Computer Vision and Image Understanding 154 (2017) 94–107

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

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

Online multi-object tracking via robust collaborative model and sample selection

Mohamed A. Naiel[a], M. Omair Ahmad[a,*], M.N.S. Swamy[a], Jongwoo Lim[b], Ming-Hsuan Yang[c]

[a] Department of Electrical and Computer Engineering, Concordia University, Montreal, QC H3G 1M8, Canada
[b] Department of Computer Science and Engineering, Hanyang University, Seoul 113-791, Republic of Korea
[c] School of Engineering, University of California, Merced, CA 95344 USA

ARTICLE INFO

Article history:
Received 5 September 2015
Revised 24 April 2016
Accepted 17 July 2016
Available online 21 August 2016

Keywords:
Multi-object tracking
Particle filter
Collaborative model
Sample selection
Sparse representation

ABSTRACT

The past decade has witnessed significant progress in object detection and tracking in videos. In this paper, we present a collaborative model between a pre-trained object detector and a number of single-object online trackers within the particle filtering framework. For each frame, we construct an association between detections and trackers, and treat each detected image region as a key sample, for online update, if it is associated to a tracker. We present a motion model that incorporates the associated detections with object dynamics. Furthermore, we propose an effective sample selection scheme to update the appearance model of each tracker. We use discriminative and generative appearance models for the likelihood function and data association, respectively. Experimental results show that the proposed scheme generally outperforms state-of-the-art methods.

1. Introduction

Multi-object tracking (MOT) is one challenging vision problem with numerous applications in automatic visual surveillance, behavior analysis, and intelligent transportation systems, to name a few. In the past decade, more attention has been paid on detecting and tracking one or more objects in videos. Recent advancement in object detection facilitates collaboration between the detection and tracking modules for multi-object tracking (Breitenstein et al., 2009).

Robust multi-object tracking entails solving many challenging problems such as occlusion, appearance variation, and illumination change. A pre-trained object detector robust to appearance variation of one specific class is often used as a critical module of most multi-object tracking methods. Specifically, one detector encodes the generic pattern information about a certain object class (e.g., cars, pedestrians and faces), and one tracker models the appearance of the specific target to maintain the target identity in an image sequence. However, an object detector is likely to generate false positives and negatives, thereby affecting the performance of a tracker in terms of data association and online model update.

In multi-object tracking, offline methods based on global optimization of all object trajectories usually perform better than online counterparts (Andriyenko and Schindler, 2011; Andriyenko et al., 2012; Brendel et al., 2011; Butt and Collins, 2012; Izadinia et al., 2012; Leal-Taixé et al., 2011; Shitrit et al., 2011; Wu et al., 2012; Zamir et al., 2012), and an experimental evaluation of recent methods can be found in Leal-Taixé et al. (2015). For instance, Brendel et al. proposed the maximum-weight independent set of a graph for data association (Brendel et al., 2011), and Zamir et al. used the generalized minimum clique graph to solve the data association (Zamir et al., 2012). In Butt and Collins (2012), the data association problem is solved by using a sliding window of three frames to generate short tracklets, and in case of inconsistencies, the algorithm uses larger tracklet optimization. The minimum-cost network flow is then used to optimize the overall object trajectories. For real-time applications, online methods (Breitenstein et al., 2009; Okuma et al., 2004; Shu et al., 2012; Wu et al., 2008) have been developed within the tracking-by-detection framework where data association between detections and trackers are carried out in an online manner.

Table 1 summarizes the multi-object tracking methods that are most related to this work. Online multi-object tracking can be carried out by using joint state-space model for multi-targets (Duffner and Odobez, 2013; Eiselein et al., 2012; Jin and Mokhtarian, 2007; Maggio et al., 2008; Okuma et al., 2004; Vermaak et al., 2003).

* Corresponding author.
E-mail addresses: m_naiel@ece.concordia.ca (M.A. Naiel), omair@ece.concordia.ca (M.O. Ahmad), swamy@ece.concordia.ca (M.N.S. Swamy), jlim@hanyang.ac.kr (J. Lim), mhyang@ucmerced.edu (M.-H. Yang).

http://dx.doi.org/10.1016/j.cviu.2016.07.003

Table 1
Representative online multi-object tracking algorithms. AM: appearance model, J/I: joint/independent, MU: model update, PF: particle filter, KF: Kalman filter, CVM: constant velocity motion model, MCMC: Markov Chain Monte Carlo, CH: color histogram, LBP: local binary patterns, BOW: bag of words, DCD: detection confidence density, SS: sample selection, PGF: probabilistic gating function, $q(.)$: proposal distribution, SGM: sparsity-based generative model, PGM: 2DPCA-based generative model, SDC: sparsity-based discriminative classifier.

Algorithm	Search and proposal distribution	J/I	Sample descriptor	Data association	MU	Likelihood function
Okuma et al. (2004)	Mixture particle filters, $q(.)\propto$ new observation and propagated particles	J	HSV CH	NA	No	Bhattacharyya similarity
Breitenstein et al. (2009)	PF with CVM	I	RGI, LBP	Boosted classifier, PGF and position	Yes	Distance between each particle and the associated detection, DCD, and Boosted classifier
Yang et al. (2009)	Bayesian filtering	I	RGB, shape, BOW	CVM, position and scale	Yes	Joint likelihood of AM features
Shu et al. (2012)	Detector based or KF	I	CH, LBP	SVM classifier, position, size	Yes + SS	KF (if no associated detection to the tracker)
Zhang et al. (2012)	Mean shift tracker or KF	I	CH, shift vector	Size, search area, tracker re-detection	Yes	Combination of Mean-shift and KF
Schumann et al. (2013)	PF with random walk or CVM	I	RGB CH	Overlap ratio	Yes	Detector confidence
Duffner and Odobez (2013)	MCMC with random walk, $q(.)\propto$ new detections and sampled particles	J	HSV CH	Overlap ratio and position, tracker re-detection	Yes	The product of the visible individual targets likelihoods
Proposed method	PF with detection based CVM, $q(.)\propto$ new associated detections and propagated particles	I	Grayscale	Overlap ratio, SGM and PGM, tracker re-detection	Yes + SS	SDC, different weights for newly created and propagated particles

For instance, a mixture particle filter has been proposed (Okuma et al., 2004) to compute the posterior probability via the collaboration between an object detector and the proposal distribution of the particle filter. However, the joint state-space tracking methods require high computational complexity. The probability hypothesis density filter (Mahler, 2003) has been incorporated in visual multi-target tracking (Maggio et al., 2007; Maggio et al., 2008) since the time complexity is linear with respect to the number of targets. However, it does not maintain the target identity, and consequently, requires an online clustering method to detect the peaks of the particle weights and applies data association to each cluster.

Numerous online multi-object tracking methods deal with each tracker independently (Breitenstein et al., 2009; Schumann et al., 2013; Shu et al., 2012; Yang et al., 2009; Zhang et al., 2012). In Breitenstein et al. (2009), a method based on a particle filter and two human detectors with different features was developed, where the observation model depends on the associated detection, the detector confidence density and the likelihood of appearance. In addition, Shu et al. (2012) introduced a part-based pedestrian detector for online multi-person tracking. This method combines the detection results with the Kalman filter, where data association is performed every frame, and the filter is used when occlusion occurs. Recently, Zhang et al. (2012) used the mean-shift trackers and the Kalman filter for multi-person tracking, where trackers are either weakly or strongly trained. We note that these methods are likely to have low recall as the detector and tracker are not integrated within the same framework.

The degeneracy problem of particle filters (Gordon et al., 1993) has been addressed in several methods (Huang and Djuric, 2004; Jinxia et al., 2012; Rui and Chen, 2001; Santhoshkumar et al., 2013) with more effective proposal distributions and re-sampling steps. Rui and Chen (2001) used the unscented Kalman filter for generating the proposal distribution, and Han et al. (2011) used a genetic algorithm to increase the diversity of the particles. Recently, the Metropolis Hastings algorithm has been used to sample particles from associated detections in the tracking-by-detection framework (Santhoshkumar et al., 2013). We note that the above-mentioned methods do not exploit the collaboration between detectors and trackers (Han et al., 2011; Rui and Chen, 2001), or do not consider the effect of false positive detections on the trackers (Santhoshkumar et al., 2013).

An adaptive appearance model is one of the important factors for effective object tracking as it accounts for appearance change (Salti et al., 2012; Wu et al., 2013). In Okuma et al. (2004), the appearance model is fixed during the tracking process and thus, may result in tracking failure. On the other hand, the trackers are updated with positive samples (Zhang et al., 2012) straightforwardly without differentiating whether they contain noise or not. As multiple objects are likely to be occluded, it is necessary to analyze the samples and reduce the likelihood of including noisy samples for model update. Recently, the appearance models (Shu et al., 2012) have been updated by the detected non-occluded object parts rather than the holistic samples.

In this paper, we propose an online multi-object tracking scheme by using a robust collaborative model for interaction between a number of single-object trackers with sparse representation-based discriminative classifiers (Wright et al., 2009; Zhong et al., 2012), and a pre-trained object detector in the particle filter framework, where every target is tracked independently to avoid the high computational complexity of the joint probability with increasing number of targets. A novel sample selection scheme is used to update each tracker by using key samples with high confidence from the trajectory of an object, where the key sample represents the association between the tracker and a detection at time, t. In addition, we present a data association method with partial occlusion handling by using diverse generative models composed of sparsity-based generative model (Zhong et al., 2012), and two-dimensional principal component analysis (2DPCA) (Yang et al., 2004) generative model. Finally, we introduce a 2DPCA generative model to re-identify lost targets. Experimental results on benchmark datasets demonstrate that the proposed scheme generally outperforms state-of-the-art methods.

2. Overview of the proposed scheme

The proposed multi-object tracking scheme consists of three main components: a pre-trained object detector, a data association module and a number of single-object trackers. Fig. 1 shows the block diagram of the proposed scheme, wherein only one single-object tracker is shown. The object detector is applied on every frame and supports the data association module with a set of detections $\mathcal{D}^t$ at time t. The object tracker adopts a hybrid motion