Robust Real-Time Tracking of Multiple Objects by Volumetric Mass Densities

Horst Possegger, Sabine Sternig, Thomas Mauthner, Peter M. Roth, and Horst Bischof

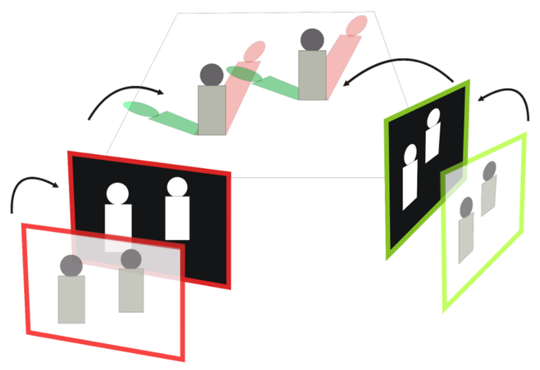
{possegger,sternig,mauthner,pmroth,bischof}@icg.tugraz.at

Institute for Computer Graphics and Vision, Austria

Graz University of Technology

Motivation

- Multi-object tracking from overlapping views
- Standard approaches use planar projections [2, 4]:
  - Common ground-plane assumption
  - Cannot handle out-of-plane motion
  - Projection artifacts



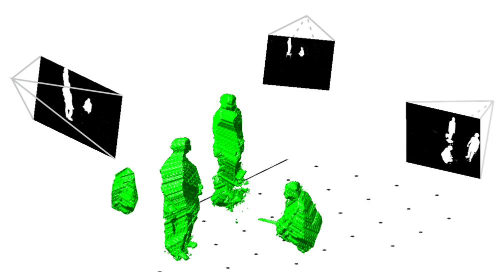
- Volumetric reconstructions [3, 6] overcome these limitations

- **Goals**:
  - Real-time capable multi-object tracking based on local mass densities of visual hull reconstructions
  - Exploit 3D knowledge to obtain a geometric tracking cue and robustly extract features to train discriminative classifiers on-line
- Evaluations demonstrate significant improvements compared to state-of-the-art
- Annotated datasets and evaluation protocol publicly available (scan QR-code)

Volumetric Mass Densities

- Visual hull representation:

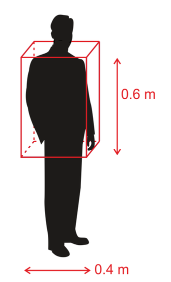
$$v_i \in \mathcal{V}_{\text{VH}} = \begin{cases} 1 & \text{if } v_i \text{ foreground} \\ 0 & \text{otherwise} \end{cases}$$



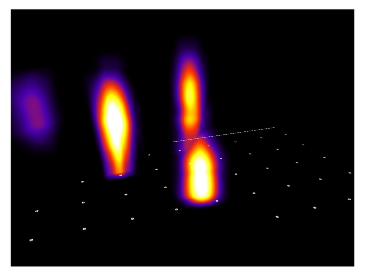
- Compute local mass density values to obtain the occupancy volume:

$$\mathcal{V}_{\text{O}} = \{ m_D(v_i) \,|\, \forall v_i \in \mathcal{V}_{\text{VH}} \},$$

$$m_D(v_i) = \frac{\sum_{v_j \in N_{v_i}} v_j}{|N_{v_i}|}$$



- Local neighborhood approximates the human torso by an axis-aligned cuboid:

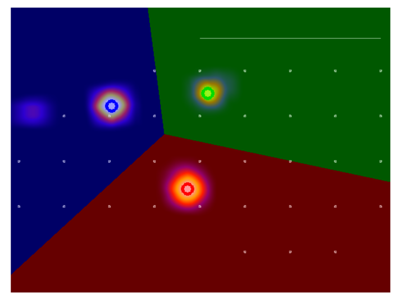


$$N_{v_i} = \Big\{ v_j \,\Big|\, |v_{j,x} - v_{i,x}| \le r \wedge |v_{j,y} - v_{i,y}| \le r \wedge |v_{j,z} - v_{i,z}| \le h/2 \Big\}$$

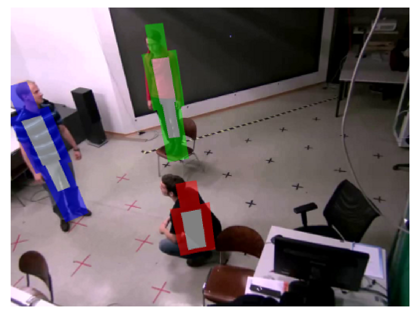
(2+1)D Tracking

- To overcome a computationally expensive 3D search, the tracking part is split into two separate steps:
- Estimate Cartesian coordinates:
  - Particle filtering on top-view occupancy map $\mathcal{M}$:

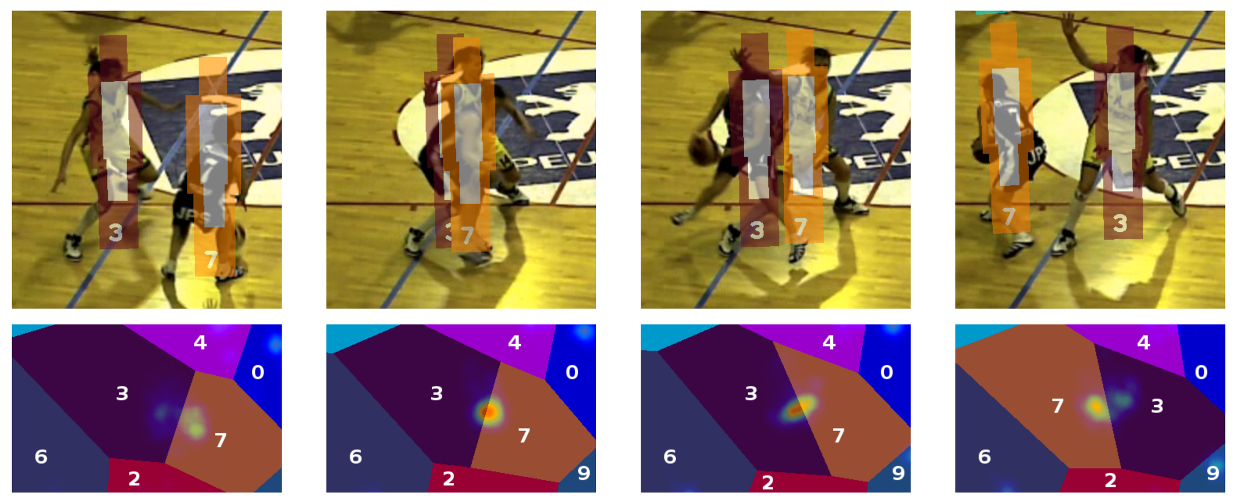
$$\mathcal{M}_{x,y} = \max m_D(v_i),$$
$$v_{i,x} = x,\ v_{i,y} = y,$$
$$v_{i,z} \in [z_{\text{min}}, z_{\text{max}}]$$



  - Second-order auto-regressive transition model and exponentially distributed likelihood model
  - Voronoi partitioning [5] allows for efficiently restricting the particle transition
- Given the (x, y) positions, obtain full 3D location by searching for the vertical mass center



Resolving Geometric Ambiguities



- Collect valuable samples $\mathbf{f}$ (HS histograms) at torso regions by exploiting 3D knowledge
- Ensure up-to-date appearance samples by FIFO updated feature bags:

$$\mathcal{F}_i = \left\{ \left\{ \mathbf{f}_l^{(c)} \right\}_{l=1}^{N_F} \right\}_{c=1}^{N_C}$$

- Cluster ambiguous hypotheses into pools $\mathcal{P}_m$ and train one-vs-all logistic regression classifiers on-demand:

$$\min_{\mathbf{w}_i} \frac{1}{2} \mathbf{w}_i^\top \mathbf{w}_i + C_i \sum_l \log \left(1 + e^{-y_{i,l} \mathbf{w}_i^\top \mathbf{f}_l} \right)$$

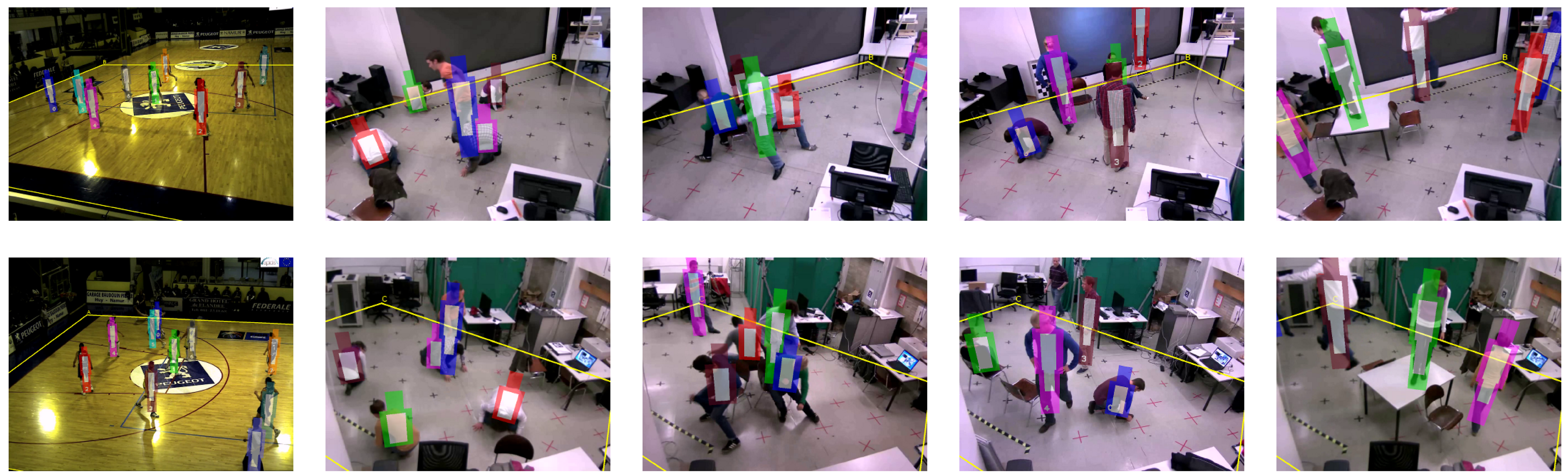
- Extract $\mathbf{f}_{\text{NMS}}$ at local maxima of conflict regions and re-assign trackers:

$$\hat{i} = \underset{i \in \mathcal{P}_m}{\text{arg max}}\ p_i \left(y_{i,\text{NMS}} = +1 \,|\, \mathbf{f}_{\text{NMS}}, \mathbf{w}_i \right)$$

Evaluation and Results

- Comparison to the state-of-the-art KSP tracker [1] on top of POM detections
- Standard *CLEAR MOT* metrics with assignment cut-off $\tau_d = 0.5$ [m]

Dataset	Frames	Algorithm	MOTP [m]	MOTA	TP	FP	FN	IDS	FPS
APIDIS	1500	Proposed	0.205	0.675	656	88	172	9	4.42
		Prop. w/o Color	0.211	0.597	625	121	202	10	6.16
		KSP/POM	0.231	0.490	607	156	220	46	80.70, 0.03
Chap	3760	Proposed	0.102	0.994	1555	2	6	1	9.89
		Prop. w/o Color	0.102	0.719	1316	193	241	4	12.67
		KSP/POM	0.167	0.952	1607	50	21	7	43.49, 0.02
Leaf 1	1800	Proposed	0.107	0.991	464	2	2	0	9.88
		Prop. w/o Color	0.107	0.721	436	83	44	7	10.34
		KSP/POM	0.169	0.976	495	6	1	5	63.84, 0.04
Leaf 2	2400	Proposed	0.097	0.916	930	41	41	0	7.65
		Prop. w/o Color	0.116	0.727	856	115	117	34	9.04
		KSP/POM	0.175	0.819	913	87	66	24	229.77, 0.05
Much	2400	Proposed	0.111	0.977	783	9	9	0	12.08
		Prop. w/o Color	0.116	0.736	694	99	99	11	13.21
		KSP/POM	0.218	0.754	770	139	32	26	185.28, 0.06
Pose	1820	Proposed	0.123	0.944	485	14	14	0	10.27
		Prop. w/o Color	0.128	0.822	456	42	44	3	12.99
		KSP/POM	0.201	0.555	427	156	31	17	132.49, 0.05
Table	1760	Proposed	0.112	0.898	599	30	28	6	8.03
		Prop. w/o Color	0.120	0.818	577	56	51	7	9.60
		KSP/POM	0.210	0.719	573	105	58	14	208.51, 0.07



APIDIS | Leaf 2 | Much | Pose | Table

References and Acknowledgments

[1] J. Berclaz, F. Fleuret, E. Türetken, and P. Fua. Multiple Object Tracking using K-Shortest Paths Optimization. *PAMI*, 2011.
[2] R. Eshel and Y. Moses. Tracking in a Dense Crowd Using Multiple Cameras. *IJCV*, 2010.
[3] L. Guan, J.-S. Franco, and M. Pollefeys. Multi-Object Shape Estimation and Tracking from Silhouette Cues. In *Proc. CVPR*, 2008.
[4] S. M. Khan and M. Shah. Tracking Multiple Occluding People by Localizing on Multiple Scene Planes. *PAMI*, 2009.
[5] M. Kristan, J. Perš, M. Perše, and S. Kovačič. Closed-world tracking of multiple interacting targets for indoor-sports applications. *CVIU*, 2009.
[6] M. Liem and D. M. Gavrila. Multi-person tracking with overlapping cameras in complex, dynamic environments. In *Proc. BMVC*, 2009.

This work was supported by the Austrian Science Foundation (FWF) under the projects Advanced Learning for Tracking and Detection in Medical Workflow Analysis (I535-N23) and MASA (P22299).