Pattern Recognition 48 (2015) 126–139

Contents lists available at ScienceDirect

Pattern Recognition

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

Object tracking based on an online learning network with total error rate minimization

Se-In Jang [a], Kwontaeg Choi [b], Kar-Ann Toh [a,*], Andrew Beng Jin Teoh [a], Jaihie Kim [a]

[a] School of Electrical and Electronic Engineering, Yonsei University, Seoul, Republic of Korea
[b] Department of Computer Science, Yonsei University, Seoul, Republic of Korea

ARTICLE INFO

Article history:
Received 9 November 2012
Received in revised form
12 March 2014
Accepted 25 July 2014
Available online 12 August 2014

Keywords:
Object tracking
Particle filter
Self-adaptation
Random projection network
Online learning

ABSTRACT

This paper presents a visual object tracking system which is tolerant to external imaging factors such as illumination, scale, rotation, occlusion and background changes. Specifically, an integration of an online version of total-error-rate minimization based projection network with an observation model of particle filter is proposed to effectively distinguish between the target object and the background. A re-weighting technique is proposed to stabilize the sampling of particle filter for stochastic propagation. For self-adaptation, an automatic updating scheme and extraction of training samples are proposed to adjust to system changes online. Our qualitative and quantitative experiments on 16 public video sequences show convincing performances in terms of tracking accuracy and computational efficiency over competing state-of-the-art algorithms.

1. Introduction

1.1. Background and motivation

Visual object tracking, which captures the location and orientation of an object in video sequence, has been a focused research field over the past few decades due to its high commercial application values [1–3]. Accordingly, a lot of works attempting problems such as deformable tracking, abrupt changes, occlusion and illumination changes can be found under computer vision, security and multimedia applications.

Generally, existing methods to object tracking can be classified into generative and discriminative approaches. Techniques under the generative approach attempt to build a model based on the object appearance of interest [4]. However, they cannot discriminate an object of interest from the background because they consider only object similarity. On the other hand, techniques under the discriminative approach attempt to separate the object from the background. This observation is evident from several recent developments in the literature [5–7] which showed better tracking accuracy than that of the generative approach.

For example, under the discriminative approach, a Support Vector Tracking (SVT) [5] technique was proposed to integrate Support Vector Machine (SVM) and optical flow. Instead of minimizing an intensity difference function between successive frames, the SVT maximizes the SVM classification score. More recently, a context tracker [8] was proposed using distracters and supporters. Distracters are regions which have a similar appearance to the object of interest. Supporters are local key-points around the object where they play an important role in verifying the genuine object. Another good example is a Multiple Instance Learning (MIL) tracker [9] which trains a discriminative classifier in an online manner to separate the object from the background. This classifier bootstraps itself by using the current tracker state to extract positive and negative examples from the current frame.

However, the discriminative approach does not cater for arbitrary object tracking and adaptation to appearance changes. For example, the OAB [9] tracks only fixed-sized rectangular object. When the object changes in size, some background patches will be included in the object region. This leads to performance degradation. Although [7,10,11] attempted to address the issue of size variation, their tracking remained focusing on rectangular objects. The SVT in [5] was applied to track only vehicles. Moreover, the SVT adopted a fixed pre-trained SVM without adaptation. In [12], a color-based meanshift method was adopted to update a color model according to appearance change. However, the color-based method can be easily attracted to other objects and background with a similar color distribution.

Apart from the above issues, the following challenges need to be addressed in order to achieve both efficiency and effectiveness in an object tracking system.

- High dimensionality: Visual objects are frequently represented by feature vectors with high dimensionality. Unfortunately,

* Corresponding author.
E-mail address: katoh@ieee.org (K.-A. Toh).

http://dx.doi.org/10.1016/j.patcog.2014.07.020
0031-3203/

learning of classifiers under such a feature space is computationally expensive. Although Haar-like features can be adopted as low dimensional feature [5,7], such methods are limited to tracking of objects with affine motion.

- Small sample size problem: Learning methods, which involve an inverse operation such as linear discriminant analysis and least squares solution, would suffer from performance instability when feature dimension is high but training samples is small. This leads to performance degradation.
- Computational cost: The online versions of existing trackers [6,7,10,11] are limited in terms of efficiency since their updating costs for high dimensional features are computationally expensive.

In this paper, we attempt to address these issues using an Online Random Projection Network [13] based on Total-Error-Rate (TER) minimization [14–16].

1.2. Contributions and organization

To address arbitrary object tracking, we propose to integrate discriminative and generative methods into an observation model of particle filter based on online learning with self-adaptation in this work. Four major contributions of the proposal are enumerated as follows:

- We propose an accurate and efficient system for object tracking under illumination and background variations. In the proposed system, the object is separated from the background through an Online version of Total-Error-Rate minimization Network (OTERN) [14,17] which utilizes a Random Projection Network (RPN) [13] while adapting to appearance changes through a Sum of Squared Differences (SSD) [18] in particle filtering.
- We propose to stabilize the sampling of particle filter based on confidence obtained from both OTERN and SSD for stochastic propagation.
- Instead of a uniformly periodic update scheme as seen in [4,6,7], we propose an automated updating scheme based on a self-adjusting threshold technique to determine whether or not to update the learning parameters. This reduces updating trials which require high computation cost under high dimensional space.
- We propose to estimate the target label among those unlabeled data accumulated during tracking to alleviate the drifting problem.

The paper is organized as follows: Section 2 provides a brief review to literatures related to object tracking. In Section 3, some preliminaries related to our proposed method are provided for immediate reference. In Section 4, we introduce the proposed method which integrates OTERN and SSD into an observation model of particle filtering with self-adaptation. Section 5 presents both qualitative and quantitative comparisons with state-of-the-art methods on publicly available video sequences. Finally, some concluding remarks are given in Section 6.

2. A brief literature review

According to [1–3], existing literatures for object tracking can be categorized into three approaches: (i) point-based approach, (ii) contour-based approach and (iii) kernel-based approach (see Fig. 1). For the purpose of positioning our proposed method, a brief overview will be covered with more attention paying to the kernel-based approach where the proposed system belongs to.

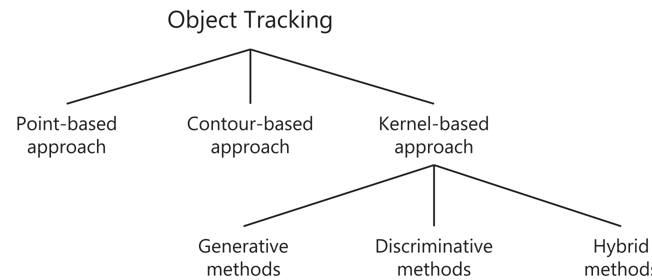

Fig. 1. Main approaches to object tracking.

2.1. Point-based approach

Under this approach, an object of interest is represented by a set of feature points such as corners of the object. According to the methods used in tracking, existing point-based trackers can be further divided into deterministic and statistical methods [1,2]. Deterministic methods adopt a one-to-one correspondence matching among all possible points as a greedy search. Representative works [19–21] showed reliable tracking performances under partial occlusion condition. However, they appeared to suffer from unpredictable perturbations which arose from intrinsic and extrinsic changes [19–21].

Different from those deterministic methods, statistical methods utilized uncertainty information obtained in the previous state to establish the correspondences between frames. In [22–24], Kalman and particle filters [25] were investigated to estimate the states of a linear system in object tracking. However, due to the lack of information for the points, several problems such as multiple objects in a scene, object drifting and significant appearance changes during tracking [22–24] remained unsolved.

2.2. Contour-based approach

Unlike the point-based approach, tracking methods under the contour-based approach provide localization of the target object with its boundaries (e.g. edges and lines). According to [26], contour-based methods are commonly referred to as Active Contour Model (ACM) or snakes. According to how the contours being described and used in tracking, existing contour-based methods can be divided into parametric ACM and geometric ACM [27].

Parametric ACM represents curves of the target object in parametric forms during curve evolution [26]. The main goal here is to minimize an energy function (e.g. intrinsic and extrinsic energy function [26]) which indicates the goodness of boundaries. As a representative work, Xu et al. [28] proposed a gradient vector flow field based on the vector diffusion of the gradient of an edge map [28] to solve problem. According to [1,2], the parametric ACM methods were observed to suffer from a similar appearance and multiple objects problems. Moreover, it showed difficulties in terms of topological adaptation such as an object splitting and merging.

In order to resolve the shortcomings observed in parametric ACM methods, geometric ACM based approach was proposed [29]. The multiple objects and topological problems were effectively handled by propagating the curves with a level set which consists of points labeled with the same values [29]. However, the contour-based approach cannot resolve those problems caused by intrinsic and extrinsic factors (e.g. appearance change (intrinsic), the background change (extrinsic) and occlusion (extrinsic)).

2.3. Kernel-based approach

In computer vision, a kernel can be associated with an object appearance such as a rectangular shape and an elliptical shape