Signal Processing 112 (2015) 146–153

Contents lists available at ScienceDirect

Signal Processing

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

3D object tracking via image sets and depth-based occlusion detection

Yan Chen, Yingju Shen, Xin Liu, Bineng Zhong *

Department of Computer Science and Technology, Huaqiao University, Jimei District, Xiamen, Fujian 361021, China

A R T I C L E I N F O

Article history:
Received 15 March 2014
Received in revised form
7 August 2014
Accepted 16 August 2014
Available online 16 September 2014

Keywords:
3D
Object tracking
Image sets
Depth information

A B S T R A C T

Object tracking has been typically formulated as an online learning problem, in which a target appearance is updated adaptively using the single-shot images tracked from the previous 2D frames. The traditional methods have the following limitations: (1) useful depth information is ignored; (2) each training and testing example is a single image, so it can be easily corrupted. In this paper, we propose a novel 3D object tracking method using image sets and depth-based occlusion detection, in which each training and testing example contains a set of image instances of an object and covers large variations in the object's appearance. To do so, we first represent each image set as its natural second-order statistic. Then, we use kernel partial least squares to adaptively learn low-dimensional discriminative feature subspace for object representation. Third, to prevent improper appearance model updating during occlusions, we exploit depth information obtained from binocular video data to detect occlusion. Finally, to alleviate the tracker drifting problem caused by model update, we exploit both the ground truth appearance information of the object labeled in the initial frames and the image observations obtained online. Extensive experiments on challenging video sequences demonstrate the robustness and effectiveness of the proposed method.

1. Introduction

Visual tracking has attracted increasing interest in computer vision and pattern recognition due to its wide potential applications, including intelligent video surveillance, human–machine interfaces, robotics, and so on. Although much progress has been made in the last two decades, designing robust visual tracking methods is still a challenging problem. Challenges in visual tracking problems include non-rigid shapes, appearance variations, occlusions, illumination changes, cluttered scenes, low resolution, etc.

To solve these challenges, most top-performing methods rely on online learning-based algorithms [1–30] to adaptively update target appearance. In these methods, visual tracking is formulated as an online binary classification problem and the target appearance models are updated adaptively using the images tracked from the previous frames. Specifically, for online learning based visual tracking, a tracker observes testing examples (typically image patches) in each frame and predicts their labels to be either foreground or background. At the end of each frame, the adaptive tracker uses the newly obtained example-label training pairs to improve its prediction rule for the following frames. However, one key issue in the online learning based tracking literature is how to generate a set of training and testing examples when estimating the object location and updating the adaptive appearance model. Traditionally, in the training

* Corresponding author. Tel.: +86 592 6162556.
E-mail address: bnzhong@gmail.com (B. Zhong).

http://dx.doi.org/10.1016/j.sigpro.2014.08.046
0165-1684/

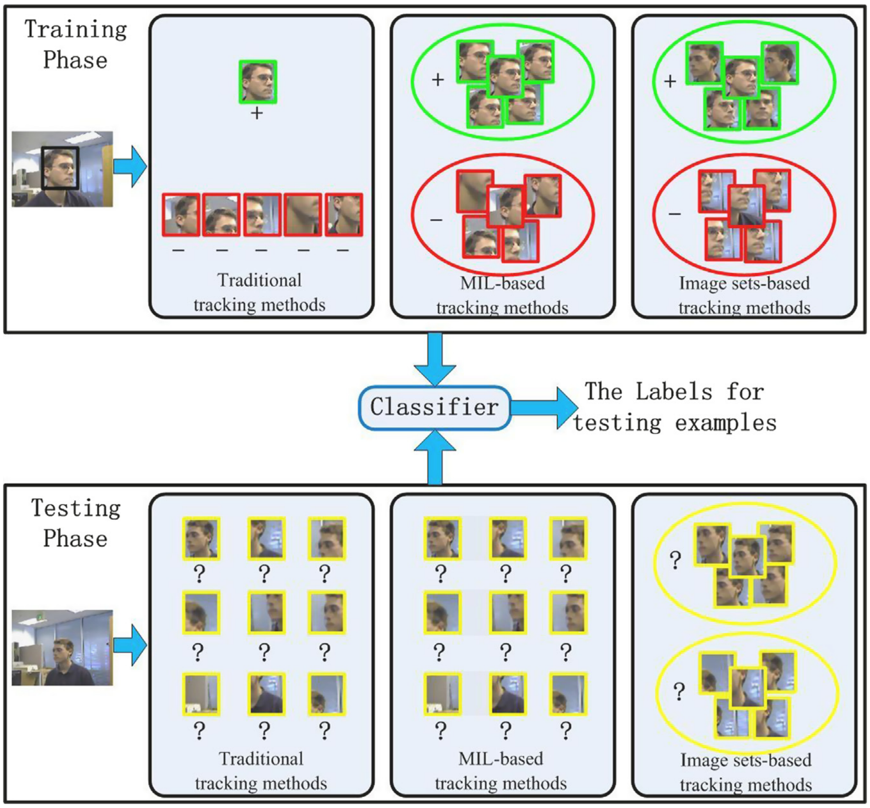

Fig. 1. Illustration of the training and testing examples used in different tracking methods. Different from the traditional tracking methods in which each training and testing example is a single image, for the image sets-based tracking methods, each training and testing example contains a set of image instances. The setting of MIL-based tracking methods can be viewed as a special case of the setting of image sets-based tracking methods where each testing example is a single image.

phase, the image patch from the current object location is selected as one positive training example, and the negative training examples are generated by selecting the image patches which are relatively far from the object location. In the testing phase, the image patches from the neighborhood around the estimation from the previous frame are selected as testing examples. Consequently, slight inaccuracies/offsets in the object location can lead to incorrectly labeled training examples, which degrades the tracker and can cause further the model drift problem.

In this paper, inspired by ideas of image set classification, we overcome this limitation through considering visual tracking from sets of images. As shown in Fig. 1, different from the traditional tracking methods in which each training and testing example is a single image, for the image sets-based tracking methods, each training and testing example contains a set of image instances. Alternatively, Babenko et al. [17] have recently proposed a Multiple Instance Learning (MIL) method for visual tracking, in which each training example is presented in set (often called “bag”), and each testing example is a single image. While traditional tracking methods based on single-shot images have achieved a certain level of success under restricted conditions, more robust visual tracking can be expected by using sets as input rather than single images. This is mainly because the image set incorporates useful data variability information, which can be efficiently exploited under more realistic conditions with significantly larger variations [31–43].

There is an interesting relation with the method of Babenko et al. [17] and the traditional tracking methods [1–16,18–30]. When each testing example is a single image, our image sets-based tracking method is equivalent to the setup of Babenko et al. Furthermore, when each training and testing example is a single image, our image sets-based tracking method is also equivalent to the setup of the traditional tracking methods.

The rest of the paper is organized as follows: Section 2 reviews the related work. The overview of the proposed tracking algorithm is described in Section 3. In Section 4, image sets-based appearance model is presented. The occlusion detection and recovery method is then described in Section 5. In Section 6, experimental results are given. Finally, the conclusions are drawn in Section 7.

2. Related work

Visual tracking is an extensively studied research topic. Please refer to the survey papers [44–46] and a recent benchmark [47] for a comprehensive survey of this topic. In the following, we first discuss image set classification and its applications in relevant research communities. Then, we review recent online tracking methods using different types of target appearance models, namely,