Engineering Applications of Artificial Intelligence 37 (2015) 307–318

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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

Online adaptive motion model-based target tracking using local search algorithm

Amir Hossein Karami [a], Maryam Hasanzadeh [a,*], Shohreh Kasaei [b]

[a] Department of Computer Engineering, Shahed University, Tehran, Iran
[b] Department of Computer Engineering, Sharif University of Technology, Tehran, Iran

ARTICLE INFO

Article history:
Received 19 May 2014
Received in revised form
16 August 2014
Accepted 25 September 2014
Available online 21 October 2014

Keywords:
Abrupt motion
Adaptive motion model
Computer vision
Local search
Sequential importance resampling particle filter
Visual tracking

ABSTRACT

An adaptive tracker to address the problem of tracking objects which undergo abrupt and significant motion changes is introduced. Abrupt motion of objects is an issue which makes tracking a challenging task. To address this problem, a new adaptive motion model is proposed. The model is integrated into the sequential importance resampling particle filter (SIR PF), which is the most popular probabilistic tracking framework. In this model, in each time step, if necessary, the particles' configurations are updated by using feedback information from the observation likelihood. In order to overcome the local-trap problem, local search algorithm with best improvement strategy is used to update particles' configurations. Then, the motion model is updated online with respect to the configurations of the best particle in the current and previous time steps. By using this adaptive model, a more robust tracking is achieved to abrupt significant motion changes. The tracker is experimentally compared to other state-of-the-art trackers on BoBoT dataset. The experimental results confirm that the tracker outperforms the related trackers in many cases by having better PASCAL score. Furthermore, this tracker improves the accuracy of the conventional SIR PF approximately 15%.

1. Introduction

Visual tracking is an important topic in computer vision which has many applications, such as surveillance (Benfold and Reid, 2011; Huang and Fu, 2011; Huang and Cheng, 2012; Rho et al., 2012) activity recognition (Choi and Savarese, 2012; Huang et al., 2014), robotics (Mekonnen et al., 2013; Richa et al., 2010) video summarization (Tavassolipour et al., 2013), and human–computer interaction (HCI) (Lupu et al., 2013; Xu and Lu, 2014). Since visual tracking is an important topic in computer vision, many approaches have been proposed in the literature, and it has been extensively studied in the recent decades. Among these methods, the particle filter (PF) is the most popular one. PF is first introduced in Isard and Blake (1998a, 1998b), which is also known as sequential importance resampling particle filter (SIR PF). PF offers a flexible framework which is able to handle non-linear and non-normal systems, and also it is able to deal with multimodal distribution and recover from tracking failure. In addition, there are many applications which use the SIR PF framework to track targets. Bruno (2004) discussed various Bayesian-based target tracking methods in the SIR PF framework for image sequences with cluttered background and random target aspect. Moreover, Danaee and Behnia (2013), in order to cope with time-varying target presence problem, extended the SIR PF through split measurements. Consequently, PF becomes the most widely used framework in visual tracking systems.

One of the most fundamental problems in visual tracking is *abrupt motion*, in which the target's motion model changes suddenly, which happens in situations such as unexpected target dynamic, target fast motion, camera switching, and low-frame-rate videos. In real world applications, many trackers are based on a smooth motion assumption or an accurate motion model to facilitate efficient tracking. In most cases, the motion of objects is governed by a first-order auto-regressive (AR1) or second-order auto-regressive (AR2) process. As a result, scenarios that contain rapid motion changes present conventional trackers with serious difficulties. This is because these trackers fail to optimally predict the new position of the target; therefore, updating the motion model potentially increases the robustness of the tracker in such situations.

In general, tracking methods can be categorized as deterministic (Yang et al., 2005) or sampling-based ones (Khan et al., 2004, 2005). In most cases, sampling-based strategies formulate the tracking problem using probability distributions of random variables to model uncertainty in the models and measurements. Deterministic methods offer an alternative means to search the solution space using non-statistical approaches. Deterministic and sampling-based methods face challenges when dealing with the large motion uncertainty induced by abrupt

* Corresponding author. Tel.: +98 21 51212046; fax: +98 21 51212021.
E-mail addresses: ah.karami@shahed.ac.ir (A.H. Karami), hasanzadeh@shahed.ac.ir (M. Hasanzadeh), skasaei@sharif.edu (S. Kasaei).
URL: http://sharif.edu/~skasaei (S. Kasaei).

http://dx.doi.org/10.1016/j.engappai.2014.09.018

motions. Enlarging the sampling variance to cover the motion uncertainty is a direct solution for the sampling-based tracking methods. However, the sampling inefficiency is a problem, which is raised here because the increase in the sampling volume may require a more expensive computational cost, particularly for the systems with the high-dimensional state-space. It is worth noting that the sampling inefficiency problem is of particular concern for the case of multi-object tracking, where the dimension of the state-space grows proportionally to the number of objects represented.

In this paper, a new adaptive motion model is described which is able to update motion model of the tracker online (*i.e.*, *on-the-fly*) by using local search algorithm. This model is integrated into PF framework because of its success which highly relies on its ability to maintain a good approximation to the posterior distribution.

The proposed tracker does not change the number of particles. Just for sequences with abrupt and fast camera motion, it updates the motion model if necessary. Consequently, the proposed method is effective and computationally efficient to track targets at real-time. Moreover, the experimental results demonstrate that the proposed tracker is robust to abrupt significant motion changes.

The remainder of this paper is organized as follows: In Section 2, some related tracking algorithms are reviewed briefly. In Section 3, the proposed method is described. The experimental results are reported in Section 4. The last section concludes the paper.

2. Related work

Visual tracking has been a popular topic for several years. In this section, only the most related tracking methods, focusing on algorithms that directly deal with the abrupt motion problem are reviewed. Please refer to Yang et al. (2011), which is a rich literature in visual tracking, for more complete reviews on recent visual tracking methods.

The first popular solution to the abrupt motion problem is hierarchical (or layered) and multiscale sampling strategy which focuses on improving the sampling efficiency in the solution space. Sullivan et al. (1999), in a Bayesian context, to facilitate efficient searching in the fine scale, combined observations from multiple scales. Inaccurate inference in the large scale which may cause the failure of searching in a fine scale is a potential problem of this method. Hua and Wu (2004) designed a collaborative searching and matching scheme based on the dynamic Markov network to overcome the potential error propagation problem. In order to achieve a more accurate Bayesian inference, they also developed a bi-directional-based belief-propagation algorithm. In contrast to multiscale-based trackers, Li et al. (2008) applied a pipeline of complementary observation models (also known as cascade PF) on the same image space to deal with the large motion uncertainty of the target object. They believed that an LFR (low frame rate) condition is equivalent to abrupt motion for a tracking system. Their approach has proved to be highly efficient in LFR conditions, but it requires an additive offline learning process and several reliable observation models.

Other approaches based on iterative optimization such as the Kanade–Lucas–Tomasi (KLT) (Tomasi and Kanade, 1991) and mean shift (Comaniciu et al., 2000) feature trackers generally require some assumptions which are often not satisfied under conditions of LFR or abrupt motion. They require feature patches or the kernels in consecutive frames to be close to or overlap with each other.

To guide the proposal distribution of particle filter, Okuma et al. (2004) proposed an offline-boosted detector which they called Boosted particle filter (BPF). Moreover, Porikli and Tuzel (2005) developed the conventional mean shift algorithm through optimizing multiple kernels at motion areas. These areas were obtained by background modeling to track a pedestrian. Their ideas required using an independent detector to amend the search of an existing tracker in case the target motion is unpredictable and fast. The common limitation of these methods is that they require a fast detector to be applied to large search areas (in most cases, the whole frame).

Detecting and connecting (Han et al., 2004; Kaucic et al., 2005; Wu and Nevatia, 2006) is another method which has the potential of dealing with abrupt motion problem. This is because it first detects the targets of interest and then constructs and connects trajectories via analyzing motion continuity, appearance similarity, *etc.* The drawback of this category is that when background is not fixed, a fast detector is required.

In another point of view, the target tracking problem can be viewed as a two-part problem: motion detection and target matching (Ke et al., 2011). The motion detection part generally entails modeling the background, and then subtracts it from a frame in a video sequence to find the moving target (Ke et al., 2011). Maddalena and Petrosino (2008) proposed a background subtraction method based on the *self-organizing map* (SOM) via *artificial neural networks* (ANNs), which can adopt artificial neural network in scenes which contain moving background and illumination changes. Huang (2011) proposed an accurate motion detection approach via a rapid matching procedure. His approach yields the optimum distinguishes between background and foreground pixels. More recently, Guo et al. (2013) proposed the concept of multilayer block-based codebook to construct the background model. With that model, the reliability and processing speed of the approach are improved against dynamic background subtraction problem. Chen and Huang (2014) presented an applicable approach in various bit-rate video streams for motion detection, which uses lower-dimensional eigen-patterns. In their work, in order to construct an accurate background model, they used the principal component analysis-based radial basis function network (PCA-based RBF network).

McKenna and Nait-Charif (2007) proposed an algorithm which could detect sequences with abrupt, fast camera motion automatically. Although they applied the same motion model for the entire sequence, they also employed the Iterative Likelihood Weighting procedure. To this end, the particles are divided into two sets. The first set is normally propagated, and the second one is iteratively propagated. After that, the particles are weighted several times in order to let extreme movements happen within one time step.

There has recently been a tendency to introduce some learning methods into tracking, in which the aim is distinguishing the tracking target from the background. Therefore, tracking can be viewed as a classification problem. Klein et al. (2010) for real-time visual tracking of arbitrary objects applied a particle filter-based approach. Their tracker uses an adaptive observation model. To this end, a robust classifier consisting of an ensemble of Haar-like center-surround features was proposed. Their proposed classifier was learned from a single positive training example with Gentle AdaBoost and is adaptive in facing new target and background appearance changes in each time step.

Frintrop (2010)'s approach to target tracking automatically detects the most discriminative parts of a target in addition to learning its appearance from a single frame. Stalder et al. (2009) by using object specific and adaptive priors presented a semi-supervised tracking approach.

Searching the whole state space in order to cover the motion uncertainty completely is another simple solution to the abrupt motion problem. However, practically, the large search space of the target state makes this infeasible, and it results in an expensive computational cost. To estimate the search space based on the target state prediction, an accurate dynamic model is required. Nevertheless, such models are often learned from the specific training data (Isard and Blake, 1998a, 1998b; North and Blake, 1998), which removes flexibility from tracking algorithms.

In comparison to other mentioned approaches, the proposed tracker offers a criterion to detect when the motion vector of the target changes. Therefore, if the tracker detects abrupt motion of the