Biologically Inspired Cognitive Architectures 18 (2016) 68–79

Contents lists available at ScienceDirect

Biologically Inspired Cognitive Architectures

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

Research article

A novel biologically inspired computational framework for visual tracking task

Alireza Sokhandan *, Amirhassan Monadjemi

Department of Artificial Intelligence, Faculty of Computer Eng., University of Isfahan, Isfahan, Iran

ARTICLE INFO

Article history:
Received 6 July 2016
Revised 28 September 2016
Accepted 28 September 2016

Keywords:
Visual tracking
Visual cortex
Saliency
Motion perception
Occlusion

ABSTRACT

Visual tracking is a process in which the location of one or several objects is estimated in a video sequence based on their appearance. High diversity of states and conditions of a moving stimulus and the existence of abundant challenges such as a cluttered background, variations in the objects appearance, and occlusion make this problem very complicated. Although a wide range of diverse algorithms have been introduced in this field, even the state-of-art ones have not been able to achieve high accuracy in facing different challenges of this problem. Nevertheless, by observing the biological world and paying attention to the living beings, particularly humans, it is revealed that the biological visual system performs the visual tracking task ideally. Therefore, inspiring from this biological system and especially its spatiotemporal processing, motion perception, attention, and saliency mechanism, in this paper a new biologically inspired visual tracking framework is proposed. The suggested framework includes a six-phase process in which the executive blocks, their connections, and information flow among them are inspired by the visual cortex. Using visual features extracted in primary layers of the visual cortex as well as the local-global and sparse representation of the object of interest make possible to employ the proposed framework in different conditions. In this paper, the theory of the suggested framework and its biological background is explained. Investigating the procedure of proposed framework from the viewpoint of employed features, spatiotemporal process, and the quiddity of local-global processing indicated that this framework enjoys the ability to manage most of the mentioned challenges, and can robustly track the object of interest in real conditions and application. This paper seeks to bridge biological and machine visions in order to bring the robustness, speed, and accuracy of biological vision into the artificial intelligence world.

Introduction

One of the most important requirements of each smart system is the ability to detect and track moving objects which help them to see, perceive, and react. The process in which the location of one or several objects is estimated in a video sequence based on their appearance, which called *visual tracking*. With the advancements in the field of microelectronics and availability of high-resolution cameras and imaging sensors as well as fast and powerful processors, many hardware limitations for tracking algorithms have been eliminated, and gradually they are the tracking algorithms themselves that lag behind in this advancing trend. Addition, because of the extensive application of visual tracking algorithms in movie and game industries, medicine, auto surveillance, smart robots, self-driving cars, etc. the need for a robust and accurate visual tracking algorithm is highly felt.

In the last two decades, a wide range of studies have been conducted in this field and various visual tracking algorithms have been introduced, trying to do these tasks as ideally as possible. Different methods of appearance and shape representation, different methods of motion estimation, using statistical methods in the phase of detection, and many other techniques (Jalal & Singh, 2012; Li et al., 2013; Smeulders et al., 2014; Yang, Shao, Zheng, Wang, & Song, 2011) are some of these efforts. But in spite of all these efforts, high diversity of states and conditions of a moving stimulus, and existence of abundant challenges such as a cluttered background, changes in the objects appearance, and occlusion (Maggio & Cavallaro, 2011) make this problem very complicated to the extent that even the state-of-art algorithms have not been able to overcome these challenges (Wu, Lim, & Yang, 2015). Therefore, introducing a robust visual tracking algorithm which is able to achieve acceptable accuracy in the face of existing challenges is

* Corresponding author.
E-mail addresses: ar.sokhandan@eng.ui.ac.ir (A. Sokhandan), monadjemi@eng.ui.ac.ir (A. Monadjemi).

http://dx.doi.org/10.1016/j.bica.2016.09.006
2212-683X/

still an open problem, and has a special importance in the field of artificial intelligence and particularly machine vision.

Excellent performance of the human visual system in visual tasks and applications such as detection, recognition, localization, tracking, etc. has inspired researchers to employ this ideal visual system in the development of different algorithms. Consequently, with the advent of neurology, psychology and in general, the cognitive science, various algorithms and models have been introduced by inspiring from the human visual system and achieved significant performances compared to the traditional image processing and machine vision algorithms. Visual tracking is among those problems which have attracted researchers' attentions because of the excellent performance of the biological visual system in this task, in such a way that even under the most challenging conditions from cluttered backgrounds and low lamination to the complicated movement of stimulus, the target has been tracked perfectly.

In the last decade, extensive researches both from the perspectives of psychology (Oksama & Hyönä, 2004) and neurosciences (Culham et al., 1998; Sakata, Shibutani, & Kawano, 1983) have been done which has resulted in several biological visual tracking models. Among the most outstanding models are those introduced by Pylyshyn (Pylyshyn & Storm, 1988) and Kahneman (Kahneman, Treisman, & Gibbs, 1992). These models are not quantitative and only introduce the psychological aspect of tracking a simple stimulus such as a single-point, and they didn't mention how it can be extended in order to track the real objects in real environments. Based on these psychological models, several computational models such as one developed by Kazanovich (Kazanovich & Borisyuk, 2006) and Vul (Vul, Alvarez, Tenenbaum, & Black, 2009) were introduced, and like their own psychological base models, they only evaluated their algorithms on the simple stimulus, and their performance in real application has not been discussed. Beside these algorithms, some other ones employed biologically inspired mechanism - especially the saliency mechanism - in the procedure of their algorithms (Frintrop, 2010; Toyama & Hager, 1996; Wang, Chen, & Xu, 2010; Yang, Yuan, & Wu, 2007). The visual tracking algorithm introduced by Mahadevan and Vasconcelos (Mahadevan & Vasconcelos, 2013) is one of the most outstanding biologically inspired visual tracking algorithms of this group. They pointed out that in the human visual system, the goal of tracking is to fix the object of interest in the fovea (the center of attention of eyes) even in the conditions where the object or observer is moving. The human visual system performs this task based on the attention mechanism, appearance features of the object, and the distinction between spatiotemporal features of the target and background. Inspired by these three principles, Mahadevan and Vasconcelos developed a two-phase framework for visual tracking. In these phases, a bottom-up and top-down saliency is applied to the visual stimulus to localize and track the object of interest. They indicated that this framework can appropriately track objects in real conditions. Their algorithm provides the possibility to use different features in different conditions. The Experimental results indicate that this algorithm has an effective performance in terms of accuracy. However, they only used the information of the last frame to train the statistical model which reduced the generality of the algorithm. Considering the single-point shape representation method, ignoring the motion information of the object, and neglecting the problem of occlusion are the main disadvantages of Mahadevan's tracking algorithm.

In this paper, a computational biologically inspired visual tracking framework is presented based on researches conducted in the mode of performing the visual tracking task in the visual cortex from the psychological, neurological, and cognitive aspects. Studies have shown that the visual tracking is classified as spatiotemporal tasks in the visual cortex, and involves almost all parts of visual cortex and its mechanisms from the shape processing in the ventral pathway to motion perception in the dorsal pathway. In the proposed framework, by investigating different mechanisms involved in the tracking process, and the communication and information flow among them, a six-phase model has been presented. Investigating the procedure of proposed framework from the viewpoint of employed features, and the quiddity of local-global processing, it is indicated that this framework has the capability of managing most of the mentioned challenges, and can robustly track the object of interest in real conditions and application.

In a nutshell, the main contribution of this paper can be summarized as follows: studying the procedure of biological visual tracking task conducted in the visual cortex from the modular view and dividing it into the several executive blocks; introducing a biologically inspired framework for visual tracking algorithm by connecting these blocks – by defining the information flow between them – and executing them in a consecutive way; investigating the presented biologically inspired computational models which can sit in the place of these blocks (with a few changes in their procedure or parameters) and suggesting a candidate implementation of the proposed framework.

The rest of this paper is organized as follows: In Section 'Visual tracking, from biological vision to machine vision' the biological visual tracking and two famous models which explain different aspect of visual cortex function on this task are briefly explained. In Section 'The inspired tracking framework', the suggested framework is introduced, and the procedure of different phases and their biological backgrounds are described. Section 'The proposed framework and visual tracking challenges' explains how the suggested framework handles different tracking challenges and finally, the conclusion is given in Section 'Conclusion'.

Visual tracking, from biological vision to machine vision

The human visual system and generally, biological vision can easily identify objects, localize them, track their motions with high accuracy. In this visual system, the goal of visual tracking is keeping the object of interest in the fovea, which requires estimating the object's motion pattern (Kowler, 2011). This ideal tracker works based on the complicated motion perception system which acts not only based on the visual information received through the eyes, but also the visual features and motion patterns of the object which are learned and stored in short-term memory (Masson & Perrinet, 2012).

Although plenty of researches have been conducted in this field in the last decade, the brain function in tracking of a moving object and the mode of handling existing challenges has not been completely identified. However, it was indicated that the attention mechanism plays a fundamental role in accessing this performance (Koch & Ullman, 1987). The goal of this mechanism is to localize the region of visual scene in which the object or the event of interest is put. This mechanism acts based on the saliency and center-surrounded operator and is the basis of a diverse group of tasks in the biological visual system. One of these tasks that has a significant role in tracking is visual search (Bravo & Nakayama, 1992). Visual search is responsible for recognizing an object of interest from among other objects existing at the scene (which are called distractors). This task is conducted based on visual features of the target, and a mechanism called feature search which identifies and utilizes that features of the target which make distinctions between the target and distractors. There can be only one feature or a collection of several features (such as color, intensity, oriented edges, etc.). Numerous studies have been conducted on the function of this mechanism which Feature Integration Theory (FIT) (Treisman & Gelade, 1980) and Guided Search (GS) (Wolfe, 2007) models the most famous ones.