Pattern Recognition Letters 57 (2015) 74–80

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

A robust perception based method for iris tracking ☆

Vittoria Bruni [a,*], Domenico Vitulano [b]

[a] Department of Scienze di Base e Applicate per l'Ingegneria, University of Rome La Sapienza, Via A. Scarpa 16, Rome 00161, Italy
[b] Istituto per le Applicazioni del Calcolo M. Picone—C.N.R., Via dei Taurini 19, Rome 00185, Italy

ARTICLE INFO

Article history:
Received 8 May 2014
Available online 16 September 2014

Keywords:
Iris tracking
human visual system
Jensen–Shannon divergence

ABSTRACT

The paper presents an application of the modified kernel based object tracking to iris tracking. Human perception rules are used for defining a proper feature space for iris that mainly accounts for the fact that eyes represent instinctive visual fixation regions. In addition, a similarity metric that is close to the way human vision compares and perceives the difference between scene components has been employed for finding the exact iris location in subsequent frames. As a result, just one iteration of the mean shift algorithm allows us to get a faithful estimation of iris location in subsequent frames. This property makes the proposed algorithm implementable on mobile devices and useful for real time applications. Experimental results performed on MICHE database show that the proposed method is robust to changes in illumination or scale, iris partial or total occlusion for some subsequent frames, blinks.

1. Introduction

Iris detection and tracking represent two important tasks not only in biometric applications. Several recent studies and applications proved that iris movement is very informative and can be used in different ways in several fields. It is not only a fingerprint for human being, but it also plays a significant role in psychology, robotics, security, and especially in neurological studies. Eye movement is actually a widely investigated topic in neurology and in vision research since it represents the basis of perceptual learning and gives information about the way the brain codes visual information, reacts to visual stimuli, learns from what it sees, selects what is important in a visual scenario [31,1,12,37,28,32]. On the other hand, neurological studies gave new flour to image processing applications since they offered new instruments for processing and understanding visual information. In fact, independently of its origin and the way it has been transmitted, image is received and processed by human brain which judges its quality and learns its content. As a result, it makes sense to process image information in a perceptual guided way. Saliency maps [25], that represent the image content in a hierarchical way from the most visible to the less visible region, are a representative example of this new way of processing image. Following this concept, it has been shown that image anomalies can be detected as those image components that are visible at the first glance [7]. In fact, they immediately capture human attention since they are perceived as foreign objects in the scene. In the same way, image content can be coded in a non uniform way according to its visual importance [40,22,39]. More in general, the use of the mechanisms that regulate human vision allows to optimize a wide class of image processing based systems and applications, like compression, restoration, printing, watermarking, segmentation, displaying, and also object tracking [40,22,39,5,6,33,8,17]. In fact, in a recent paper the perceptual interpretation (version) of a well known tracker allowed to achieve satisfying and quite impressive results [5,6]. Inspired by this work, in this paper we are interested in studying to what extent the use of some features that are related to the way iris is perceived by human eye can help in its tracking, making tracking robust and at the same time quite fast. Actual iris tracking algorithms are based on optimization procedures like mean-shift, Kalman filtering, particle filtering or their combination in order to compensate the limits of each technique [24,2,43,20,42,35]. The major effort consists of making those trackers robust to changes in illumination, eye closure, face orientation, distance from the camera, etc. It often means to define specific and distinctive features for the iris as its geometrical appearance (circle or ellipse), the presence of a bright pupil effect, the high contrast with respect to the white part of the eye, the color appearance and so on. That is why some models look at the iris contours, characterize their profile and track them in subsequent frames. However, those methods, even though performing, can be computationally expensive since more distinctive target features, especially the ones related to the geometrical appearance, can require additional and not negligible computational load. In addition they might be not robust to occlusions or changes in illumination or scale. One of the most interesting neurological results is that human eyes are attracted by very few points (regions) of the observed scene [18,27,30]. In addition, human eye reacts to a moving

☆ This paper has been recommended for acceptance by Nappi Michele.
* Corresponding author: Tel.: +39 06 49766648; fax: +39 06 4957647.
E-mail address: vittoria.bruni@sbai.uniroma1.it, bruni@iac.rm.cnr.it (V. Bruni).

http://dx.doi.org/10.1016/j.patrec.2014.09.001
0167-8655/

stimulus which can be modeled as a combination of luminance modulated (first order motion) and contrast modulated (second order motion) sinusoidal stimulus [36,23,19,29]. With regard to face processing, eyes belong to these attentional points [21] and iris represents its prominent and highly contrasted component; moreover iris is a moving and textured eye component. These concepts have been confirmed also by recent studies in biometric [12] in which eye movements (saccades) and fixations are deemed special features of individuals to be used for human identification. Hence, the main aim and the novelty of the presented work is the use of perceptual concepts to track the organ that is responsible for visual perception. To this aim, mean shift algorithm is still used in the optimization process, but some neurological results are exploited for defining a proper feature space which takes into account the sensitiveness to a moving and textured stimuli as well as the measures that better characterize fixation points, namely luminance and contrast. Finally, the Jensen–Shannon divergence (JSD) is used as similarity metric due to its correlation with the visual system when comparing different objects [4,8,3]. JSD is able to recognize iris in subsequent frames even if iris appearance changes due to partial eye closure, iris movement, change in illumination, distance from the camera. In fact, JSD is less sensitive to changes in iris geometrical features thanks to its dependence on a more global measure, like the distribution of the luminance and contrast in iris region.

As a result, the focus of the paper, and its main difference with the existing literature about the topic, is not the iris but the way it is perceived by the observer. Human eye is sensitive to iris since it belongs to a region of the face that is fixated by human eye at the first glance and that is instinctively used by human eye in the recognition process. The additional perceptual property of the iris is that it is a moving and highly contrasted component of the scene and then it attracts human attention more than other parts of the eye. As a result the use of specific characteristics of fixation points in terms of luminance and temporal contrast as well as the use of a metric that weights image information in a way that is consistent with the one employed in the vision process, allow us to define a very simple and almost inexpensive tracker that, on the one hand is able to compete with existing trackers in terms of tracking precision, without exploiting specific objective characteristics of the iris, and, on the other hand, to outperform them in terms of required computational load, making it suitable for real time applications and easily implementable on mobile devices.

Experimental results show very promising tracking results. The combination of the contrast/luminance based feature space and the similarity metric allows mean shift to converge after one iteration with a considerable reduction of the computing time, without introducing annoying drifting effects. In addition, the tracker is robust to blink, changes in illumination, zoom, scale and contrast and it does not require Kalman filtering or more specific morphological and/or geometrical iris features to get satisfying performance.

The remainder of the paper is the following. Next section first gives in depth motivations for the feature space and the similarity metric, then describes the tracking algorithm. Section 3 presents some experimental results and gives implementation details. Finally, Section 4 draws the conclusions and gives guidelines for future research.

2. Perception based feature space and similarity metric

In this section visual perception is used for defining a proper feature space for iris and for selecting a proper similarity metric to use in iris tracking process. It is worth observing that the visual features are defined just for tracking and not for iris detection. Iris detection is a delicate topic and it is out of the scope of this paper. In the paper we assume that iris has been detected in the first frame of the video sequence using an existing iris detector [9,10] or it has been manually indicated by the external user.

2.1. Luminance and contrast based iris feature

Several neurological studies proved that during the observation process few points attract human attention and those few points are necessary for understanding the scene content in the early vision [18,27,30]. Those fixation points depend on both the image subject and the luminance and contrast characteristics of the scene content [18,27,11]. With regard to face processing, it has been shown that human eye has a sort of preference for the regions containing eyes and nose, with particular and immediate importance for the left side [21]. It is due to the fact that the right part of the brain is the one sensitive to low frequencies and actually the early vision corresponds to a more global vision of the scene (a sort of adaptive low pass filtering). Although the complexity of human vision, luminance and contrast remain the two main measurements that are involved in the observation process, especially in the early vision, and it is true for both static and moving scenes. The two interesting aspects of this kind of vision process are that luminance and contrast seem to have a sort of independence in correspondence to fixation points [18,27]. Specifically, light adaptation (luminance gain) and contrast gain are the two rapid mechanisms that control the gain of neural responses in the early vision and they operate almost independently. It means that luminance and contrast can be considered as two independent sources and the visual stimulus is the linear combination of these two sources. More precisely, for each point (x, y) of the image I, the visual stimulus can be modeled as

$$S(x, y) = \frac{1}{2}L(x, y) + \frac{1}{2}C(x, y) \quad (1)$$

where $L(x, y)$ and $C(x, y)$ represent the luminance and the visual contrast measures at the image point (x, y). The visual contrast is, in general, measured as a sort of a normalized spatial variation of the luminance [41]. More common definitions are the ones based on the Weber's law, i.e. $C_W(x, y) = (\sigma_L(x, y))/(\mu_L(x, y))$ where $\sigma_L(x, y)$ and $\mu_L(x, y)$ respectively are the image local standard deviation and the mean of the luminance, which measures the visibility of an object with a uniform luminance with respect to a uniform background, or the one based on the definition of the Michelson contrast, i.e. $C_M(x, y) = (\max L(x, y) - \min L(x, y))/(\max L(x, y) + \min L(x, y))$, which measures the visibility of a sinusoidal stimuli. In this work, the Michelson visual contrast has been considered due to the textured nature of the iris and the fact that there is a close relationship between the distributions of Michelson contrasts in natural images and the contrast response functions of neurons [34].

However, the persistence of vision is another characteristic of human vision. It derives from the fact that neurons in the visual system summate information over both space and time. It turns out that the receptive fields have spatio-temporal characteristics and not only spatial [36]. It means that the evolution of luminance and contrast over time in a fixed location cannot be neglected. In particular, it is possible to define a first order motion stimulus, that is related to the spatio-temporal variation of the luminance, and a second order motion stimulus which is related to the motion of a contrast modulated texture [36,23,19,29]. Also in this case the two stimuli are almost independent and the global visual stimulus is the sum of the two different spatio-temporal stimuli. In general, the second order motion is detected and modeled as the ratio between the temporal variation of the luminance and its spatial variation. Based on these observations, the feature space, that can be seen as the image of the perceived stimuli, can be roughly modeled as the weighted sum of the luminance and contrast components whenever the contrast assumes a temporal significance. In other words, the stimulus at time t can be written as

$$S(x, y, t) = \frac{1}{2}L(x, y, t) + \frac{1}{2}C(x, y, t) \quad (2)$$

with $C(x, y, t) = (L(x, y, t) - L(x, y, t-1))/(L(x, y, t) + L(x, y, t-1))$. In addition, since the perception of scene content at the first glance and