## Efficient Spatio-Temporal Data Association Using Multidimensional Assignment for Multi-Camera Multi-Target Tracking

Moonsub Byeon msbyeon@snu.ac.kr Songhwai Oh songhwai@snu.ac.kr Kikyung Kim koreaton@snu.ac.kr Haan-Ju Yoo neohanju@snu.ac.kr Jin Young Choi jychoi@snu.ac.kr

This paper proposes a novel multi-target tracking method which solves a data association problem using images from multi-cameras. In multi-cameras, two combinatorial problems should be solved at the same time: spatial data association between cameras and temporal data association between frames. The *spatio-temporal data association* problem is a well known NP-hard problem even in a small number of cameras or frames (more than 3). Current existing methods [1, 2] simplify the spatio-temporal data association model (shortest path) and 3D location estimation. However, the complexity grows exponentially with the number of cameras.

In this work, the spatio-temporal data association problem is formulated as a multidimensional assignment problem (MDA). To achieve a fast, efficient, and easily implementable algorithm, we solve the MDA problem iteratively by solving a sequence of bipartite matching problems using random splitting and merging operations. Hence, the proposed algorithm can be considered as a guided random search to find the global optimum through repeated random local searches (bipartite matchings). In addition, we design a new cost function considering 3D reconstruction accuracy, motion smoothness, visibility from cameras, starting/ending at entrance and exit zone, and false positive. Our approach reconstructs 3D trajectories that represent people's movement as 3D cylinders whose locations are estimated considering all adjacent frames (See Figure 2).

The observations during 1, ..., T frames and at cameras 1, ...K form a KT-partite (hyper) graph  $G = (V, E) = (\mathbf{I}_{11} \cup ... \cup \mathbf{I}_{KT}, E)$ , where vertices V are partitioned into  $K \times T$  different independent sets  $\mathbf{I}_{11}, ..., \mathbf{I}_{KT}$ and each hyperedge in E contains at least one vertex in each partite set. Trajectory hypotheses set  $\mathbf{T}$  can be defined as a set of all hyperedges E. We represent each trajectory hypothesis  $\mathcal{T}_n \in \mathbf{T}$  as a matrix whose entry in the *k*-th row and *t*-th column corresponds to an observation index at the *t*-th frame of the *k*-the camera. The problem of finding a set of disjoint trajectory hypotheses with a minimum sum of costs can be formulated as the MDA problem which is equivalent to the problem of minimizing the sum of costs of hyperedges containing one element per partite set in the hypergraph G. With binary decision variables  $x_{\mathcal{T}_n} \in \{0, 1\}$  deciding whether the trajectory  $\mathcal{T}_n$  is in the association hypothesis  $\mathbf{H}$ , and cost function  $c: \mathbf{T} \to \mathbb{R}$ , the objective function and disjointness constraints are given by

$$\min \sum_{\mathcal{T}_n \in \mathbf{T}} c(\mathcal{T}_n) x_{\mathcal{T}_n} \quad s.t. \sum_{\mathcal{T}_n \in \mathbf{T}_{[kt],i}} x_{\mathcal{T}_n} = 1 \begin{cases} \mathbf{T}_{[kt],i} = \{\mathcal{T}_n \in \mathbf{T} | [\mathcal{T}_n]_{k,t} = i\} \\ k = 1, 2, ..., K \\ t = 1, 2, ..., K \\ i = 1, 2, ..., m_{kt} \end{cases}$$
(1)

where  $\mathbf{T}_{[kt],i}$  is a subset of all trajectory hypothesis set  $\mathbf{T}$  whose value in the *k*-th row and *t*-th column has the detection index *i* and cost  $c(\mathcal{T}_n)$ denotes the cost of the *n*-th trajectory  $\mathcal{T}_n$ .

Denoting an association hypothesis at *iter*-th iteration as  $\mathbf{H}^{iter}$ , we can get a new association hypothesis  $\mathbf{H}^{iter-1}$  by improving the previous association hypothesis  $\mathbf{H}^{iter-1}$ . The key idea is that we randomly split the previous association hypothesis  $\mathbf{H}^{iter-1}$  and optimally re-merging by solving the approximated matching problem. Our split/merge strategy is expected to find the solution of which quality is better than or equal to the previous one, because it splits the previous association hypothesis  $\mathbf{H}^{iter-1}$  while maintaining the disjointness of the trajectory and re-merges the split trajectories to form a new association hypothesis  $\mathbf{H}^{iter}$  with a less than or

Perception and Intelligence Laboratory, Department of Electrical and Computer Engineering, Seoul National University, Seoul, Korea



Figure 1: An example of 5 frames and 2 cameras. Our data association strategy is to split and re-merge the previous association hypothesis **H** iteratively. Trajectory splitting is done by the Random Split Mask determined randomly at each iteration and merging is done to search the optimal bipartite matching between association hypotheses splits  $\tilde{\mathbf{H}}^{I}$  and  $\tilde{\mathbf{H}}^{J}$ . The resulting optimal merged association hypotheses is  $\tilde{\mathbf{H}}^{*}$ .



Figure 2: Examples of tracking result by our method.

equal to the previous cost (See an example in Figure 1). Our cost function is a summation of five individual terms:

$$c(\mathcal{T}_n) = c_{\mathcal{T}_n} = c_{rec} + c_{mot} + c_{vis} + c_{tse} + c_{fpt}, \qquad (2)$$

 $c_{rec}$ : measures 3D reconstruction error between a 3D cylinder model and its observations from multi-cameras,

 $c_{mot}$ : evaluates how well trajectory describes real motion of a person,  $c_{vis}$ : designed for modeling whether a person is visible at each camera,  $c_{tse}$ : prevents a trivial solution which includes too many track fragments,  $c_{fpt}$ : prevents a trivial association hypothesis where a trajectory hypothesis is considered as a false positive.

The experiments illustrate the proposed method shows the state-ofthe-art performance in challenging multi-camera datasets and the computational efficiency with 8 times faster computation than the existing BIP approach. In particular, the proposed high-order motion model and 3D trajectory estimation with 3D cylinder model can reduce the possibility of ID switches.

- M. Hofmann, D. Wolf, and G. Rigoll. Hypergraphs for Joint Multiview Reconstruction and Multi-object Tracking. In *CVPR*, 2013.
- [2] L. Leal-Taixe, G. Pons-Moll, and B. Rosenhahn. Branch-and-price global optimization for multi-view multi-target tracking. In *CVPR*, 2012.