

Grouping Model for People Tracking in Surveillance Camera

Xiaojing Chen

Department of Computer Science
University of California, Riverside

xchen010@ucr.edu

Bir Bhanu

Center for Research in Intelligent Systems
University of California, Riverside

bhanu@cris.ucr.edu

Abstract

Person tracking and analysis from images or videos captured by closed-circuit television (CCTV) play an important role in forensics applications. In this extended abstract, we describe a tracking model by group analysis. This framework improves over state-of-the-art tracking algorithms by leveraging an online learned social grouping behavior model. This group model is practical in real-world applications where group changes (e.g., merge and split) are natural among pedestrians.

1. Motivation

Automatic tracking of multiple targets is crucial for many security based applications such as visual surveillance, and anomaly detection. The goal of multi-target tracking is to recover trajectories of all targets while maintaining consistent identity labels. There are many challenges for this problem, such as illumination and appearance variation, occlusion, and sudden change in motion. As great improvement has been achieved in object detection, data association-based tracking (DAT) has become popular recently [4]. In the DAT framework, often a pre-learned detector is applied on each frame to produce detection responses of all targets, and short-term tracking results (i.e., tracklets) are generated by associating responses from consecutive frames that have high probability to contain the same target. These tracklets are further linked to produce long-term tracking results.

However, the DAT based tracking methods face problems such as identity switch and track fragmentation still exist in current association based tracking approaches, especially under challenging conditions where appearance or motion of the target changes abruptly and drastically, as shown in Fig. 1.

Nevertheless, there is often other useful high level contextual information in the scene which can be effectively used to mitigate the aforementioned shortcomings. For instance, sociologists have found that up to 70% of pedestri-



Figure 1. Examples in which grouping information is helpful under the challenging conditions for tracking in a video. The same color indicates the same target.

ans tend to walk in groups in a crowd, and people in the same group are more likely to have similar motion pattern and be spatially close to each other for better group interaction [3]. Moreover, pedestrians in the crowd often either consciously or unconsciously follow other individuals with similar destination to facilitate navigation [2]. It is also observed in many real world surveillance videos that if two people are walking together at certain time then it is very likely that these two people will still walk together after a short time period.

2. Approach

Based on the above observations, we propose an elementary grouping model with non-linear motion context to compensate the errors caused by using basic appearance model and linear motion model. A grouping graph is constructed based on input tracklets with high confidence, where each node represents a pair of tracklets that form an elementary group (a group of two targets) and each edge indicates that the connected two nodes (two elementary groups) have at least one target in common. The group trajectories of any

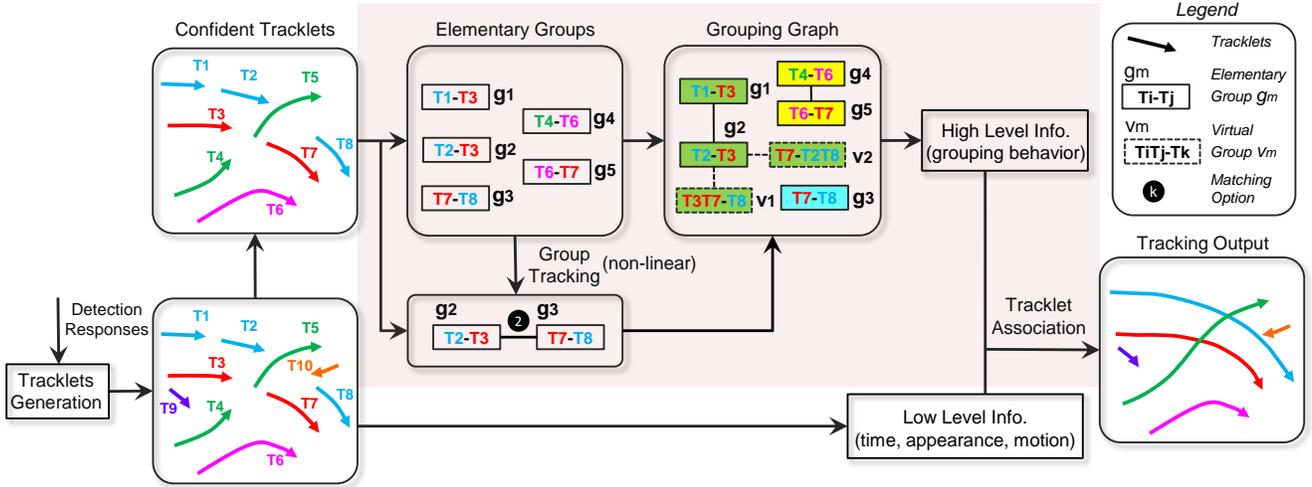


Figure 2. Block diagram of our tracking system. After initial tracklets are generated by linking detection responses, confident tracklets are selected to form elementary groups. The relationships between elementary groups are identified by group tracking with non-linear motion context. Then a disjoint grouping graph is constructed, from which high level information (i.e., grouping behavior) is extracted. Finally, tracklet association is carried out based on affinity model that combines both high level and low level information. Tracklets with the same color contain the same target.

two linked nodes are used to estimate the probability of the other target in each group being the same person. Neighboring tracklets that have time overlap and similar motion pattern are possible candidates for elementary groups. Relationships between elementary groups are further discovered with the help of group tracking, in which a non-linear motion map is used to explain large time gap between two elementary groups. An overview of the proposed method is shown in Fig. 2. The elementary grouping model is summarized in Fig. 3.

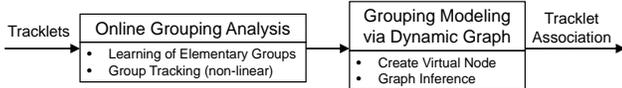


Figure 3. Overview of the elementary grouping model.

The size of a group may change dynamically as people join and leave the group, but a group of any size can always be considered as a set of elementary groups. Therefore, focusing on finding elementary groups instead of the complete group makes our approach capable of modeling flexible group evolution [1] in the real world. Note that the social group in our case refers to a number of individuals with correlated movements and does not indicate a group of people who know each other.

3. Results

We evaluate our approach on four datasets: the CAVIAR dataset, the TownCentre dataset, the PETS2009 dataset, and

the UNIV dataset. Comparison with other state-of-the-art methods using various evaluation metrics demonstrates the advantages of our method, especially in handling group evolution (e.g., merge and split). Regarding computational cost, our proposed algorithm is an order of magnitude faster than the SGB model [4] on all datasets.

References

- [1] N. Ghosh and B. Bhanu. Evolving bayesian graph for three-dimensional vehicle model building from video. *IEEE Transactions on Intelligent Transportation Systems*, 15(2):563–578, April 2014.
- [2] D. Helbing and P. Molnár. Social force model for pedestrian dynamics. *Physical Review E*, 51:4282–4286, May 1995.
- [3] M. Moussaid, N. Perozo, S. Garnier, D. Helbing, and G. Theraulaz. The walking behaviour of pedestrian social groups and its impact on crowd dynamics. *PLoS ONE*, 5(4):e10047, 04 2010.
- [4] Z. Qin and C. Shelton. Improving multi-target tracking via social grouping. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1972–1978, June 2012.