Pattern Recognition 48 (2015) 580–590

Contents lists available at ScienceDirect

Pattern Recognition

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

Multi-target tracking by learning local-to-global trajectory models

Shun Zhang, Jinjun Wang *, Zelun Wang, Yihong Gong, Yuehu Liu

Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, China

ARTICLE INFO

Article history:
Received 14 March 2014
Received in revised form
10 July 2014
Accepted 13 August 2014
Available online 3 September 2014

Keywords:
Local-to-global
Trajectory model
Markov random field
Belief propagation
Iterative update

ABSTRACT

The multi-target tracking problem is challenging when there exist occlusions, tracking failures of the detector and severe interferences between detections. In this paper, we propose a novel detection based tracking method that links detections into tracklets and further forms long trajectories. Unlike many previous hierarchical frameworks which split the data association into two separate optimization problems (linking detections locally and linking tracklets globally), we introduce a unified algorithm that can automatically relearn the trajectory models from the local and global information for finding the joint optimal assignment. In each temporal window, the trajectory models are initialized by the local information to link those easy-to-connect detections into a set of tracklets. Then the trajectory models are updated by the reliable tracklets and reused to link separated tracklets into long trajectories. We iteratively update the trajectory models by more information from more frames until the result converges. The iterative process gradually improves the accuracy of the trajectory models, which in turn improves the target ID inferences for all detections by the MRF model. Experiment results revealed that our proposed method achieved state-of-the-art multi-target tracking performance.

1. Introduction

Vision based multi-target tracking aims at locating all targets of interest, inferring their trajectories and maintaining their identities from image observations in a video sequence. It is important for many computer vision applications such as video surveillance, robotics and activity analysis. Compared with single target tracking task, multi-target tracking is significantly more difficult because it has to face the following challenging problems. First, targets may be partially or completely occluded by other targets or foreground objects and become unobservable for a short time, which may confuse tracking algorithms and may result in target losses or ID switches between the targets. Second, since targets may enter or leave the scene at any moment, the number of targets is usually unknown and may vary over time. And last but not least, it is difficult to distinguish targets with similar appearances, especially when they are close to, or partially occlude each other.

In recent years, many of the successful tracking methods perform tracking by detection. These methods apply a pre-learned object detector to conduct object detections in every frame through the video stream. The tracking problem is solved by the *data association* [1–8] which strives to establish a unique identity for each target, and to simultaneously estimate the motion patterns of all targets and the assignment of detections to targets, by linking similar detections across frames. This method is essentially more flexible and robust in complex environments, since it can not only naturally handle re-initialization in tracking when a target is lost, but also avoid excessive model drift. However, it poses another difficult challenges. Since the detector's output is only partly reliable, missed detections (false negatives) and incorrect detections (false positives) may happen frequently in the detection process, which provides misleading information to association algorithms. If the targets are overlapped, the task is further complicated and it is much more difficult to retrieve the real targets among those detections and assign the labels of detections for each of them in every frame.

To deal with this data association problem, many of the previous methods [9–14] associated the detections locally, *i.e.* using local information from a few neighboring frames or frame by frame. These methods generally integrated several cues of the image information such as appearance, motion, size and location [15–19] to measure the similarity between detections from two consecutive frames. Given only the image information in a small time window, the local association methods are difficult to tackle the long-term occlusion due to the ambiguous and noisy observations, which make them incline to result in tracking failures (*e.g., trajectory fragmentation and identity switches).*

In contrast to the local tracking methods, many of the latter approaches [20–22,7,23,24] have achieved great progress by using global inference over all trajectories simultaneously in a longer period. As they consider more global information, these association

* Correspondence author. Tel./fax: +86 29 83395146.
E-mail address: jinjun@mail.xjtu.edu.cn (J. Wang).

http://dx.doi.org/10.1016/j.patcog.2014.08.013
0031-3203/

approaches can do better to overcome errors by detections and association. With the increasing of the frame numbers, however, the hypothesis search space of those alternatives grows exponentially, so that these global association algorithms usually require lots of computation which make them unsuitable for real-time processing. Moreover, they also typically assume that all detections are correct which is not always accurate [7].

Some more recent approaches [25–29] combined the local linking method and the global association method in a kind of hierarchical framework, which first linked easy-to-connect detections in neighboring frames into tracklets, and then progressively linked tracklets into longer ones based on some global association approaches, such as the Hungarian algorithm [30] and the network flow [31]. They presented a formulation with two separate optimization problems: linking detections are solved using the local information in lower stages and linking tracklets are solved using the global information in higher stages. Splitting the problem in two phases has, obviously, several disadvantages because the available evidence is not fully exploited [32]. Furthermore, transforming from the local process to the global process is generally specified heuristically and experimentally.

In this paper, we propose a unified multi-target tracking method to progressively grow local tracklets into long trajectories by combining the local and global information. Unlike many previous hierarchical methods which heuristically formulate the data association method as two separate optimization schemes, we present a unified algorithm for finding the joint optimal assignment. We online learn the trajectory model (TM) for each target to maximize a joint object function. By repeating the proposed algorithm, the trajectory model can be learned more accurately from local to global information. The trajectory model is composed of a set of cues including appearance, velocity, size and position. We segment the whole long video stream into multiple non-overlapping sliding windows. For each temporal sliding window, the trajectory models are firstly initialized by very limited, local image information (*e.g.* information from the first frame). The initial trajectory models are used to link those visually resembling, easy-to-connect detections in neighboring frames into separated tracklets. Then the appearance, motion and some other parameters of trajectory models are updated by those reliable tracklets, such that the trajectory models become more accurate. The updated trajectory models are reused to link the separated tracklets that belong to the same target. As the iterative process continues, the trajectory models become accurate, and the broken tracklets get connected to form longer trajectories.

The main contributions of this paper include the following:

- A unified framework which online learns the local-to-global trajectory models.
- An iterative algorithm to alternately update the trajectory models and link detections or tracklets into longer fragments.
- Formulating the data association problem as inferences of target IDs for all the detections using the MRF model.
- And employing the loopy belief propagation (LBP) algorithm to optimize the MRF model so as to generate separated tracklets, which has low complexity compared with most state-of-the-art methods.

Our experimental results show superior tracking performance on several public datasets and computational speed of the proposed method.

The rest of this paper is organized as follows. Section 2 briefly describes related works in the literature. Section 3 introduces an overview of our approach. Sections 4 and 5 present the proposed multi-target tracking framework and its inference algorithm, respectively. Section 6 provides the performance evaluation and comparison results, and Section 7 concludes the paper.

2. Related work

Fostered by the recent progress in object detection techniques such as [33–35], there is a rich body of multi-target tracking works based on tracking by data association approaches. They applied an object detector learned off-line or on-line to yield the per-frame detections. However, the detector posed many misleading errors including missed detections, false alarms and inaccurate detections. Detection based tracking method must overcome the failures of the detector, and the difficulties caused by occlusions, initialization and termination of targets and similar appearance among multiple targets. To overcome the data association problem, there are three main strategies: associating the detections locally, associating them globally, and combining the local method and global association method.

Much effort [9–14] has been made to deal with the data association problem locally. A frame-by-frame tracking method in [9] presented an affinity measure between detections based on cues from position, size, and color and used a greedy algorithm to match the detection responses and hypotheses. The trajectory initialization and termination were based on the evidence collected from the detection responses. This association method depends a lot on the detector and easily results in tracking failures when the targets are occluded for a long period. In [10], a mixture particle filter method was used to associate the Adaboost detections of an unknown number of objects. It assigned a mixture particle filter to each target, and constructed the proposal distribution for the particle filter from a mixture of the Adaboost detections in the current frame and the dynamic model predicted from the previous time step. The detection responses were used to generate new particles and evaluate existing particles. However, increasing the number of particles requires more computational cost. The local association methods are likely to lead to drift when multiple targets are close to each other, since the noisy target detections significantly increase the difficulty of data association.

To consider associations beyond the local basis, a rich body of data association approaches propose inference over multiple targets by seeking to resolve the drift problem in a longer period. Multi-Hypothesis Tracking (MHT) [36] and Joint Probabilistic Data Association Filters (JPDAF) [37] are among the earliest widely used techniques for global data association. They maintain multiple hypotheses until enough evidence can be collected to resolve the ambiguities. Recently, a variety of global data association approaches [20–22,7,23,24] which try to simultaneously optimize all trajectories by diverse optimization algorithms have been developed. For example, [24] used Viterbi algorithm to get optimal object sequences, [38,23,39] used the Hungarian algorithm to simultaneously optimize all trajectories, [22] used Quadratic Boolean Programming to couple the detection and estimation of trajectory hypotheses, [2] formulated multi-target tracking as the maximum weight-independent set problem, [21,40] proposed a linear programming approach to search multiple paths, [41,42] involved the minimization of complex energy functions and relied on brute-force algorithms to search for locally optimal solutions, and [7] used a min-cost network flow to model the MAP data association problem. Although these approaches have been demonstrated to improve tracking performance, they are computationally exponential both in memory and time.

The hierarchical association framework [27,43,44,29,25,28] combined the local linking methods and global association methods for better tracking performance in detection based tracking literature. These methods tackled the tracking problem by progressively connecting short detections or tracklets into longer ones. They typically split the data association into two separate optimization problems: linking detections locally in lower stages and linking tracklets globally in higher stages.