Engineering Applications of Artificial Intelligence 31 (2014) 90–100

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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

An efficient eye detection and tracking system based on particle swarm optimization and adaptive block-matching search algorithm

Rehab F. Abdel-Kader*, Randa Atta, Sheren El-Shakhabe

Electrical Engineering Department, Faculty of Engineering, Port Said University, Port Fouad 42523, Port Said, Egypt

ARTICLE INFO

Available online 23 July 2013

Keywords:
Eye detection and tracking
Deformable template matching
Particle swarm optimization

ABSTRACT

The problem of eye detection and tracking in video sequences is very important for a large number of applications ranging from face recognition to gaze tracking. Eye detection and tracking are challenging due to a variety of factors such as eye-blinking, partially closed eyes, and oblique face orientations which tend to significantly limit the efficiency of most eye trackers. In this paper, an efficient eye detection and tracking system is presented to overcome these limitations. The proposed system switches between the particle swarm optimization (PSO) based deformable multiple template matching algorithm and the adaptive block-matching search algorithm to improve the efficiency and robustness of the tracking system. For eye detection, PSO-based deformable multiple template matching is employed to estimate the best candidate of the center of the eyes within an image of the video sequence with the highest accuracy. For eye tracking the block-matching algorithm with adaptive search area is utilized to reduce the computational time required to perform the PSO-based algorithm. Experimental results on the standard VidTIMIT database show that the proposed method outperforms the deformable template matching based methods such as genetic and PSO. Moreover, it achieves better performance compared to model-based methods such as the statistical active appearance model (AAM) method and the edge projections based method in terms of accuracy and computational complexity.

1. Introduction

In recent decades, human eye detection and tracking has emerged as a challenging task that spans multiple disciplines such as image analysis and computer vision. This is due to the wide range of applications in vision based man–machine interaction technology. Robust, dynamic, and real-time eye-tracking provide new means for man–machine interaction that can greatly improve the underlying user experience. The main idea is to make the computer perceive various natural user communicative cues such as gaze directions to activate or command devices. New applications for eye-tracking communication interfaces are being considered in complex technological environments, such as hospital operating rooms, airplane cockpits, and industrial control units (Kim et al., 2000; Park and Lim, 2001). Consumer applications have been developed for assisting people with disabilities (Betke et al., 2002; Chau and Betke, 2005), monitoring driver vigilance (Ishikawa et al., 2004; Ji and Yang, 2001; Wang et al., 2010), gaze tracking for gaming design (Corcoran et al., 2012), and for identifying Website Keyobjects (Velásquez, 2013).

Eye detection and tracking techniques can be classified into two main categories: intrusive methods (head-mounted) which require direct contact with the eyes and non-intrusive (remote) methods which avoid any physical contact with the user. Typically non-intrusive methods are more attractive since the camera and the light sources are placed at a distance from the user, usually below or around the computer screen, making the user less constrained to uncomfortable equipment. In non-intrusive methods, eye detection is first performed. Various methods of eye detection have been proposed to detect and localize eyes in still images (Kumar et al., 2002; Motwani et al., 2004; Tivive and Bouzerdoum, 2005; Turkan et al., 2008; Wu and Zhou, 2003; Zhou and Geng, 2004; Zhu et al., 2002; Wang and Chen, 2006). Turkan et al. (2008) proposed a method for eye localization in face images with frontal pose and upright orientation. In this method, a given face region is filtered by a highpass wavelet-transform filter. Candidate points for each eye are detected after analyzing horizontal projections and profiles of edge regions in the highpass filtered image. All the candidate points are then classified using a support vector machine (SVM). Locations of each eye are then estimated according to the most probable ones among the candidate points. Once the locations of the eyes are detected, tracking them in successive images is performed. Broadly speaking, non-intrusive methods can be classified into active infrared (IR) illumination based approaches and traditional image-based

* Corresponding author. Tel.: +20 1001172820.
E-mail addresses: rehabf98@gmail.com, rehabfarouk@eng.psu.edu.eg (R.F. Abdel-Kader), r.atta@eng.psu.edu.eg (R. Atta), sherenzakaria@eng.psu.edu.eg (S. El-Shakhabe).

0952-1976/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.engappai.2013.06.017

eye tracking passive approaches. The former approaches are simple and effective. They exploit the spectral properties of the pupil under near IR illumination. Many techniques (Ebisawa, 1998; Ji and Yang, 2001; Morimoto and Flickner, 2000) have been developed based on this principle. They are based on an active IR light source to produce the dark or bright pupil effects. Most of these methods require distinctive bright/dark pupil effect to work well.

On the other hand, traditional image-based eye tracking approaches detect and track eyes based on exploiting the differences between frames and assume that the eyes have distinctive properties in its shape and intensity dissimilar from other facial features. According to Zhu and Ji (2005), traditional image-based approaches are classified into three categories: Appearance based methods (Bacivarov et al., 2008; Kawato and Tetsutani, 2004; Zhu and Ji, 2005), feature based methods (Lucas and Kanade, 1981; Sirohey et al., 2002; Tian et al., 2000), and template based methods (Akashi et al., 2007; Al-Mamun et al., 2009; Tan et al., 2003). In appearance-based methods, the eyes are detected based on their photometric appearance. These methods require large sets of training data for training the classifier (e.g. neural network or the support vector machine). Training data should represent the eyes of the different subjects under various face orientations and illumination conditions. Appearance-based methods can be combined with active IR illumination methods as in (Zhu and Ji, 2005) to solve common eye tracking issues such as the presence of glasses and varying lighting conditions. Methods based on active appearance models were proposed for eye tracking. Bacivarov et al. (2008) proposed a statistical active appearance model (AAM) designed to recover the eye blinking parameters of an eye model as well as tracking the location of the eyes in image sequences for various subjects with various expressions or poses. Kawato and Tetsutani (2004) presented a new algorithm where blinking can be detected based on the differences between two consecutive video frames. Head movements are also taken into consideration. For eye tracking, a template of “Between-the-Eyes” is used instead of the eye template.

In the feature-based methods, the eyes are tracked based on exploring their characteristics to identify a set of distinctive features around the eyes such as edges and color intensity variations in the iris and eye corners. Tian et al. (2000) proposed a method of tracking the eye location, detecting the eye state, and estimating the eye parameters for each frame in a video sequence. Manual initialization of the eye template in the first frame is required in this method. The iris is detected and tracked by intensity and edge information and then the eye state is obtained by the iris detection. If the iris is detected, the eye is open. Then, for open eyes eyelid center points are tracked using the tracking method developed by Lucas and Kanade (1981). According to Sirohey et al. (2002), methods of tracking the motion of the head and the independent motions of the irises and upper eyelids in an image sequence are described. These methods require no manual initialization and do not need any prior assumptions about the relative positions of the irises and eyelid.

In template-based methods, a generic eye model is initially designed and template matching algorithm is utilized to search the image for the reference eye template. Considering the possible changes in the face orientation, conventional template matching technique cannot be applied even when the templates are updated frame-by-frame with patterns of current located positions. This is because the tracking points will gradually migrate out from the eyes. Therefore the tracked points do not always represent the centers of the eye pattern. Deformable template-matching (Tan et al., 2003) is a variant of the traditional template matching algorithm in which the reference eye template is allowed to translate, rotate and deform to fit the best representation of the eye shape in the image. Although this method can yield high tracking accuracy, it is computationally expensive, and requires a good image contrast for the method to converge correctly. In Al-Mamun et al. (2009) an eye detection algorithm based on genetic deformable template matching was presented to avoid exhaustive blind search associated with traditional template matching techniques. Akashi et al. (2007) proposed a population inheritance genetic algorithm (GA) for eye detection and tracking to overcome the high computational cost associated with applying GA with the deformable template matching in each frame in a video sequence. These methods do not require any prior knowledge about the eye geometry or potential eye locations on the facial image as in Turkan et al. (2008) and Bacivarov et al. (2008) since they use artificial eye templates.

Swarm intelligence based algorithms have gained significant attention in the last decade due to their robustness, flexibility, and the increasing demand on their applications. Particle swarm optimization (PSO) (Eberhart and Kennedy, 1995; Eberhart and Shi, 2001, 1998) is one of the popular swarm intelligence based methods that has been applied successfully as an optimizer in many domains such as training artificial neural networks, linear constrained function optimization, wireless network optimization, data clustering, and many other areas where GA can be applied. In Perez et al. (2010) a PSO-based method was proposed to generate templates for frontal faces and iris localization in real time. The method was applied to frontal view faces to generate a set of templates for different face sizes. Eye positions were used to align faces in order to make the PSO algorithm converge to a solution. One alternative for this method is to use faces from different subjects to generate a generic PSO template. Another alternative is to create personalized templates by using a set of faces from the same subject to create the PSO templates. The method also allows the control of the maximum number of points in the template thus reducing the computational time, allowing real-time processing. The PSO templates were tested and compared to other methods based on anthropometric templates and Adaboost for face localization. Furthermore, it was compared to a method using combined binary edge and intensity, and to a method based on SVM classifiers for iris localization. Results showed that the PSO templates exhibit better spatial selectivity for frontal faces resulting in a better performance in face and iris localization.

In summary, the template-based methods based on either GA or PSO used for eye detection (Al-Mamun et al., 2009; Perez et al., 2010) and eye tracking (Akashi et al., 2007) have some limitations. The efficiency of these methods is limited especially under various undesirable conditions such as eye-blinking, partially closed eyes, and oblique face orientations. Furthermore, they are inappropriate for real-time implementation since they are computationally expensive. To overcome the shortcomings of these methods, in this paper an efficient eye detection and tracking system is proposed. The proposed system combines PSO-based inheritance deformable multiple template matching and adaptive block matching search algorithms. The former algorithm is used to find the center of an eye position within an image in a video sequence. Since this step is the most computationally expensive, it is dynamically performed when the eye position could not be detected by the latter algorithm. In the proposed system, multiple templates are used and the system can select among them to increase the detection accuracy. For eye tracking, adaptive block-based matching search algorithm is presented to reduce the computational complexity required to perform the PSO-based algorithm. In this algorithm, the appropriate size of search area is predicated from the previously tracked group of frames (GOF). Therefore, the predicted value may vary from a GOF to another during the eye tracking process and it can then significantly affect the performance of the algorithm.