J. Vis. Commun. Image R. 35 (2016) 67–77

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

Recognition of the gaze direction: Anchoring with the eyebrows ☆

Laura Florea *, Corneliu Florea, Constantin Vertan

The Image Processing and Analysis Laboratory (LAPI), University "Politehnica" of Bucharest, 313 Splaiul Independeței, Bucharest, Romania

ARTICLE INFO

Article history:
Received 2 September 2015
Accepted 5 December 2015
Available online 12 December 2015

Keywords:
Eye and eyebrow landmarks
Gaze recognition
Vision processing
Eyebrow influence on gaze
Shape processing
Face analysis
Real time
Multistage fusion landmark localization

ABSTRACT

In this paper we investigate the accuracy of estimating a person's direction of gaze from remote imaging. The problem is addressed by a person independent, multistage fusion approach for eye landmark localization, followed by eye region analysis for actual gaze recognition. We test the proposed landmark localization system on three databases, showing superior accuracy than state of the art solutions. Finally, we show that, inspired by human perception, by incorporating the location of eyebrows, superior performance is achievable when estimating the gaze direction. Given the found results, we argue that computer vision systems for gaze recognition should mimic the human perception and incorporate the eyebrows.

1. Introduction

Non-verbal communication is a fundamental part of the inter-human interaction and there have been considerable efforts for creating automatic systems able to mimic such trait [1]. Psychologists noted the importance of efficient perception of people's gaze direction as part of non-verbal messaging [2]. Kobayashi and Kohshima [3] noted that humans are the only primates which exhibit dark iris against a white sclera through the palpebral fissure (the aperture between the eyelids), thus providing a number of potential cues to gaze direction. They also suggested that the combination of a dark iris with light sclera evolved to provide a cue for gaze direction (an advantage in a highly social and cooperative species such as humans), in contrast to the darker sclera of other primates that may have evolved for the opposite purpose, to make gaze direction obscure, such that predators cannot tell that they have been spotted. Ando [4] showed that the brightness altering of the scleral areas has deep effects on perceived eye gaze direction.

The potential of the gaze as a social signal has been in sight of the academic research for a long period. Humans are processing the gaze direction of the interacting person by specific neuronal circuitry involving areas from the posterior superior temporal sulcus region, that has been found [5] to respond to the inferred intentionality of social cues. The review of Frischen et al. [2] evaluates the social impact of gaze direction and concludes that many opportunities arise upon the understanding of the perceived direction of gaze. They asserted that "people's eyes convey a wealth of information about their direction of attention and their emotional and mental states". They further note that "eyes and their highly expressive surrounding region can communicate complex mental states such as emotions, beliefs, and desires" and "observing another person's behavior allows the observer to decode and make inferences about a whole range of mental states such as intentions, beliefs, and emotions". All these findings motivate the necessity of building systems that infer one's direction of gaze.

In this paper we propose a new system for identification of gaze direction in a framework that implies remote[1] imaging under passive illumination (i.e. standard video camera). This topic has been previously addressed and it has been concluded [6] that better separation is possible on the horizontal direction than on the vertical direction. To address this imbalance, we introduce the position of the eyebrow as anchors to improve the assessment on the vertical direction. The applications are multiple, varying from human computer interaction [7], to psychological investigation such as understanding the process of reading [8] which can be helped by automated solution.

☆ This paper has been recommended for acceptance by M.T. Sun.

* Corresponding author.
E-mail addresses: laura.florea@upb.ro (L. Florea), corneliu.florea@upb.ro (C. Florea), constantin.vertan@upb.ro (C. Vertan).

[1] We use "remote" as opposed to head-mounted device and not in the sense of "being very far".

http://dx.doi.org/10.1016/j.jvcir.2015.12.003
1047-3203/

1.1. Humans and gaze direction

Although many existing studies discuss various aspects of human perception or control of the gaze direction, we will refer here to only two of them.

Firstly, we recall that gaze movement when performing non-visual tasks is an indicator of the person behavioral intent [9]. On the other hand, Laeng and Teodorescu [10] showed that even when performing non-visual tasks, voluntary control of the eye movements is possible. Thus, to avoid additional incertitude related to the topics in debate, we consider that the recording digital system used for automatic gaze direction estimation should be as non-invasive as possible; wearable devices being thus excluded. Another consequence is that correlation with head pose (trait which is discussed in the review by Murphy-Chutorian and Trivedi [11] and exploited by Valenti et al. [12]) is subject to voluntary control and thus it will not be taken into account in the current work.

Secondly, following the conclusion reached by Hansen and Ji [6] (supported also by the findings in [13] or [14]), there is need for improved precision in the estimation of the vertical direction of gaze. Humans have less trouble in identifying the direction of gaze; the results of Sadr et al. [15] showed that, for face analysis, the eyebrows are as important as the eyes. Dakin and Watt [16] showed that one "barcode" processed by us when scanning others faces is the ensemble of eye together with the eyebrow. At last, Watt et al. [17] showed that the perception of eyebrow regulates the perception of the gaze direction: the critical distance (i.e. the distance at which we correctly perceive others' direction of gaze) is reduced by the lowering of the eyebrows; in other words, we use the eyebrows as relative anchors to relate the position of the iris inside the sclera on the vertical axis, while estimating the direction of gaze.

Concluding, we consider that these findings motivate the construction of a gaze direction estimator based on passive illuminated and remotely recorded images, that should locate both the eyes and the eyebrows landmarks before inferring the actual gaze direction.

1.2. Related work

From a computer vision point of view, the estimation of the direction of gaze is typically approached using an eye tracker [18,6]. Various solutions may be separated by (a) the position of the tracker, which may be head mounted or remote and (b) the illumination source domain, which may be active (using infra-red light emitting source) or passive (within the visible spectrum). Based on the previous described motivation, we avoid distinct, obvious devices (i.e. wearable), while for illumination source, in order to broaden the applicability, we rely on passive illumination sources.

Gaze estimation in remotely acquired passive illuminated scenes was discussed in multiple works. Wang et al. [19] select recursive discriminant features from a topographic image feature pool to train an Adaboost that locates the eye direction. Cadavid et al. [20] train a Support Vector Machine (SVM) with spectrally projected eye region patches to identify the direction of gaze. Wolf et al. [21] use the eye landmark localizer provided by Everingham and Zisserman [22] to initialize the fit of the eye double parabola model and localize the direction of gaze. Unfortunately none of the mentioned solutions reports results on public databases or make the code publicly available. All these methods start by localizing eye landmarks and follow by analysis of the located eye regions for the gaze direction determination. We use the same model and one of our main contribution is a system for localizing landmarks in the eye area; beside the different technical approach, we further differ by the fact that we use eyebrow landmarks too.

Furthermore, the proposed approach, as it is trained and tested on different persons, does not require any person dependent calibration. Also no-calibration gaze recognition is proposed in the works of Sun et al. [23,24], but in contrast to us, they rely on consumer depth images (e.g. Kinnect acquired), thus limiting the applicability to such devices.

1.2.1. Face landmarking

Facial landmarking originates in the classical holistic approaches of Active Shape Models (ASMs) [25], Active Appearance Models (AAMs) [26] and Elastic Graph Matching [27].

Building on the ASMs/AAMs versatility, a multitude of extensions did appear. For instance in [28], higher independence between the facial components is encouraged while the actual fitting step is further optimized by a Viterbi optimization process. Another example is by adding a monotonically decreasing inert factor function to the traditional iterations to form the so-called D-ASM [29,30].

In the later period, the ASM underlying holistic fitting switched to independent models built on top of local part detectors to form the so-called Constrained Local Models (CLMs) [31]. The CLM model was extended with full voting from a random forest in [32], by probabilistic interpretation for optimization of the shape parameters in [33].

All other notable works assume a connected spatial model, requiring the need for approximate matching algorithms. This is a rich class of methods, including some of the most recent and accurate solutions. Thus Valstar et al. [34] complemented the SVM regressed feature point location with conditional MRF to keep the estimates globally consistent. Zhu and Ramanan [35] relied on a connected set of local templates described with HOG. Yu et al. [36] used 3D deformable shape models to iteratively fit over 2D data and identify without respect to pose the facial landmarks positions. Martinez et al. [37] trained SVM regressors with selected Local Binary Patterns to formulate initial predictions that are further iteratively aggregated for improving accuracy. In [38], the relation between fiducial points is encoded directly in the localization system, which is based on deep convolutional networks.

Despite their evolution, many facial landmarking methods do not localize the iris center. As this is critical for our application, we refer here to the maximum isophote method proposed by Valenti and Gevers [39], to the gradient accumulation in circular shapes proposed by Timm and Barth [40] and to the random forest trained with image patches introduced by Markus et al. [41].

The aggregation of eye landmarks for the recognition of gaze direction in the 7 cases related to specific hypothesis of the Neuro-Linguisitc Programming was previously proposed by us in [14], extending the method from [42]; this method segments the eye region (bounding box) previously identified by landmarks in order to extract its components; the analysis of the components leads to gaze direction estimation. Radlak et al. [43] employed Hybrid Image Projections functions as defined by Zhou and Geng [44] followed by either a SVM or by a random forest (RF) to recognize the direction of gaze.

The here-proposed method is build upon the framework introduced by Florea et al. in [13]. Yet, it differs by using independent left/right eye models, extending the models to the eyebrow, a new solution to aggregate the landmarks information (which, this time, include the eyebrow one) in the gaze direction and more extensive and intensive testing; furthermore the gaze recognition system is completely new, as we use the double parabola model, eyebrows and Support Vector Machine for classification.

1.3. Paper structure and contribution

We propose a system (schematically presented in Fig. 1) for the automatic recognition of the direction of gaze. First the landmarks of the eye and eyebrow regions are located using a multi-stage fusion approach described in Section 2. The landmarks positioning