Computers in Human Behavior 68 (2017) 450–455

Contents lists available at ScienceDirect

Computers in Human Behavior

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

Full length article

Aggregating physiological and eye tracking signals to predict perception in the absence of ground truth

Enkelejda Kasneci [a,*], Thomas Kübler [a], Klaus Broelemann [b], Gjergji Kasneci [b]

[a] University of Tübingen, Department of Computer Science, Perception Engineering Group, Sand 14, 72076, Tübingen, Germany
[b] Schufa Innovation Lab, Kormoranweg 5, 65201, Wiesbaden, Germany

ARTICLE INFO

Article history:
Received 25 July 2016
Received in revised form
28 November 2016
Accepted 30 November 2016
Available online 2 December 2016

Keywords:
Eye tracking
Hazard perception
Physiological signals
Response aggregation
Eye tracking
Pupil dilation
ECG
Galvanic skin response

ABSTRACT

Today's driving assistance systems build on numerous sensors to provide assistance for specific tasks. In order to not patronize the driver, intensity and timing of critical responses by such systems is determined based on parameters derived from vehicle dynamics and scene recognition. However, to date, information on object perception by the driver is not considered by such systems. With advances in eye-tracking technology, a powerful tool to assess the driver's visual perception has become available, which, in many studies, has been integrated with physiological signals, i.e., galvanic skin response and EEG, for reliable prediction of object perception.

We address the problem of aggregating binary signals from physiological sensors and eye tracking to predict a driver's visual perception of scene hazards. In the absence of ground truth, it is crucial to use an aggregation scheme that estimates the reliability of each signal source and thus reliably aggregates signals to predict whether an object has been perceived. To this end, we apply state-of-the-art methods for response aggregation on data obtained from simulated driving sessions with 30 subjects. Our results show that a probabilistic aggregation scheme on top of an Expectation-Maximization-based estimation of source reliabilities can predict hazard perception at a recall and precision of 96% in real-time.

1. Introduction

Eye movements and physiological signals such as heart rate and galvanic skin conductance are measured in a variety of use-cases to derive information about a subject. More specifically, since the latter two signals are considered as strong indicators of cognitive load and stress, they have been analyzed in several applications to understand user behavior in complex tasks, and in particular during driving. In fact, sudden stress in safety-critical situations, as they may occur during driving, arouses the sympathetic nervous system. The subject transpires and skin conductance and heart rate change as a result (Backs, Lenneman, Wetzel, & Green, 2003; Helander, 1978; Lenneman & Backs, 2009; Lewis & Phillips, 2012; Mehler, Reimer, & Coughlin, 2012; Reimer, Mehler, Coughlin, Godfrey, & Tan, 2009; Son et al., 2011; Taylor, 1964). With advances in eye-tracking technology and analytical approaches, additional means have become available to assess performance and stress level during driving. More specifically, since changes in pupil diameter have been considered as a measure of emotional arousal and autonomic activation, pupil analysis has been employed in several studies (Benedetto et al., 2011; Bradley, Miccoli, Escrig, & Lang, 2008; Potamitis et al., 2000). The general assumption during driving is that visual perception of scene features such as signs, pedestrians, and obstacles requires foveated vision, i.e., an object of interest in the scene is considered as perceived if it has been fixated by the driver (Fletcher & Zelinsky, 2009). Although peripheral vision is considered as sufficient for some subtasks, such as keeping the vehicle centered in the lane (Summala, Nieminen, & Punto, 1996), it has been reported that peripheral vision is insufficient for the detection of traffic hazards (Maltz & Shinar, 2004).

Recent studies investigating the correlation between hazard fixation and its perception have however reported that the direction of a driver's gaze towards an upcoming hazard does not a priori imply its perception (Kasneci, Kasneci, Kübler, & Rosenstiel, 2015, pp. 411–434; Tafaj, Kübler, Kasneci, Rosenstiel, & Bogdan, 2013). Moreover, several studies have reported that subjects have shown adequate hazard detection although the target object has not been fixated (Kasneci et al., 2014; Kübler et al, 2015a, b). Thus, in some cases, peripheral vision may be sufficient for hazard perception

* Corresponding author.
E-mail address: enkelejda.kasneci@uni-tuebingen.de (E. Kasneci).

http://dx.doi.org/10.1016/j.chb.2016.11.067
0747-5632/

(Tafaj et al., 2013).

In driving scenarios, determining whether a hazard was perceived by the driver early on can lead to significant support of automatic accident avoidance systems. To approach this challenge, signals from different sensors need to be processed in an online fashion and aggregated according to their reliability. The reliability of a sensor, however, depends not only on the type of the sensor but also on the subject. For example, eye-tracking signal is sensitive to make-up and changing illumination, while skin conductance and heart rate are vulnerable to loosened or detached electrodes. Deriving a binary decision about hazard perception from raw sensor data is a challenging problem of its own and requires device specific filtering, synchronization and processing. We will threat this necessary preprocessing step as abstract throughout the manuscript and work only on the readily preprocessed binary decision label in order to demonstrate the proposed concepts in a more general, device independent way. For details on how the data used throughout the manuscript was preprocessed, please refer to (Kübler et al., 2014).

In this paper we address the problem of how to combine multiple physiological (i.e., heart rate and galvanic skin response) signals with eye-tracking measurements for an automated detection of target perception. Especially in the absence of ground truth, a viable aggregation method has to jointly infer the reliability of the source delivering the signal and the true event taking place (e.g., whether the target was perceived) (Fig. 1).

From a theoretical viewpoint the problem can be formalized as follows. Given binary signals or responses $X_{i1}, \ldots, X_{in} \in \{0,1\}$ from n sources (e.g., sensors), 1 meaning that the event $e_i, 0 \le i \le t$, has taken place and 0 that e_i has not occurred, how can we aggregate the responses in a way that we can learn the latent truth (i.e., whether the event occurred or not).

In the data mining literature, there is a vast body of work addressing the aggregation of responses in order to find the latent ground truth. Some of these approaches can be adapted to the aggregation of sensor signals. However, only a few are applicable in real-time. In this paper we analyze the most popular algorithms in

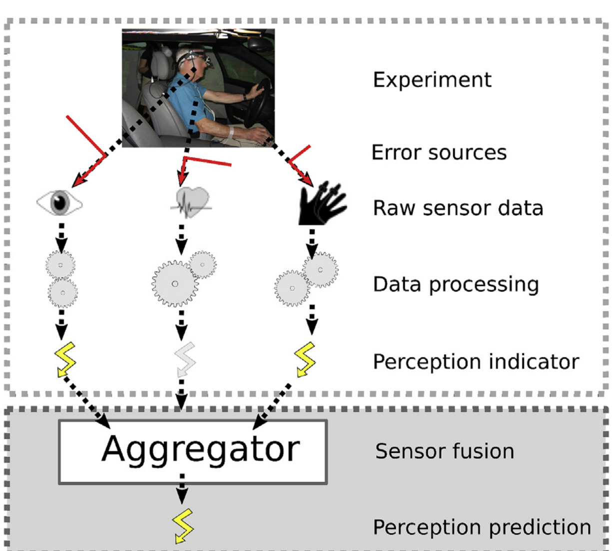

Fig. 1. This manuscript discusses the aggregation of physiological sensor data (eye-tracking, heart rate, skin conductance) that has already been preprocessed to a binary perception indicator (displayed as a thunderbolt). Aggregation and reliability estimation of the individual perception indicators is performed (bottom box). For information on recording and preprocessing of the data (top box) see, for example (Kübler et al., 2014).

this realm and provide a practical guidance for their real-time application in driving scenarios.

In the following we will use the terms signal, response, answer and claimed value interchangeably.

The rest of the paper is organized as follows: Section 2 gives an overview of related work in the area of latent truth discovery and reliable response aggregation. Section 3 provides a practical framework for the real-time application of popular truth discovery and aggregation algorithms. An extensive analysis and evaluation of the algorithms on real-world data collected from driving experiments with human subjects is presented in Section 5. The data was carefully labeled by experienced annotators as described in Section 4.

2. Related work

From an abstract viewpoint, there are 3 categories of latent truth discovery methods:

Bayesian Inference algorithms use prior distributions for the truth and reliability parameters and jointly estimate truth and source reliability by fitting the parameters to the available data based on the assumed prior distributions.

Fix-point and Expectation Maximization algorithms start with an initial guess on the truth and reliability parameters and simplifying assumptions are used to iteratively fit the parameters to the available data.

Semi-Supervised Learning algorithms start with a set of known ground truth labels. This initial ground truth and other assumptions are exploited to learn the reliability of sources. In turn, the reliability estimations can be used to estimate the latent truth.

In the following paragraphs, we give an overview of the above three groups by highlighting representative approaches.

2.1. Bayesian Inference

TruthFinder (Yin, Han, & Philip, 2008) models the influence between claimed values and applies Bayesian analysis to iteratively estimate source reliabilities and the latent truth. AccuSim (Dong, Berti-Equille, & Srivastava, 2009; Li, Dong, Lyons, Meng, & Srivastava, 2012) integrates the similarity between claimed values into the Bayesian inference approach and proposes an extension of the algorithm AccuCopy in which also source similarities – in terms of which source might have copied from which other source – are considered. The more a source has copied from other sources, the more its weight is reduced.

A Bayesian approach to knowledge corroboration is proposed by Kasneci, Van Gael, Herbrich, & Graepel (2010); Kasneci, Van Gael, Stern, Graepel (2011), where a latent truth discovery model integrates the logical dependencies between facts in a knowledge base and crowd opinions to derive the underlying correctness of the facts in the knowledge base.

Latent Truth Model (LTM) (Zhao, Rubinstein, Gemmell, & Han, 2012) is a probabilistic graphical model that applies collapsed Gibbs sampling to estimate the false positive and the false negative rate of sources by optimizing for the most probable answers.

Another Bayesian inference approach for continuous responses is presented in Zhao & Han, 2012.

2.2. Fix-point algorithms and expectation maximization

In 2-Estimates (Galland, Abiteboul, Marian, & Senellart, 2010) the assumption that there is one and only one true value for each