Subgraph Decomposition for Multi-Target Tracking

Siyu Tang[1], Bjoern Andres[1] Mykhaylo Andriluka[1,2] Bernt Schiele[1]
[1]Max Planck Institute for Informatics, Saarbrücken, Germany [2]Stanford University, USA

Multi-target tracking can be formulated as an optimization problem with respect to a graph whose nodes correspond to detection hypotheses and whose edges connect detection hypotheses that hypothetically describe the same target. A commonly employed objective of the optimization is to select a subset of nodes and edges in such a graph to maximize similarity of connected detection hypotheses, while maintaining constraints that prevent splits and merges of tracks.

By far the most common approach is to choose the initial graph such that detection hypotheses are connected only across time (not within the same time frame) and to constrain the solution such that connected components of selected detection hypotheses are paths (that do not branch). With respect to a linear objective function, this problem is a Minimum Cost Disjoint Paths Problem with respect to the initial graph. It is used, explicitly or implicitly, in many modern tracking algorithms including [2, 3, 4].

While being intuitive, the Disjoint Paths formulation has a notable caveat: Typical target detectors yield, for each time frame, many similar (and typically equally plausible) detections of the same target. Within the Disjoint Paths formulation, it becomes necessary to choose, for each time frame and target, one best out of many similar (and plausible) hypotheses. Various recipes are proposed in the literature to address this challenge. E.g., [2] rely on a greedy iterative procedure that finds one track at a time and then removes corresponding hypotheses, or [4] performs several rounds of optimization that merge detections into tracklets and then into full tracks. Unfortunately, all these methods depend on parameters that need to be tuned carefully, as noted in [2, 4].

Embracing the possibility of having multiple plausible hypotheses per target and frame motivates us to formulate multi-target tracking as a Minimum Cost Subgraph Multicut Problem. The feasible solutions of this formulation are such that possibly multiple hypotheses per track and time frame are selected and clustered, resulting in an overall rigorous and elegant approach to link, cluster and track targets *jointly* across space and time. To illustrate the similarities and differences to prior work we implement a version of a tracking algorithm based on the Minimum Cost Disjoint Path Problem. Although conceptually simple, its output is already on par with the state of the art for public benchmark sequences, as we show in our paper.

This paper makes the following contributions: *First*, to our knowledge, our work is the first to propose a Subgraph Multicut model for the multi-target tracking problem jointly solving the spatial *and* temporal associations of detection hypotheses. *Second*, we provide an in-depth analysis and comparison of the Subgraph Multicut and the Disjoint Paths models. Our results suggest that the Subgraph Multicut model has considerable advantages due to the fact that state-of-the-art object detectors output multiple hypotheses per target. *Third*, besides proposing an exact solver, we also provide a heuristic solution based on the Kernighan-Lin algorithm [1], which makes the method applicable to large sequences. Finally we perform extensive experiments and present superior results compared to the state-of-the-art.

Subgraph Multicut Problem We formulate multi-target tracking as a Minimum Cost Subgraph Multicut Problem (Def. 1). The formulation is with respect to an *undirected* graph $G = (V,E)$ whose nodes V are all hypothesized detections of an entire video and whose edges E connect pairs of detection hypotheses that hypothetically describe the same target, including pairs in the same video frame.

The feasible solutions of the Minimum Cost Subgraph Multicut Problem (Def. 1) define subgraphs $G' = (V',E')$ of G which are encoded by $x \in \{0,1\}^V$, the characteristic function of the subset $V' = \{v \in V \mid x_v = 1\} \subseteq V$ of nodes, and $y \in \{0,1\}^E$, a characteristic function defining the subset $E' = \{vw \in E \mid y_{vw} = 1\} \subseteq E$ of edges. More specifically, the subgraph

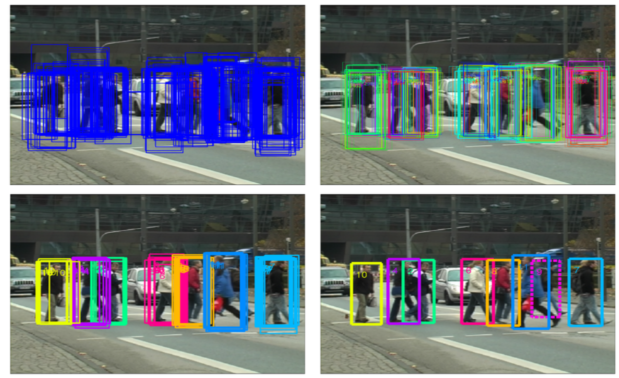

Figure 1: Overview of the Subgraph Multicut tracking method: (clockwise) detection hypotheses, overlapping tracklet hypotheses, hypotheses decomposition (clustering jointly across space and time) and final tracks (dotted rectangles are interpolated tracks).

G' is constrained (by Def. 1) such that each connected component (V'',E'') of G' contains all edges $E'' = \binom{V''}{2} \cap E$.

The objective function of the Minimum Cost Subgraph Multicut Problem is linear in the coefficients of x and y:

Definition 1 With respect to an undirected graph $G = (V,E)$, $c \in \mathbb{R}^V$ and $d \in \mathbb{R}^E$, the 01-linear program written below is called an instance of the *Minimum Cost Subgraph Multicut Problem*.

$$\min_{\substack{x\in\{0,1\}^V \\ y\in\{0,1\}^E}} \quad \sum_{v\in V} c_v x_v + \sum_{e\in E} d_e y_e \tag{1}$$

$$\text{subject to} \quad \forall e \in E \; \forall v \in e: \; y_e \leq x_v \tag{2}$$

$$\forall C \in \text{cycles}(G) \; \forall e \in C: \; (1-y_e) \leq \sum_{e' \in C \setminus \{e\}} (1-y_{e'}) \tag{3}$$

Here, the constraints (2) state that an edge can only be selected if both its nodes are selected. The cycle constraints (3) state, firstly, that every component of the selected subgraph G' is also a component of G and, secondly, that every edge of G whose nodes are in the same component of G' is also in G. In the context of multi-target tracking this implies that if a detection hypothesis is connected (spatially or temporally) to another detection hypothesis, all neighbors of the first hypothesis have to be connected to all spatial and temporal neighbors of the second hypothesis as well.

We solve instances of the Subgraph Multicut problem exactly by Integer Linear Programming (ILP), using Branch-and-Cut, as well as heuristically, by fixed points of the Kernighan Lin Algorithm.

Experiments We evaluate the performance of the proposed Subgraph Multicut model on three publicly available sequences: TUD-Campus, TUD-Crossing and ParkingLot. We perform extensive experiments and analysis on the TUD-Crossing sequence and present quantitative, superior results compared to other competitive methods on all three sequences.

[1] B. W. Kernighan and Shen Lin. An efficient heuristic procedure for partitioning graphs. *Bell Systems Technical Journal*, 49:291–307, 1970.

[2] Hamed Pirsiavash, Deva Ramanan, and Charless C. Fowlkes. Globally-optimal greedy algorithms for tracking a variable number of objects. In *CVPR 2011*.

[3] Aleksandr V. Segal and Ian Reid. Latent data association: Bayesian model selection for multi-target tracking. In *ICCV*, 2013.

[4] Li Zhang, Yuan Li, and R. Nevatia. Global data association for multi-object tracking using network flows. In *CVPR 2008*.

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.