EG23

SIC Saarland Informatics Campus

Non-Separable Multi-Dimensional Network Flows for Visual Computing

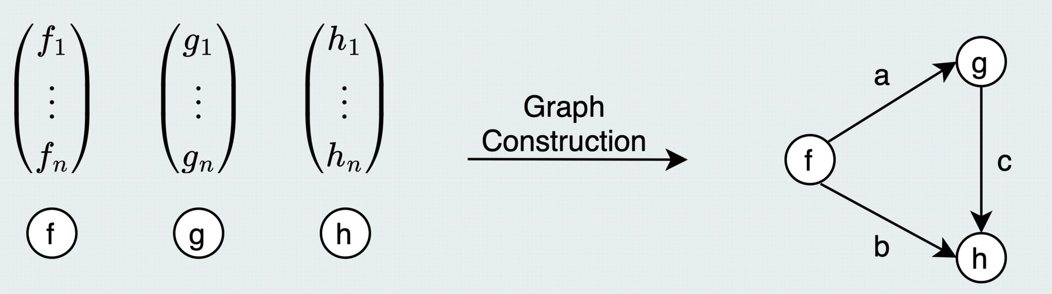
V. Ehm[1,2] D. Cremers[1,2] F. Bernard[3]

[1] TU Munich [2] Munich Center for Machine Learning [3] University of Bonn

MOTIVATION

- When creating graphs with **scalar-valued edges information is lost** and thereby the expressiveness limited.
- For example: **high-dimensional data** (*e.g.* feature descriptors) is **mapped to a single scalar value** (*e.g.* the similarity between two feature descriptors).

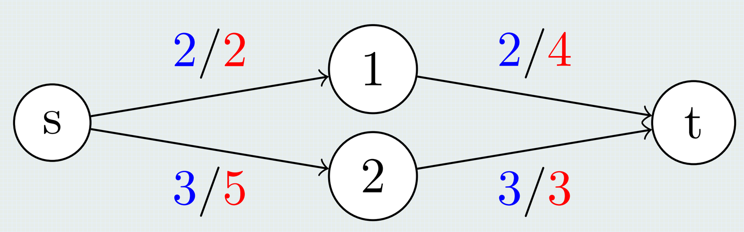
→ Information about individual feature dimensions is not explicitly available after the mapping.



BACKGROUND

Flows in Graphs

- Harris *et al.* [1] define maximum flow.



- Garg *et al.* [2] extend this formulation to multi-commodity flows.
- Li *et al.* [3] use a binary variable to ensure that only one path per commodity can be used to reach the sink.

→ We introduce a non-separable maximum multi-commodity flow formulation, i.e. only one path for all commodities.

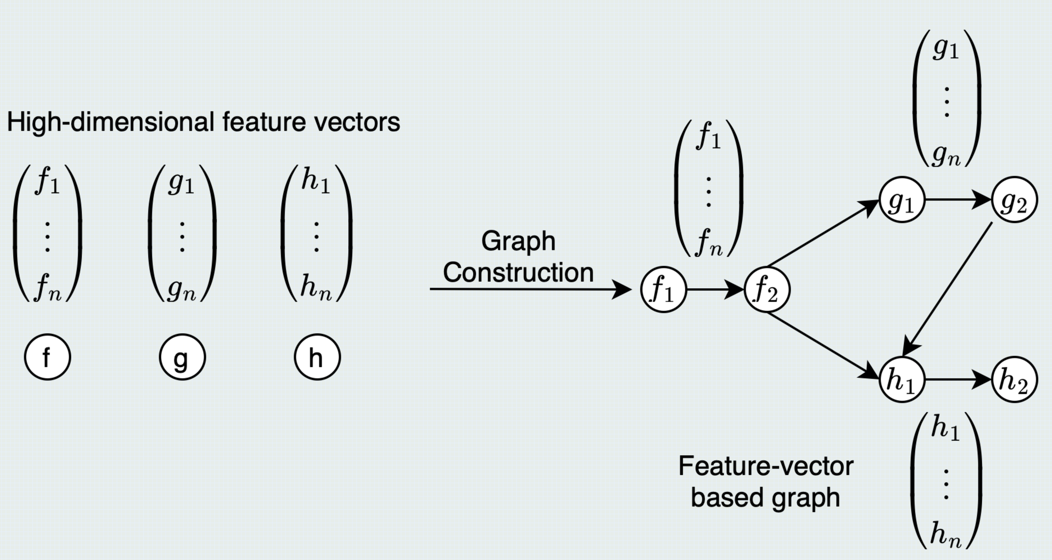
Network Flows in Multi-Object Tracking (MOT)

- Traditional MOT approaches [4,5] use minimum cost flow algorithms through scalar graphs.

→ Instead of minimum cost, we use maximum multi-commodity flow formulation.

OVERVIEW

- We present a **non-separable multi-commodity flow formulation**, where a flow unit with all its commodities can not be separated throughout the graph.
- As proof of concept, we apply the algorithm in the context of **multi-object tracking.**
- Our approach can increase the **robustness to noise** in the multi-object tracking setting.



METHODOLOGY

Non-Separable Multi-Dimensional Network Flow

We define our approach as a mixed-integer program, with a binary variable and two special constraints:

- **Node count:** A single incoming and a single outgoing flow vector may have non-zero flow. → Flow vectors can not be separated through nodes.
- **Total count:** We fix the number of flow entities leaving the source node and entering the target node to a specific value.

$$\begin{aligned}
\underset{\substack{f_{uv}\in\mathbb{R}^k\\ b_{uv}\in\{0,1\}}}{\text{maximize}} \quad & \sum_{v:s\to v} f_{sv}^T\mathbf{1} & \\
\text{subject to} \quad & f_{uv} \le b_{uv}c_{uv}, & \forall (u,v)\in E \text{ (capacity)},\\
& \sum_{u:u\to v} f_{uv} = \sum_{w:v\to w} f_{vw}, & \forall v\neq s,t \text{ (flow cons.)},\\
& \sum_{u:u\to v} b_{uv} = \sum_{w:v\to w} b_{vw} = 1, & \forall v\neq s,t \text{ (node count)},\\
& \sum_{u:s\to u} b_{su} = \sum_{v:v\to t} b_{vt} = d, & \forall u,v\neq s,t \text{ (total count)},\\
& f_{uv} \ge \mathbf{0} & \forall (u,v)\in E \text{ (non-neg.)}.
\end{aligned}$$

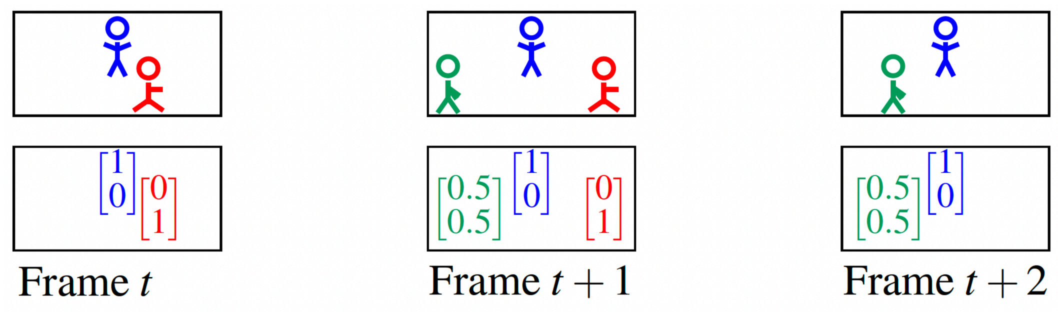
Variables

- V: nodes with a source node $s \in V$ and a sink node $t \in V$
- $E \subseteq V \times V$: set of (directed) edges
- $c_{uv} \in \mathbb{R}_+^k$: capacity vector with k commodities
- $f_{uv} \in \mathbb{R}^k$: multi-dimensional flow
- $b_{uv} \in \{0,1\}$: decision variable (whether an edge is active)
- d: number of flow entities

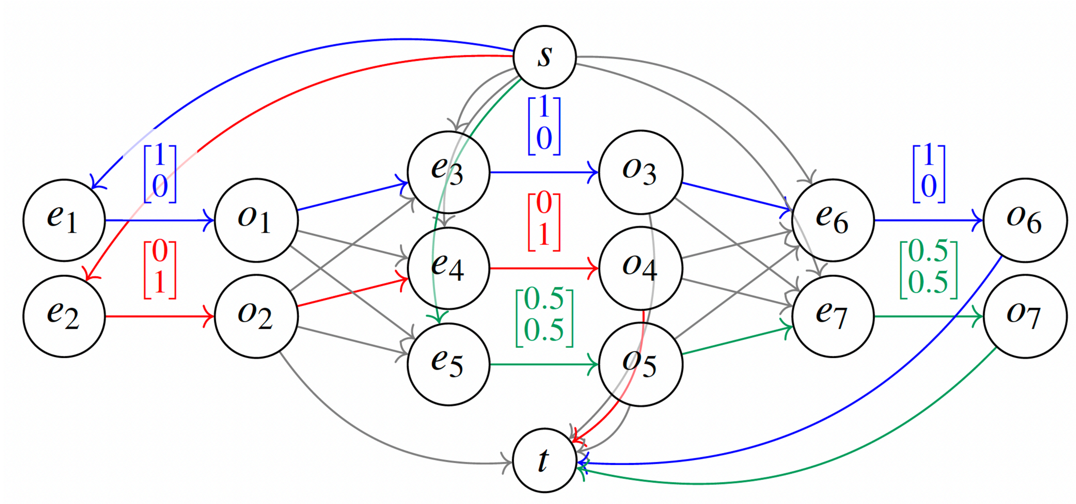
Application to Multi-Object Tracking

We define three different types of edges: observation edges (between objects), transition edges (between frames), and enter/exit edges (source/sink connection).

The objects of the three sample frames are represented by feature vectors.



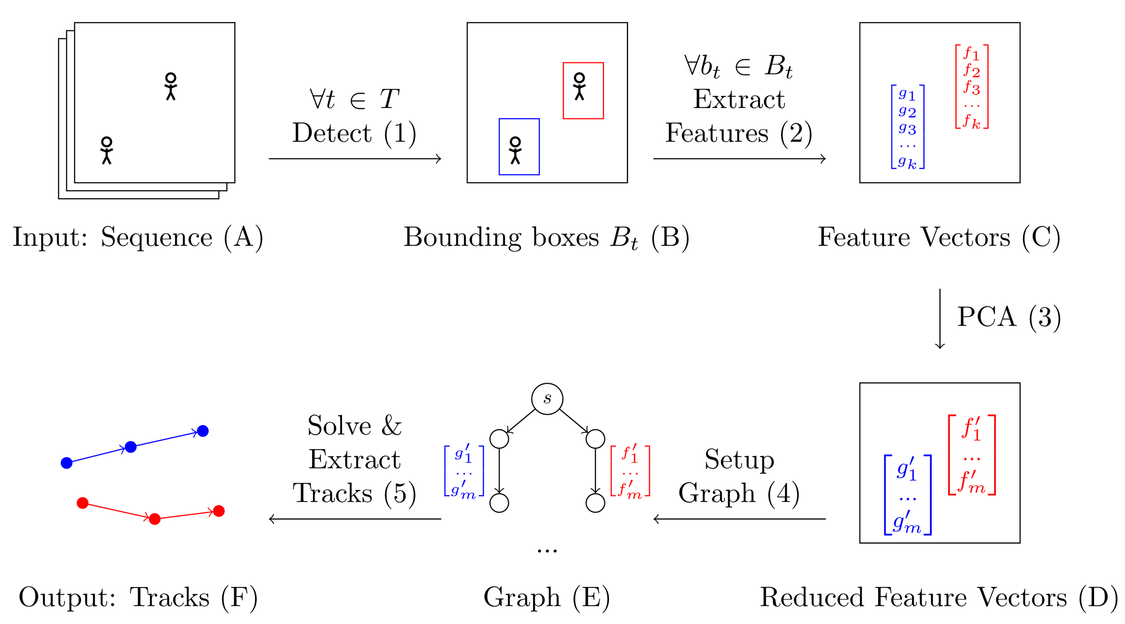
These vectors are set as capacity on the corresponding object edges. All other edges have infinite capacity.



RESULTS

Experimental Setup

Features from bounding boxes are reduced and set as edge capacity in graph to extract object tracks.



Dataset

We choose training sequences (2,4,5,9,10 and 11) of the MOT16 benchmark[6] and provide the ground truth boxes and the ground truth number of individual objects.

Feature Descriptors

We use two different feature descriptors: color histograms (RGB) and deep features (ResNet18 architecture [7]).

Evaluation Metric

As a metric we normalize the identity switches (IDSW) by the total number of ground truth boxes (GT):

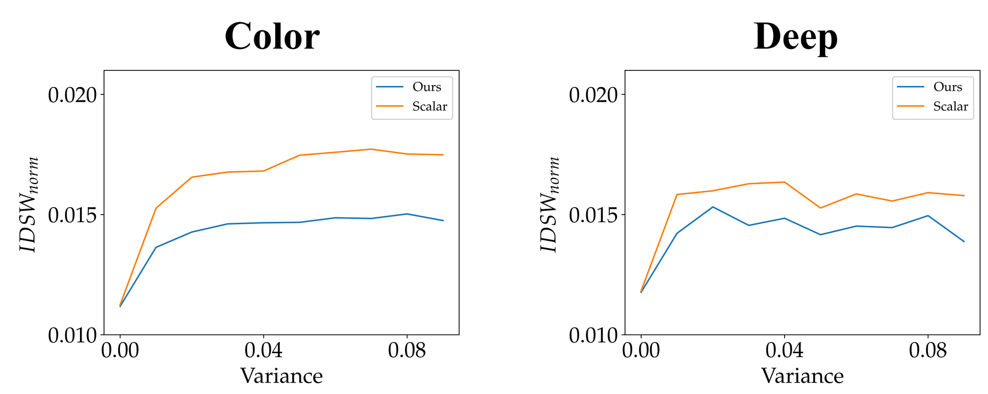
$$IDSW_{norm} = \frac{\sum_t IDSW_t}{\sum_t GT_t}$$

Robustness Evaluation

To evaluate the robustness, we add random Gaussian noise with different variances to every image.

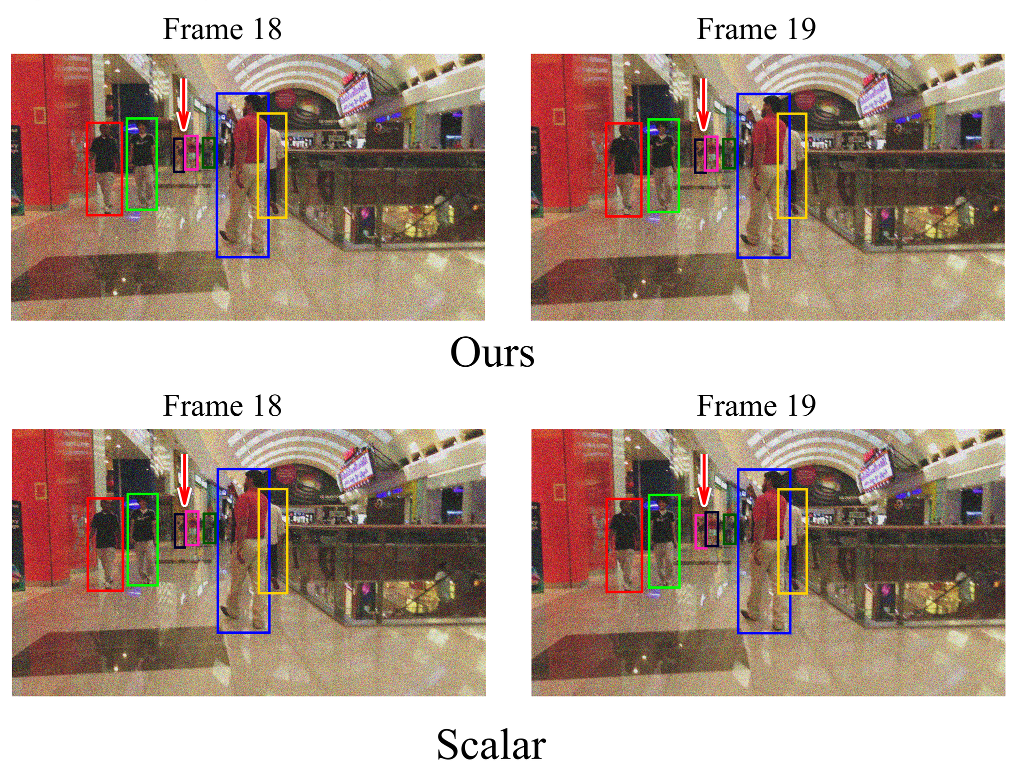
Quantitative Results

Our method performs better on noisy images than the scalar method with different feature descriptors.



Qualitative Example (Seq 11)

The scalar method has an identity switch, our method has no identity switch (red arrow).



UNIVERSITÄT BONN

Munich Center for Machine Learning

EG23

REFERENCES

[1] HARRIS T., ROSS F.: Fundamentals of a method for evaluating rail net capacities. Tech. rep., RAND CORP SANTA MONICA CA,1955.
[2] GARG N., VAZIRANI V. V., YANNAKAKIS M.: Approximate max-flow min-(multi) cut theorems and their applications. SIAM Journal on Computing 25 (1996).
[3] LI X., ANEJA Y. P., BAKI F.: An ant colony optimization metaheuristic for single-path multicommodity network flow problems. Journal of the Operational Research Society 61 (2010).
[4] LEAL-TAIXE L., PONS-MOLL G., ROSENHAHN B.: Everybody needs somebody: Modeling social and grouping behavior on a linear programming multiple people tracker. In ICCV workshops (2011).
[5] ZHANG L., LI Y., NEVATIA R.: Global data association for multi-object tracking using network flows. In CVPR (2008).
[6] MILAN A., LEAL-TAIX. L., REID I., ROTH S., SCHINDLER K.: Mot16: A benchmark for multi-object tracking. arXiv preprint arXiv:1603.00831 (2016).
[7] HE K., ZHANG X., REN S., SUN J.: Deep residual learning for image recognition. In CVPR (2016).