

Intelligent Querying in Camera Networks for Efficient Target Tracking

Anil Sharma

Indraprastha Institute of Information Technology, Delhi, India
 anils@iiitd.ac.in

Abstract

Visual analytics applications often rely on target tracking across a network of cameras for inference and prediction. A network of cameras generates immense amount of video data and processing it for tracking a target is highly computationally expensive. Related works typically use data association and visual re-identification techniques to match target templates across multiple cameras. In this thesis, I propose to formulate this scheduling problem as a Markov Decision Process (MDP) and present a reinforcement learning based solution to schedule cameras by selecting one where the target is most likely to appear next. The proposed approach can be learned directly from data and doesn't require any information of the camera network topology. NLPR MCT and DukeMTMC datasets are used to show that the proposed policy significantly reduces the number of frames to be processed for tracking and identifies the camera schedule with high accuracy as compared to the related approaches. Finally, I will be formulating an end-to-end pipeline for target tracking that will learn a policy to find the camera schedule and to track the target in the individual camera frames of the schedule.

1 Introduction

Target tracking in camera networks is an essential task for scene understanding, traffic flow analysis, anomalous behavior detection, and many other visual analytics applications. The multiple cameras in the camera network are deployed for surveillance in buildings, airports, roads, etc., for automated tracking of known and unknown targets such as pedestrians or vehicles. Tracking a target in a single camera frame requires to identify the location of the target at all time-steps. Tracking in a network of cameras is more complicated as it not only requires to identify the location of the target in each camera but also the camera schedule which is not a trivial task. These approaches are driven by the tasks of visual detection, tracking and re-identification methods. Multiple methods using deep learning [Ristani and Tomasi, 2018], estimation of camera link models with transition time modeling [Lee *et al.*, 2018] were proposed for multi-target multi-camera (MTMC)

tracking. A few works also focus on estimation of the camera schedule. For example, formulation of camera scheduling as control's strategy [Qureshi and Terzopoulos, 2006], camera selection in a distributed camera network [Tessens *et al.*, 2014; Zhiqiang *et al.*, 2014]. However, we focus on learning a policy for querying in the camera network. We use reinforcement learning to learn the policy directly from data without assuming any information about the topology of the camera network.

Target tracking has also been extensively explored in the literature using various approaches [Ristani and Tomasi, 2018; Cai and Medioni, 2014; Lee *et al.*, 2018; Chen *et al.*, 2017; Chen and Bhanu, 2017]. Many approaches use a two-step framework to enable tracking in multiple cameras. First, single camera tracking (SCT) tracks the target in a single camera field-of-view (FOV) using appearance and temporal features. Second, inter-camera tracking (ICT) resolves the camera handovers when the target transitions from one camera to another by matching the single camera tracks of the target. ICT is also performed when the target gets occluded in a single camera FOV. The matching of target template is achieved using data association and visual re-identification techniques. During ICT, multiple cameras are searched by making re-identification queries to find the target. All-pair matching across all cameras is an NP-hard problem and does not scale with number of cameras and number of targets. Hence, to minimize this search space, the related approaches make use of the camera network topology [Chen *et al.*, 2017], some model the transition time using probabilistically [Lee *et al.*, 2018; Javed *et al.*, 2008]. Recently, [Lee *et al.*, 2018] proposed a two-step method that learns the camera-link model based on appearance features of the target. They also model the transition time using a Gaussian distribution to reduce the search space during ICT. Whereas, in our method, we do not make assumption of the camera network topology and do not explicitly model the transition time of the target.

Various related methods handle the target handovers using visual re-identification by matching the target's current template with the potential target templates in the candidate cameras. This becomes a challenging task in non-overlapping camera field of views because of the non-deterministic and unknown transition time of the handovers. Ideally, one should not select any camera during this transition period because

querying a selected camera would incur higher computational cost and may lead to a potential false alarm which will adversely affect the tracking performance. Therefore, an intelligent camera selection strategy will provide benefits in computational cost as the target is matched in a very few camera frames and will also reduce false alarms. In our experiments, we showed that the policy implicitly learns the camera network topology and it outperforms the various baseline approaches for camera selections.

In this thesis, I assert that the transition time of handovers is time-varying and dependent on characteristics of target like its speed and destination. Such characteristics can be captured directly from the data and as a result, we propose a policy that learns to schedule cameras to make re-identification queries to look for the presence of the target. We model our inter-camera tracking (ICT) approach as a reinforcement learning problem that learns a policy to select an action of querying one of the candidate cameras or not querying any camera. I propose to formulate the tracking problem as a Markov Decision Process (MDP) and present reinforcement learning based solutions to schedule cameras by selecting one where the target is most likely to appear. The purpose is to learn a policy which can result in scalable processing of the large amount of video data. Using such a policy, the inter-camera handovers can then be performed on the selected cameras using any target association or re-identification method. This formulation doesn't require the knowledge of camera network but using various experiments, we have shown that the learned policy implicitly learns the network topology.

The policy is expected to select right camera at any given state which will provide the camera schedule and the transition time. Also, the camera networks generate enormous amount of video data and scalable processing of this data is of vital importance for practicable solutions. Performing template matching in a single candidate camera will provide computational benefits as compared to all-pair matching. We have shown that our policy processes fewer number of frames as compared to other related approaches. The transition time can be inferred from the sequence of null camera selection. The camera schedule should provide us the period of visibility (when present in at least one camera) and invisibility (when not present in any of the camera). We learn such a policy using reinforcement learning. In our experiments, we have shown that such a policy can be learned directly from data from the trial-and-error based approach. In our experiments, we have evaluated our approach with real data (NLPR MCT dataset [Chen *et al.*, 2017] and DukeMTMC dataset [Ristani and Tomasi, 2018]) where the targets are restricted to pedestrians, and compare them to the state-of-the-art methods.

The specific contributions of this thesis are:

1. In this thesis, I have formulated the camera scheduling problem as an MDP. I have demonstrated a method for target tracking in a camera network.
2. I have proposed a reinforcement learning based approach to solve above formulation which learns a policy to select a camera where the target is most likely to appear. Through various experiments, I have shown that the policy can be learned directly from data and unlike

other methods; it doesn't assume any information of the camera network topology.

3. Lately, I have been formulating the scheduling task that self learns the policy using reinforcement learning without any ground truth information.

2 Future Work

In the future work, I would be working on an end-to-end pipeline for the target tracking in a camera network. To achieve this, the policy has to learn to select cameras and the corresponding location of the target in individual cameras (to enable single camera tracking) and during camera handovers (to enable inter camera tracking).

References

- [Cai and Medioni, 2014] Yinghao Cai and Gérard Medioni. Exploring context information for inter-camera multiple target tracking. In *IEEE Winter Conference on Applications of Computer Vision*, pages 761–768, March 2014.
- [Chen and Bhanu, 2017] Xiaojing Chen and Bir Bhanu. Integrating social grouping for multitarget tracking across cameras in a crf model. *IEEE Transactions on Circuits and Systems for Video Technology*, 27(11):2382–2394, Nov 2017.
- [Chen *et al.*, 2017] Weihua Chen, Lijun Cao, Xiaotang Chen, and Kaiqi Huang. An equalized global graph model-based approach for multicamera object tracking. *IEEE Transactions on Circuits and Systems for Video Technology*, 27(11):2367–2381, Nov 2017.
- [Javed *et al.*, 2008] Omar Javed, Khurram Shafique, Zee-shan Rasheed, and Mubarak Shah. Modeling inter-camera space-time and appearance relationships for tracking across non-overlapping views. *Comput. Vis. Image Underst.*, 109(2):146–162, February 2008.
- [Lee *et al.*, 2018] Young-Gun Lee, Zheng Tang, and Jenq-Neng Hwang. Online-learning-based human tracking across non-overlapping cameras. *IEEE Transactions on Circuits and Systems for Video Technology*, 28(10):2870–2883, Oct 2018.
- [Qureshi and Terzopoulos, 2006] Faisal Z. Qureshi and Demetri Terzopoulos. Surveillance camera scheduling: a virtual vision approach. *Multimedia Systems*, 12(3):269–283, Dec 2006.
- [Ristani and Tomasi, 2018] Ergys Ristani and Carlo Tomasi. Features for multi-target multi-camera tracking and re-identification. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.
- [Tessens *et al.*, 2014] Linda Tessens, Marleen Morbee, Hamid Aghajan, and Wilfried Philips. Camera selection for tracking in distributed smart camera networks. *ACM Trans. Sen. Netw.*, 10(2):23:1–23:33, January 2014.
- [Zhiqiang *et al.*, 2014] Song Zhiqiang, Zhou Xianzhong, Li Wei, Li Huaxiong, and Wang Huiping. Scheduling strategies of relay tracking for network-based multiple unmanned ground vehicles. *Proceedings of the 33rd Chinese Control Conference*, pages 7943–7947, 2014.