Signal Processing 112 (2015) 154–161

Contents lists available at ScienceDirect

Signal Processing

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

# Robust depth-based object tracking from a moving binocular camera

Liujuan Cao*, Cheng Wang, Jonathan Li

*School of Information Science and Engineering, Xiamen University, China*

ARTICLE INFO

*Article history:*
Received 18 March 2014
Received in revised form
20 August 2014
Accepted 25 August 2014
Available online 10 September 2014

*Keywords:*
3D Object tracking
Depth information
Local patch-based appearance model

ABSTRACT

Depth is a rich source of information and has been successfully utilized in numerous computer vision applications. However, it is often ignored in object tracking. In this paper, in contrast to traditional 2D image-based tracking method, we propose a novel 3D object tracking method from a moving binocular camera. To effectively handle the deformable targets, a target is first represented by a local patch-based appearance model. Then, to handle the partial occlusions, we design a simple yet effective scheme to detect and recovery occlusions using depth information obtained from a moving binocular camera. Therefore, the proposed method can simultaneous capture target appearance changes and alleviate the drifting problem. The experimental results demonstrate the effectiveness of the proposed method.



## 1. Introduction

Object tracking is one of the basic tasks in numerous computer vision applications, such as intelligence video surveillance, human machine interfaces, special effects in motion pictures, indexing for multimedia, and so on. Robust object tracking will greatly improve the performance of object intelligence video surveillance, human machine interfaces and activity analysis. However, designing robust object tracking methods is still an open issue, especially considering various complicated variations that may occur in dynamic scenes, e.g., illumination variations, pose changes, background clutters, occlusions, etc.

To achieve the goal of robust object tracking, a large number of 2D image-based methods using different features and learning schemes have been proposed over the years. However, the 2D image-based tracking methods are easily corrupted by the noises and cannot effectively handle the occlusions. Differently, depth information obtained from a binocular moving camera can provide potentially useful information to deal with the occlusions.

In this paper, we present a novel 3D object tracking method from a moving binocular camera that learns a robust patch-based target appearance model and explicitly handles the occlusions by using depth information. The key idea is firstly to utilize a local patch-based appearance model to represent the target. Since the target is represented by a local patch-based appearance model, the proposed tracking method can effectively handle the deformable targets. Then, to handle the partial occlusions, we use a simple yet effective scheme to detect and recovery occlusions using depth information obtained from a moving binocular camera, which is more robust than existing 2D image-based tracking methods. Experimental results on challenging video sequences demonstrate the robustness of the proposed 3D tracking method by comparing it with several state-of-the-art tracking methods.

The rest of the paper is organized as follows: Section 2 reviews the related work. The overview of the proposed

* Correspondence to: Haiyun Park, Xiamen University, 361005, Xiamen, Fujian, China. Tel.: +86 592 6162556.
*E-mail address:* caoliujuan@xmu.edu.cn (L. Cao).

http://dx.doi.org/10.1016/j.sigpro.2014.08.041
0165-1684/

tracking algorithm is described in Section 3. The local patch-based target appearance model is described in Section 4. The tracking process and occlusion detection scheme are presented in Section 5. A summary of the proposed tracking method is given in Section 6. In Section 7, experimental results are given. Finally, the conclusions are drawn in Section 8.

## 2. Related work

To achieve robust object tracking even in complexity dynamic scenes, a number of tracking methods have been proposed.

In the tracking literature, one popular technique is to track object using fixed appearance models [1–3]. These methods assume that object will look nearly identical in each new frame. Thus, an appearance model of the object from the first frame can be always used to describe object appearance. However, these fixed appearance model-based tracking methods cannot achieve long-term robust tracking in dynamic scenes, which often requires addressing difficult target appearance update problem. To handle this problem, a number of authors have formulated the problem of visual tracking as an online learning problem, in which the target appearance is updated adaptively using the images tracked from the previous frames. Collins et al. [4] present a method to adaptively select one color feature from several different color spaces to construct adaptive appearance models, which can best discriminate the object from the current background. In Ref. [5], Avidan proposes a method using an adaptive ensemble of classifiers for object appearance model maintenance and tracking. Unfortunately, one inherent problem of online learning-based trackers is drift, a gradual adaptation of the tracker to non-targets.

To alleviating the drifting problem, a number of top-performing tracking methods have recently been proposed. Matthews et al. [6] propose a method by making sure the current tracker does not stray too far from the initial appearance model. Within the semi-supervised learning framework, Grabner et al. [7] treat all incoming samples as unlabeled data. One of the key limitations of the above methods is that very large changes would cause the failures. In [8] and [9], a multiple instance boosting based technique and a co-training based technique are respectively proposed to deal with the drifting problem. In [10] and [11], a structured output support vector machine is applied to object tracking. In [12], Gall et al. propose a Hough forests-based visual tracking method. Lu and Hager [26] propose a model adaptation method driven by feature matching and feature distinctiveness that is robust to drift. Oron et al. [27] develop another method to deal with the drift problem by automatically estimating the amount of local (dis)order in an object. In [28], a discriminative metric is learned for robust visual tracking. In addition, with the popularity of low-rank subspaces and sparse representations in image processing and machine learning, a variety of low-rank and sparse representations based tracking methods have been recently proposed [29–31].

Rather than using only 2D images, a number of authors have used 3D or stereo-based information to improve the performances of the computer vision systems [14–17,32–34,36–38]. For object tracking literature, Hu et al. [13] propose a principal axis-based stereo tracking method. Ess et al. [14] propose a robust multi-person tracking method from a mobile platform, in which depth information is used to verify object candidates obtained by object detection. Some authors focus on tracking the human body by using RGB-D cameras [15–17]. In [32], Choi and Christensen propose a RGB-D object tracking method using a particle filter on GPU. Kooa et al. [33] propose a novel model-free approach for tracking multiple objects from RGB-D point set data. In [34], Ren et al. propose a probabilistic framework for simultaneous tracking and reconstruction of 3D rigid objects using an RGB-D camera. However, their performances are severely impaired in an outdoor environment due to complicated illumination changes. Moreover, Kinect requires a minimum and a maximum distance from objects to the cameras in order to obtain accurate depth values.

Please refer to [18–21] for more complete reviews on the tracking methods.

## 3. Overview of the proposed tracking method

In this section, we develop our depth-based tracking algorithm from a moving binocular camera. The flowchart of the proposed tracking method is shown in Fig. 1.

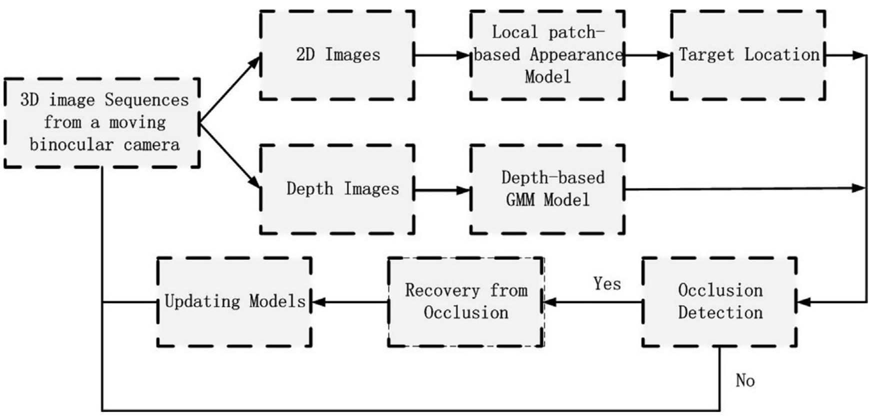


**Fig. 1.** The flowchart of the proposed tracking method.