye-data processing methods to the control of the robot. It then outlines the principle of gaze-trajectory prediction.

3.1 Teleoperation

As shown in Fig. 1, teleoperation consists of the execution by the robot of a scaled version of the trajectory drawn on a screen by a human user. The trajectory is displayed in real time on the computer screen.

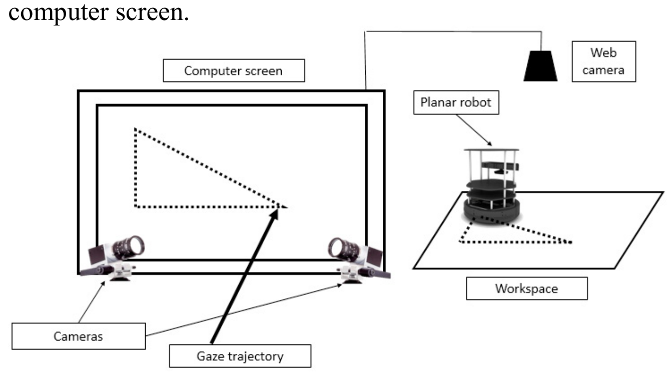

Fig. 1. Teleoperation scheme

The acquisition method is based on cascading three algorithms:

- The first algorithm filters the raw eye-tracking data. A double exponential filter is chosen because it is a good trade-off between precision and implementation simplicity. Furthermore, the fact that it is an infinite impulse response (IIR) filter makes it suitable for real-time applications.
- The second is a cartographic algorithm (Zhao and Saalfeld, 1997). This routine reduces the number of points and still conserves the topology of the trajectory. It is used to make the teleoperation more practical; indeed, the robot does not have to consider too many points.
- The third algorithm is a fixation-detection algorithm. A fixation happens when the user stares at a specific area. It produces an undesirable noise that cannot be removed by the double exponential filter. The fixation-detection algorithm detects the fixations and removes the associated noise. This paper uses a fixation-detection method based on a dispersion threshold (Salvucci and Goldberg, 2000).

The three algorithms are implemented in a cascade manner in real-time operation. Note that the method handles 3 known eye movement patterns: saccades, smooth pursuit movements, and fixations. Saccades generate rectilinear trajectories. Smooth pursuit movements imply more complex trajectories. Fixations are handled as explained before. When the trajectory is entirely