Signal Processing: *Image Communication* 43 (2016) 28–41

Contents lists available at ScienceDirect

Signal Processing: *Image Communication*

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

Top–down visual attention integrated particle filter for robust object tracking

Wanyi Li [a,*], Peng Wang [a], Hong Qiao [b,c]

[a] *Research Center of Precision Sensing and Control, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China*
[b] *State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China*
[c] *CAS Centre for Excellence in Brain Science and Intelligence Technology (CEBSIT), Shanghai 200031, China*

ARTICLE INFO

Article history:
Received 29 November 2014
Received in revised form
26 September 2015
Accepted 3 January 2016
Available online 2 February 2016

Keywords:
Object tracking
Visual attention
Saliency
Particle filter
Abrupt motion
Occlusion

ABSTRACT

Numerous tracking methods have been proposed and work well under many challenging conditions. However, there are still some problems need to be solved, such as abrupt motion and longtime occlusion. Visual attention mechanism enables humans to efficiently select the visual data of most potential interest and results in robust object tracking. Inspired by this fact, this paper presents a top–down visual attention computational model based on frequency analysis and integrates it into particle filter to solve the above mentioned problems. Given an image sequence, target-related salient regions are detected by the proposed top–down visual attention. Then the target is tracked by the proposed local and global search processes in which the salient regions are incorporated into particle filter. Comparison experiments on challenging sequences demonstrate the effectiveness of the proposed method.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Object tracking is a significant computer vision task and has many practical applications such as visual surveillance, robotics and unmanned vehicles. Despite extensive researches on this topic [1–7], achieving robust tracking performance still remains a huge challenge.

The particle filter based tracker [8,9] is a sampling-based tracking method, which can cope well with the non-linear and non-Gaussian tracking problems. Pérez et al. [10] propose a method to embed the color-based image features into a particle filter tracking framework. A Markov Chain Monte Carlo (MCMC) based particle filter for tracking multiple interacting targets has been proposed in [11], which uses the MCMC method to sample directly from the posterior distribution of the target position. Sherrah et al. [12] presented an adaptation of the particle filter to track people in surveillance video. In this algorithm, detection is based on automated background modelling and the tracks of objects are created by a labelling method. Kwon and Lee [13] presented a compound tracker named as Visual Tracking Decomposition (VTD), which utilizes the basic distinctive components of the observation, motion, and tracking models to efficiently construct compound models. Shan et al. [14] proposed a mean shift embedded particle filter (MSEPF) to improves the sampling efficiency, which move particles to local peaks in the likelihood by incorporating the mean shift optimization into particle filtering. To deal with abrupt motion difficulties, a Wang–Landau Monte Carlo sampling based tracking algorithm (AWLMC) is presented in [15]. This method is more likely to get a global optimum by sampling in the global state space, but it needs lots of particles.

A drawback of the particle filter is that it usually needs a large number of samples to estimate the state of target accurately, especially when abrupt motion and longtime occlusion occur. The increasing number of samples will

* Corresponding author. Tel.: +86 010 82544767.
E-mail address: wanyi.li@ia.ac.cn (W. Li).

http://dx.doi.org/10.1016/j.image.2016.01.001
0923-5965/© 2016 Elsevier B.V. All rights reserved.

result in increasing computational complexity. To obtain candidate regions where target may appear before tracking is one of effective ways to alleviate the above issues.

Robust object tracking is just a basic function of the human visual system (HVS). In the tracking process of HVS, visual attention plays a critical role, which directs the processing resources to the potentially most relevant visual data, especially directs our gaze rapidly towards the objects of interest. As a result, humans can easily achieve robust object tracking. Furthermore, some biological vision literature has suggested that (i) tracking is implemented by attentional mechanisms [16,17] and (ii) attentional processes may facilitate tracking through anticipation or act as an error recovery mechanism [18,19].

Due to the importance of visual attention in tracking process of human visual system and the drawback of the particle filter, this paper proposes a Top–down visual attention integrated particle filter (TAIPF) for robust object tracking. Given an image sequence, the potentially most relevant salient regions with respect to the target are first detected by our proposed top–down visual attention computational model based on frequency analysis. The top–down information related to the target object for the visual attention model is learned at the beginning of tracking process. Then the target is tracked by local and global search processes in which the salient regions are incorporated into particle filter. When the target is with smooth motion or small abrupt motion only, the local search is performed by considering salient regions overlapped with particles sampled by traditional particle filter. When very large abrupt motion or longtime occlusion occurs, global search is performed by sampling particles around each detected salient region to recover tracking.

The main contributions of this paper are twofold.

Firstly, a top–down visual attention computational model based on frequency analysis is presented, which uses the target model as a top–down information and introduces it into frequency analysis attention model. The proposed visual attention model is utilized to construct the top–down saliency map and detect candidate regions where target may appear.

Secondly, to deal with abrupt motion and longtime occlusion, a novel robust object tracking framework named as Top–down visual attention integrated particle filter (TAIPF) is proposed. We integrate salient regions detected by the proposed top–down visual attention model into particle filter and track the target by two search processes, i.e., local search and global search. In the local search process, a set of particles are first sampled by traditional particle filter, and then several salient regions overlapped with pre-sampled particles are selected. After some particles are sampled around selected salient regions, we merged them with pre-sampled particles. In the global search process, particles are sampled around each detected salient region. For both two search processes, the best particle is taken as searched result. Comparison experiments on challenging sequences and public dataset demonstrate the effectiveness and robustness of the proposed method.

The remainder of the paper is organized as follows. First, the relevant work is reviewed in the next section. Then we introduce our tracking algorithm in details in Section 3. Experimental results and analysis are presented in Section 4, and finally we draw conclusions in Section 5.

2. Related work

In this section we discuss the most relevant work including computational modeling of visual attention and saliency based approaches for tracking. More thorough reviews on object tracking can be found in [1–4].

2.1. Computational modeling of visual attention

Visual attention [20] (also referred as visual saliency in literature) is one of the key mechanisms of human visual system that enables humans to efficiently determine the most relevant parts within the large amount of visual data. It is described as a perceptual quality that makes a region of image stand out relative to its surroundings and to capture attention of the observer [21]. In the past decades, many computational models have been proposed to simulate humans' visual attention [22]. These approaches can be divided into two groups: the bottom–up approaches and the top–down approaches.

2.1.1. The bottom–up approaches

Bottom–up attention is purely data-driven and guides the gaze to salient regions in a scene. Regions attracting bottom–up attention are always those with strong contrast or certain uniqueness. Therefore, the approaches in the bottom–up group often aim to detect the unique or rare visual subsets. For example, Itti et al. [23] proposed a classical cognitive concepts inspired visual attention model which is based on a center-surround contrast calculation. Other published bottom–up visual attention models include information theoretic models [24,25], graphical models [26,27] and spectral analysis models [21,28,29], etc. Information theoretic models premise that localized saliency computation serves to maximize information sampled from one's environment and select the most informative parts of a scene and discarding the rest. Graphical models represent an image as a weighted graph and utilize Hidden Markov Models (HMM), Dynamic Bayesian Networks (DBN), or Conditional Random Fields (CRF) to calculate visual saliency. Spectral analysis models derive saliency in the frequency domain in a bottom–up manner and provide state-of-the-art performance in finding salient regions with efficient computation. However, introducing top–down information with considerable importance for tracking to spectral analysis attention models has not been broadly investigated up to now.

2.1.2. The top–down approaches

Top–down attention is driven by cognitive factors such as pre-knowledge, context, expectations, motivations, and current goals. A few types of top–down information that can drive an attention model, such as pre-knowledge, context information, have been realized in computational system. For instance, based on Itti's basic model [23], the VOCUS attention model [30] uses pre-knowledge about a target to weight the feature maps and perform visual search. Torralba et al. [31] use context information about the scene to guide the gaze, e.g., to search for people on the street level of an image rather than on the sky area. Zhu et al. [32] propose a top–down computational model for goal-driven saliency detection based on the coding-based classification framework, in which

Download English Version:

https://daneshyari.com/en/article/537174

Download Persian Version:

https://daneshyari.com/article/537174

Daneshyari.com