Neurocomputing 267 (2017) 581–596

Contents lists available at ScienceDirect

Neurocomputing

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

Two-level superpixel and feedback based visual object tracking

Jun Wang[a], Weibin Liu[a,*], Weiwei Xing[b], Shunli Zhang[b]

[a] Institute of Information Science, Beijing Jiaotong University, Beijing 100044, China
[b] School of Software Engineering, Beijing Jiaotong University, Beijing 100044, China

ARTICLE INFO

Article history:
Received 20 November 2016
Revised 8 May 2017
Accepted 12 June 2017
Available online 30 June 2017

Communicated by Xiang Bai

Keywords:
Visual tracking
Superpixel
Feedback
Two-level superpixel

ABSTRACT

While numerous superpixel-based tracking algorithms have been proposed and demonstrated successfully, there still remain some challenges, such as determining the number of superpixels, mining and exploiting the structural information of superpixels and handling the drifts. In this paper, we propose a tracking method with two-level superpixels and a novel update strategy based on feedback to deal with the challenges mentioned above. Firstly, Bilateral filter is introduced to filter out outliers and improve the boundary capability of object as well as segmentation of superpixels. Then two-level superpixel is proposed to determine superpixel number automatically through iterating instead of setting superpixel number empirically which affects the robustness of tracking algorithm. Moreover, a novel measuring method which considers color similarity and relative positions of superpixels is proposed to make a better use of structural information of superpixels and improve tracking performance by adding relative position of superpixels into the appearance model. Finally, a feedback based update strategy is presented to handle drifts existing in tracking by calculating the adaptation of appearance model and updating the parameters like superpixel number and relative position of superpixels. Experiments on challenging sequences and comparisons to state-of-the-art methods demonstrate the feasibility and effectiveness of the proposed tracking algorithm.

1. Introduction

Visual object tracking is one of the classical fundamental challenge in computer vision and artificial intelligence and has a wide range of applications in real life, such as scene surveillance, video retrieval, automatic transmission, human-computer interaction, etc. Yet, there still remain many challenging issues to solve due to illumination variation, occlusion, background clutter, deformation, scale and pose change and camera motion (as shown in Fig. 1).

To deal with the challenges, numerous sophisticated trackers with various features (such as histogram feature, textural feature, Histogram of Oriented Gradient (HOG) [1], etc.) and appearance models have been proposed, which can generally be divided into two categories: tracking using matching [2–4] and tracking using discriminative classification [5–9]. To address the appearance variation in sequences, many representations of appearance model have been developed which can be divided into three levels: top level, middle level and bottom level, according to the extracted features. Among the three levels, middle-level object representation can also be divided into two categories generally: the patches [10–13] and superpixels [8,9,14–16]. For top-level based appearance model, all the pixels within the bounding box are considered as a whole which makes the tracker more robust to background clutter and partial occlusion, but more sensitive to severe occlusion and deformation because of the changing of global statistical characteristics. On the contrary, the bottom-level based appearance model extracts some particular pixels (known as keypoints) to represent the object, thus the change of statistical feature has less influence over the representation. This kind of appearance models can handle occlusion and deformation better, but tend to fail under the scenarios when background clutter happens. Superpixel-based appearance model has the advantages of both top-level and bottom-level appearance model to solve the alternative problem above. Simple Linear Iterative Clustering (SLIC) [17] superpixel which considers the difference of pixels themselves instead of spatial distance of pixels has been widely used according to its convenience and effectiveness.

However, there still exist several deficiencies of superpixel-based trackers. Firstly, the number of superpixels needs to be set empirically which varies with different tracking sequences and affects the robustness and efficiency of the tracking algorithm. If it is set too large which means too many superpixels are involved to represent the object, the computational and representation complexity will increase, and it is also detrimental to the global

* Corresponding author.
E-mail addresses: wbliu@center.bjtu.edu.cn, wbliu@bjtu.edu.cn (W. Liu).

http://dx.doi.org/10.1016/j.neucom.2017.06.031
0925-2312/

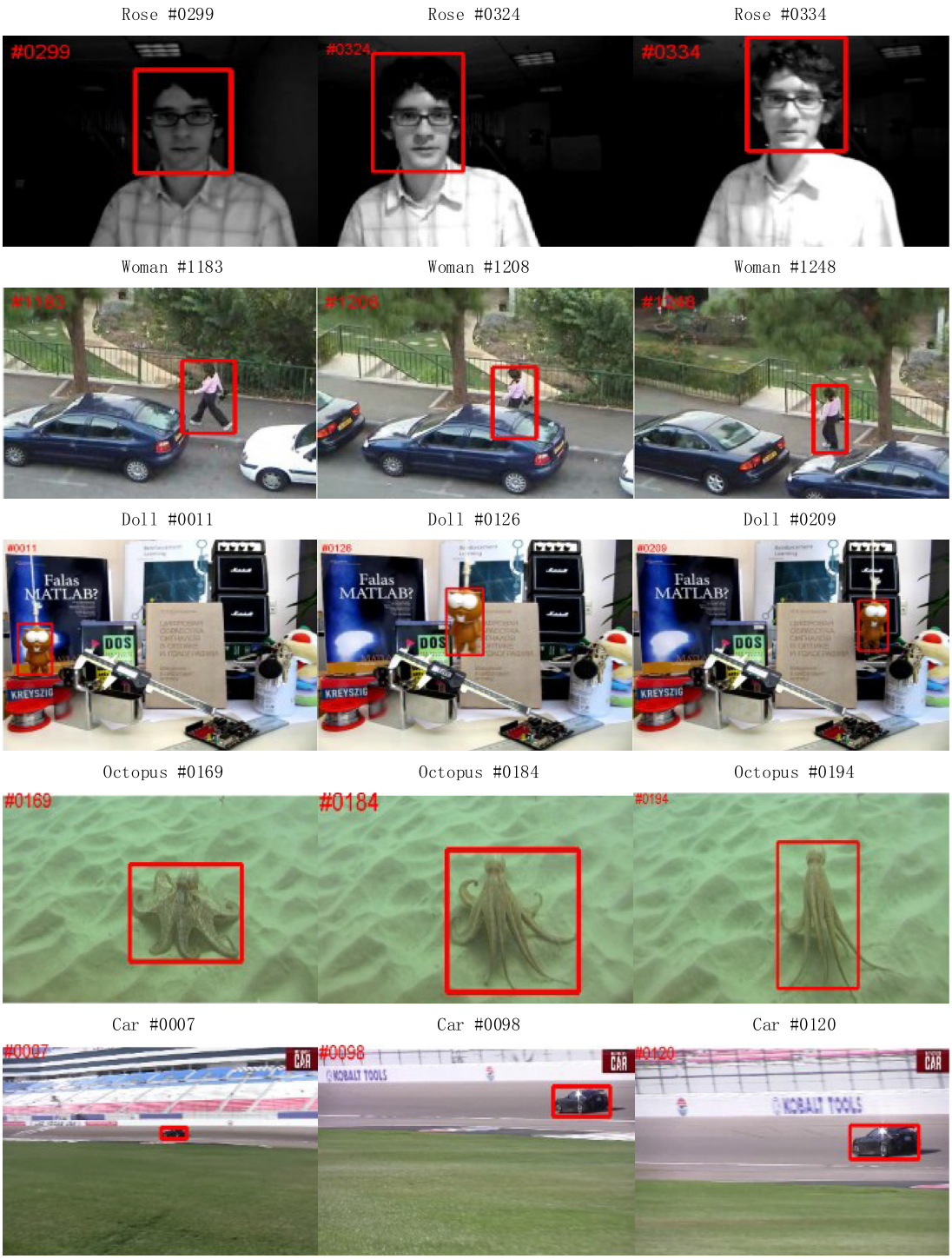

Fig. 1. Five common challenges encountered in visual object tracking. The challenges from top to bottom are illumination variation, occlusion, background clutter, deformation and camera motion respectively. Red rectangles are the ground truth of each video sequence marked by human being. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

characteristics of object. On the other hand, if the number is set too small, the trackers will show better performance under the scenarios of background clutter, and computational complexity decreases in the process of feature extraction, but fail to handle the situations when deformation and occlusion occur. Secondly, superpixel centers obtaining through the means of all pixels within their superpixels are sensitive to the noises, and general Gaussian filter will blur the images and weaken the boundaries of object. Thirdly, conventional confidence map calculated by superpixel-based tracker [8,9] does not take structural information of superpixels into consideration. Finally, most online update strategies are open-loop system whose updating process only depends on the threshold of iteration and lacks certain evaluation criterion [18–22].

Motivated by the above observation, we proposed a two-level superpixel and feedback based tracker, which represents target at two superpixel levels to form the confidence map and calculate the state of object, and then updates confidence map and appearance