Branch-and-price global optimization for multi-view multi-object tracking

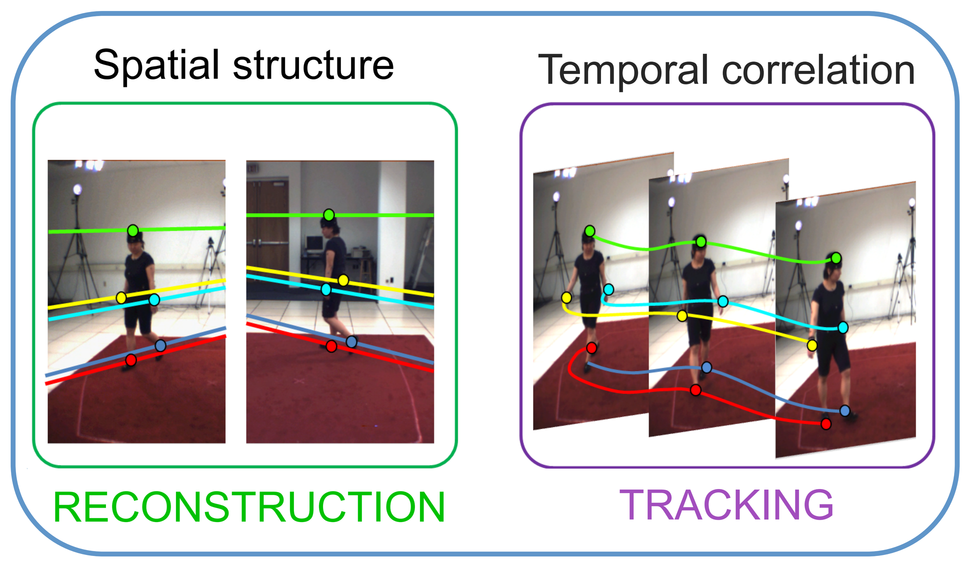
Laura Leal-Taixé Gerard Pons-Moll Bodo Rosenhahn

Institute for Information Processing (TNT), Leibniz University Hannover, Germany

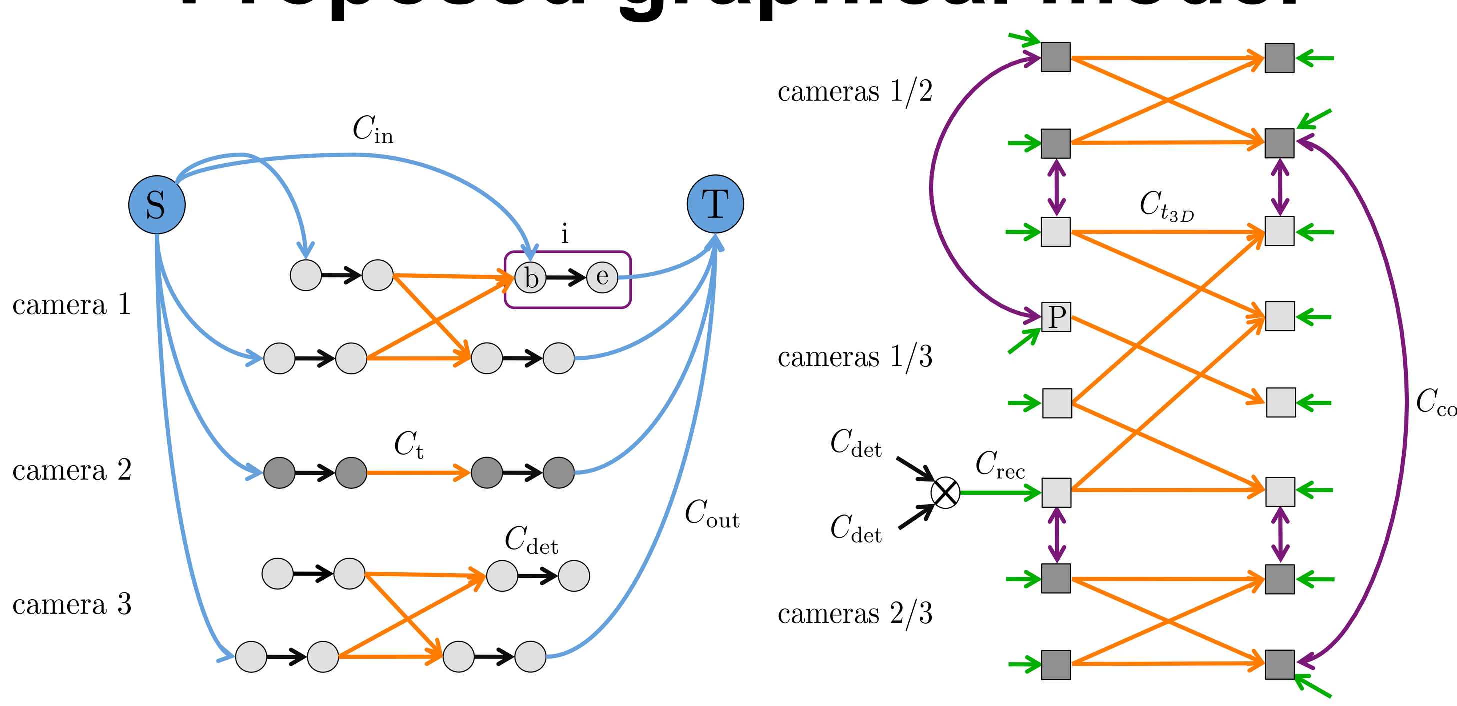
{leal,pons,rosenhahn}@tnt.uni-hannover.de

Goal

A **global optimum solution** to track multiple objects in multiple views



Proposed graphical model



2D layer 3D layer

Entrance/exit edges: determine when trajectory starts/ends

Detection edges: confident detections are likely to be in the path of the flow, and therefore, part of the trajectory.

$$C_{\det}(i_v) = \log\left(1 - P_{\det}(\mathbf{p}_{i_v})\right)$$

Temporal 2D edges: encode temporal dynamics of targets.

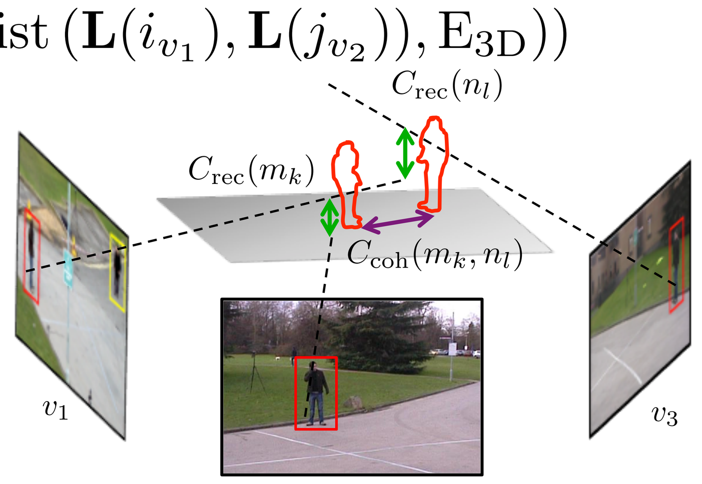
$$C_t(i_v, j_v) = -\log\left(\mathcal{F}\left(\frac{\|\mathbf{p}_{j_v} - \mathbf{p}_{i_v}\|}{\Delta t}, V_{\max}^{2D}\right) + B_f^{\Delta f - 1}\right)$$

Reconstruction edges:

★ $C_{\text{rec}}(m_k) = \log\left(1 - \mathcal{F}\left(\text{dist}\left(\mathbf{L}(i_{v_1}), \mathbf{L}(j_{v_2})\right), \mathrm{E}_{3D}\right)\right)$

Camera coherency edges:

★ $C_{\text{coh}}(m_k, n_l) = \log\left(1 - \mathcal{F}\left(\|\mathbf{P}_{m_k}, \mathbf{P}_{n_l}\|, \mathrm{E}_{3D}\right)\right)$



Temporal 3D edges: encode 3D temporal dynamics.

★ $C_{t_{3D}}(m_k, n_k) = \log\left(1 - \mathcal{F}\left(\frac{\|\mathbf{P}_{m_k} - \mathbf{P}_{n_k}\|}{\Delta t}, V_{\max}^{3D}\right)\right)$

★ Cascade of prizes: having the same identity in 2D is beneficial if the 3D information matches.

Multi-commodity flow LP formulation

Formulate MAP problem as a **Linear Program** using flow flags $f(i) = \{0, 1\}$.

Objective function: $\mathcal{T}* = \underset{\mathcal{T}}{\text{argmin}}\ \mathbf{C}^{\mathrm{T}}\mathbf{f} = \sum_i C(i) f(i)$

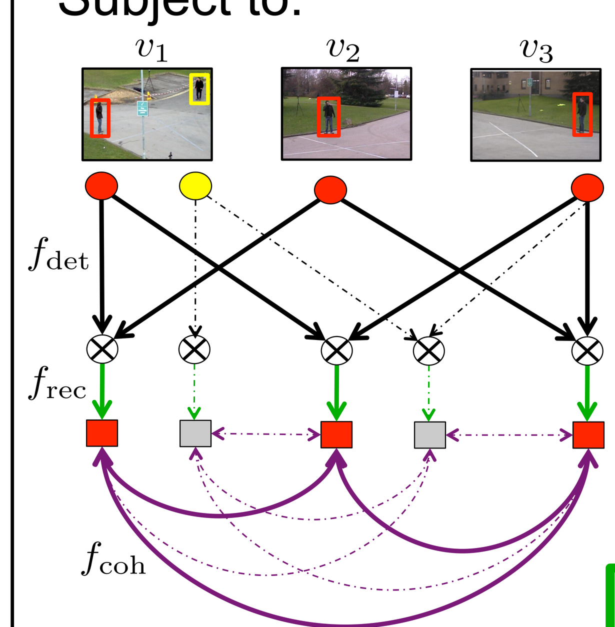
Subject to:

$$f_{\det}(i_v) = f_{\text{in}}(i_v) + \sum_{j_v} f_t(j_v, i_v)$$
$$f_{\det}(i_v) = \sum_{j_v} f_t(i_v, j_v) + f_{\text{out}}(i_v)$$

Flow conservation at the nodes

$$f_{\text{rec}}(m_k) = f_{\det}(i_{v_1}) f_{\det}(j_{v_2})$$
$$f_{\text{coh}}(m_k, n_l) = f_{\text{rec}}(m_k) f_{\text{rec}}(n_l)$$
$$f_{t_{3D}}(m_k, n_k) = f_{\text{rec}}(m_k) f_{\text{rec}}(n_k)$$

Activation constraints of the form $f_{ab} = f_a f_b$ cannot be used in a Linear Program!



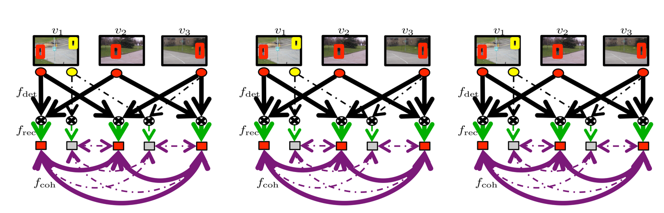
Cascade of prizes

$$f_{ab} - f_a \leq 0 \quad f_{ab} - f_b \leq 0 \quad f_a + f_b - f_{ab} \leq 1 \checkmark$$

But we want to activate the prizes **only if** the two 2D nodes are activated **by the same object.**

$$0 \leq \sum_{i_v} f_{in}(i_v) \leq 1$$
$$0 \leq \sum_{i_v} f_{out}(i_v) \leq 1 \quad \forall v$$

How to deal with multiple objects? Use a **multi-commodity flow formulation.**



Multiple copies of the graph

$$\sum_n f^n(i) \leq 1$$
$$n = 1 \ldots N_{\text{obj}}$$

Binding constraints

Much more complex LP, cannot be solved with standard techniques!!

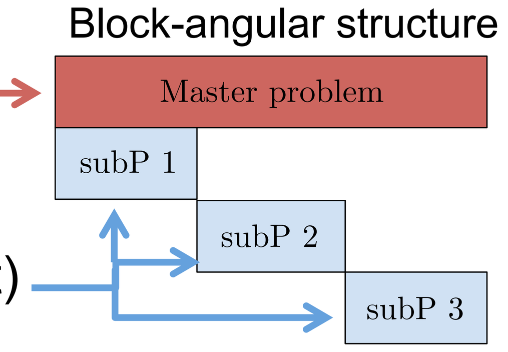
Dantzig-Wolfe decomposition

Objective function: $\min_{\mathbf{f}} \mathbf{C}^{\mathrm{T}}\mathbf{f} = \sum_{n=1}^{N_{\text{obj}}} (\mathbf{c}^n)^{\mathrm{T}}\mathbf{f}^n$

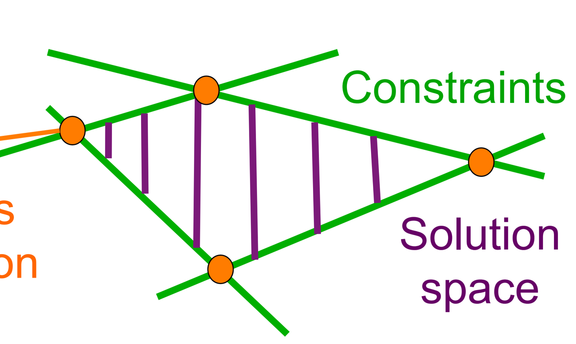
Subject to: $\mathbf{A}_1\mathbf{f} \preceq \mathbf{b}_1$ Hard constraints (binding)

$\mathbf{A}_2^n\mathbf{f}^n \preceq \mathbf{b}_2^n$ Easy constraints (for each object)

Block-angular structure



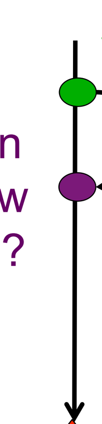
Convert the problem to a Master Problem and N_{obj} subproblems, using the representation theorem $\mathbf{f}^n = \sum_{j=1}^{J} \lambda_j^n \mathbf{x}_j^n$



Column generation

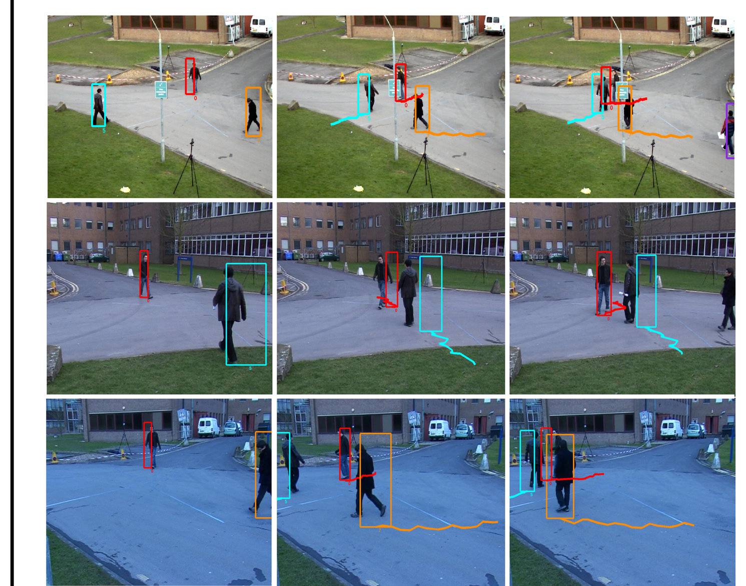
1. Select a subset of columns to form the *restricted master problem*, solve it with chosen method (e.g. Simplex, KSP).
2. Calculate the optimal dual solution μ
3. Price the rest of the columns $\mu(\mathbf{A}_1^n\mathbf{f}^n - \mathbf{b}_1^n)$
4. Find the columns with negative cost and add them to the restricted master problem. This is done by solving column generation subproblems.

$$\min_{\mathbf{f}} \quad (\mathbf{c}^n)^{\mathrm{T}}\mathbf{f}^n + \mu(\mathbf{A}_1^n\mathbf{f}^n - \mathbf{b}_1^n) \quad \text{s.t.} \quad \mathbf{A}_2^n\mathbf{f}^n \preceq \mathbf{b}_2^n$$



Results

Multiple people tracking: PETS 2009 dataset



CLEAR metrics, proposed method outperforms state-of-the-art with only 2 views.

	DA	TA	DP	TP	miss
Zhang et al. [24] (1)	68.9	65.8	60.6	60.0	28.1
GTR(2)	51.9	49.4	56.1	54.4	31.6
GRT (2)	64.6	57.9	57.8	56.8	26.8
TR (2)	66.7	62.7	59.5	57.9	24.0
RT (2)	69.7	65.7	61.2	60.2	25.1
Berclaz et al. [4] (5)	76	75	62	**62**	–
Proposed (2)	**78.0**	**76**	**62.6**	60	**16.5**
TR (3)	48.5	46.5	51.1	50.3	20
RT (3)	56.6	51.3	54.5	52.8	23.5
Proposed (3)	73.1	71.4	55.0	53.4	**12.9**

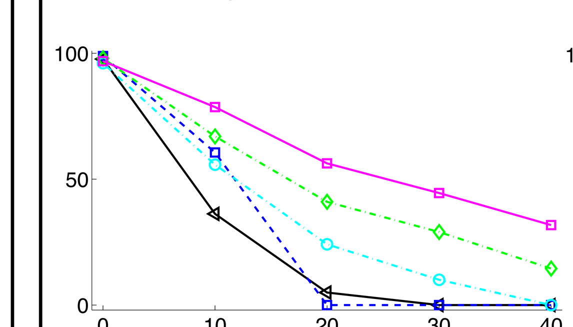
Even with calibration noise, our algorithm is able to track the red pedestrian which is occluded in 2 of the 3 views.

Much better performance than Reconstruction-Tracking or Tracking-Reconstruction.

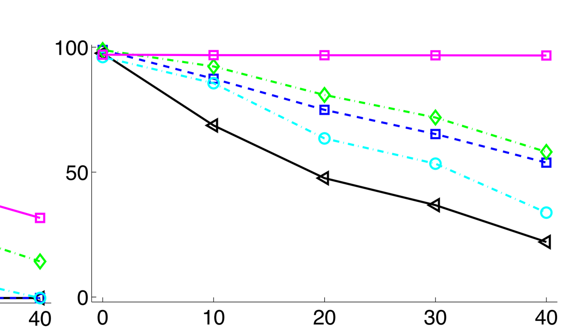
3D human pose tracking: HumanEva dataset

Ground truth 2D joint positions with 40% simulated outliers, much more robust performance than comparing algorithms.

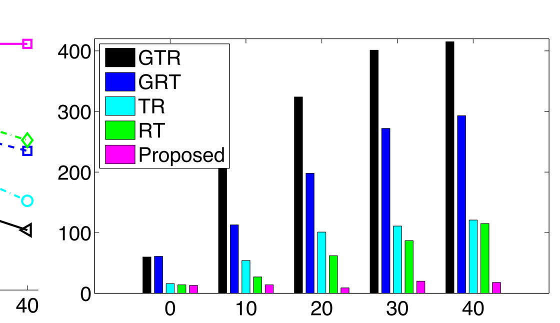
Tracking-Reconstruction Reconstruction-Tracking Proposed method



Tracking accuracy (%), simulated missing data



Tracking accuracy (%) with simulated outliers



ID switches with simulated outliers

Conclusions

- Jointly track multiple targets in multiple views.
- Proposed graph structure solves the problem as a global optimization including both temporal correlation and spatial information enforced by the configuration of the cameras.
- Branch-and-price: powerful tool to find the solution exploiting the special block-angular structure of the problem.
- **Code available!** `http://www.tnt.uni-hannover.de/~leal/`

This work was partially funded by the German Research Foundation, DFG project RO 2497/7-2.