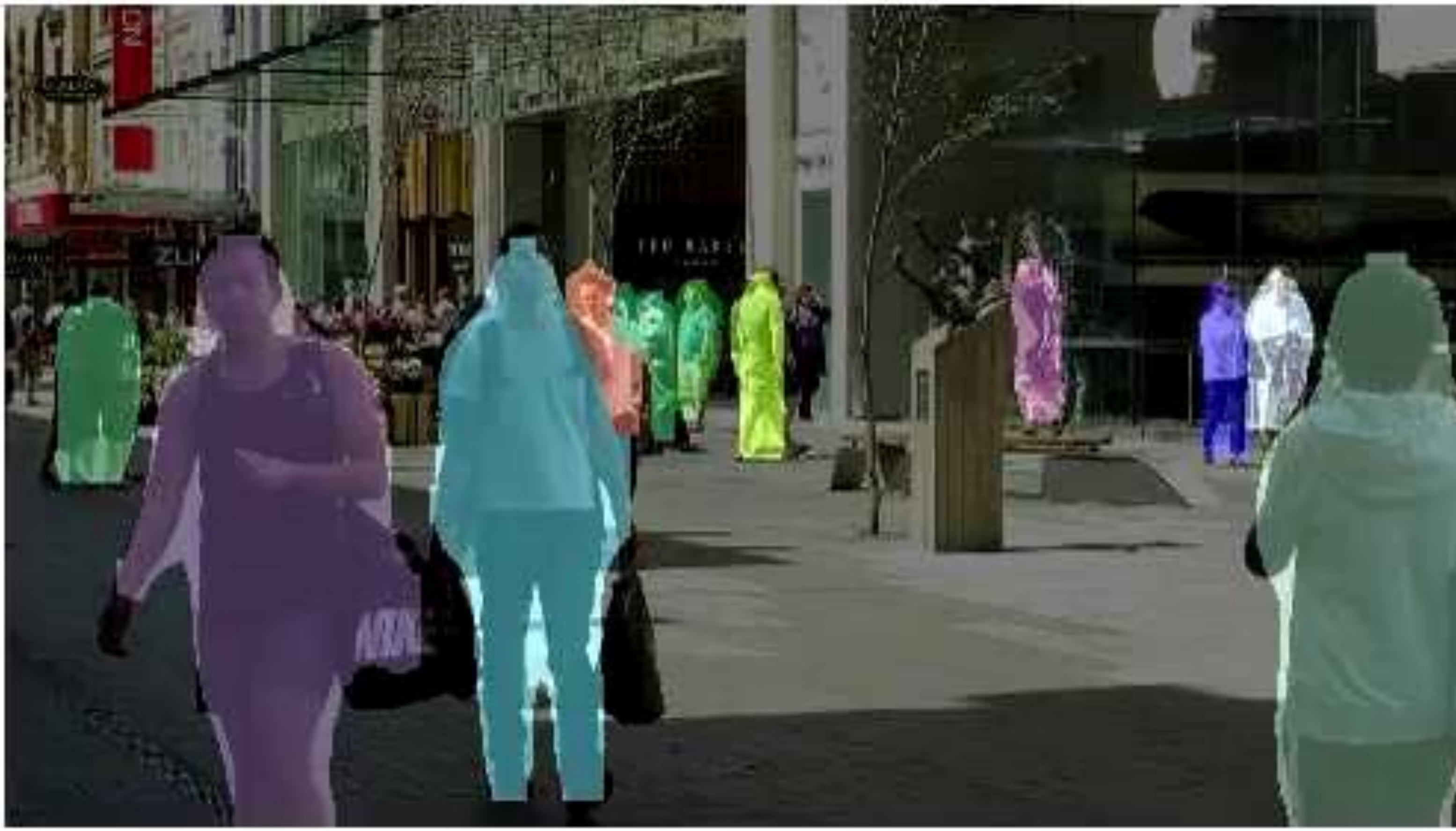
Instance Flow Based Online Multiple Object Tracking

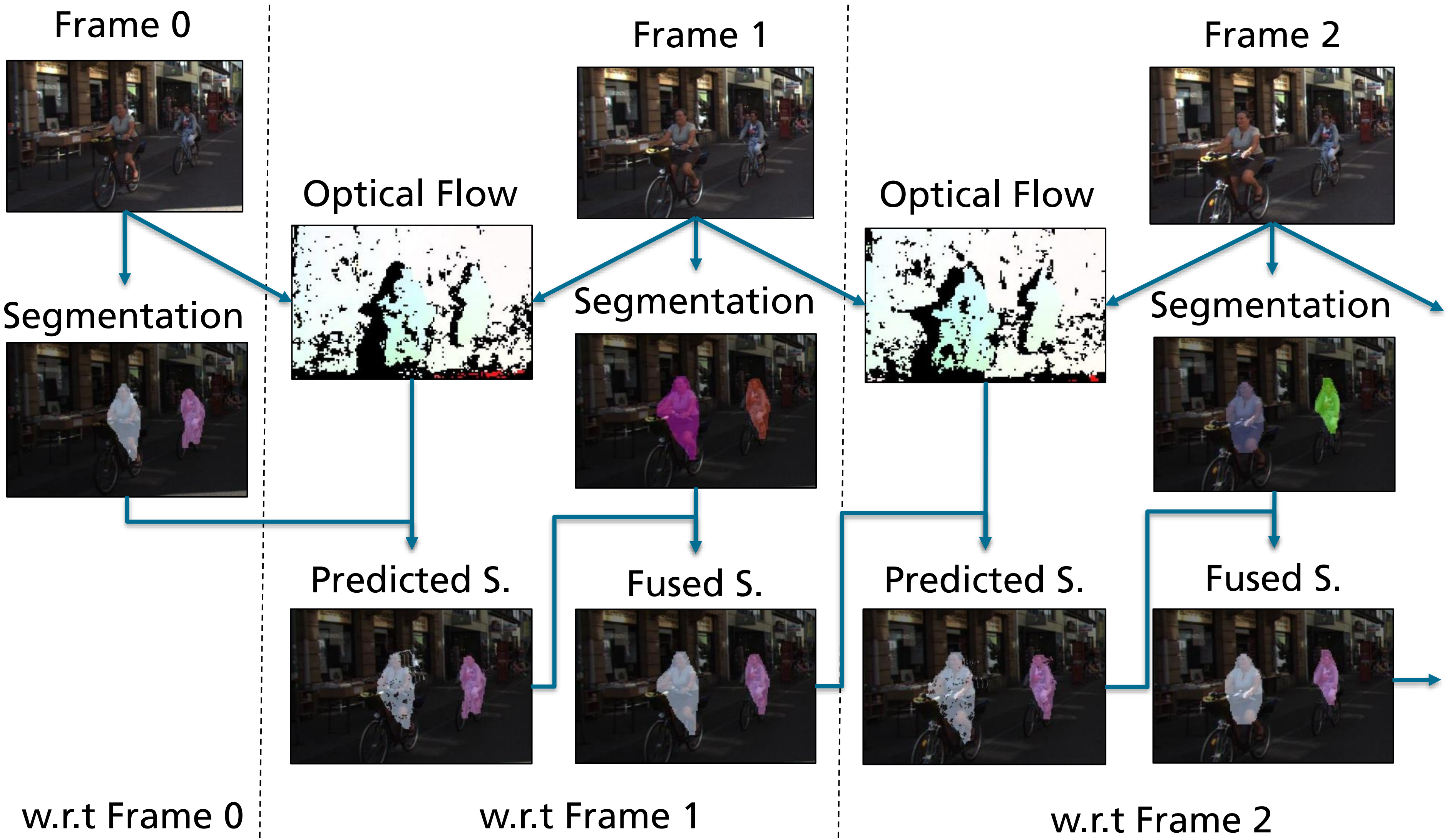
sebastian.bullinger@iosb.fraunhofer.de | 19.09.2017

Fraunhofer IOSB

Instance Flow Based Online Multiple Object Tracking



Approach


Frame 0
Frame 1
Frame 2
Optical Flow
Optical Flow
Segmentation
Segmentation
Segmentation
Predicted S.
Fused S.
Predicted S.
Fused S.
w.r.t Frame 0
w.r.t Frame 1
w.r.t Frame 2

Approach

I. Computation of instance-aware semantic segmentations
II. Computation of optical flow
III. Prediction of instance-aware semantic segmentations using optical flow
IV. Computation of an affinity matrix of predictions and instance segmentations
V. Computation of segmentation associations with Kuhn-Munkres algorithm
VI. Update tracker state using segmentation associations

III. Prediction of Object Segmentation Instances

- Semantic Segmentation S_t of Image I_t
  - $S_t(x,y) = (c,i)$
    - c = category id, i = object instance id
- Occupied Pixels of Instance i in frame t
  - $S_{t,i} = \{(x,y) \mid (x,y) \in \{1,\dots,w\} \times \{1,\dots,h\},\ S_t(x,y) = (c,i)\}$

$S_{t,1}$ $S_{t,2}$

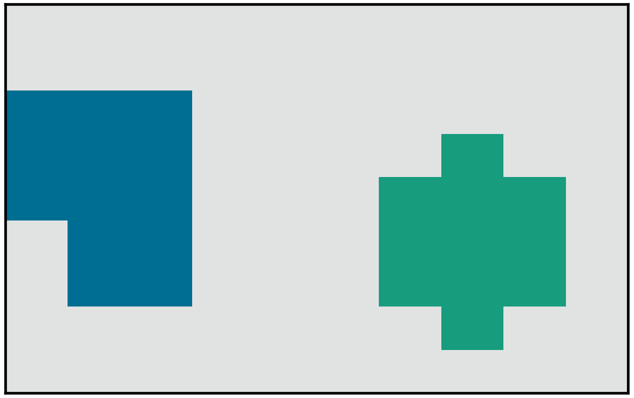
Fraunhofer IOSB

III. Prediction of Object Segmentation Instances

- Valid optical flow positions of instance i
  - $F_{t,i}^{(v)} = S_{t,i} \cap F_t^{(v)}$

III. Prediction of Object Segmentation Instances

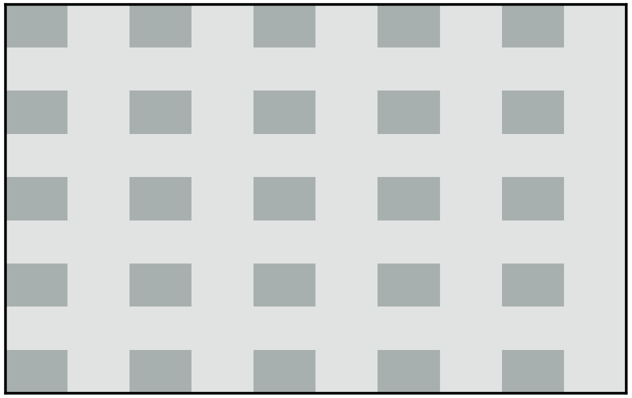
- Valid optical flow positions of instance i
  - $F_{t,i}^{(v)} = S_{t,i} \cap F_t^{(v)}$



$S_{t,1}$ $S_{t,2}$

III. Prediction of Object Segmentation Instances

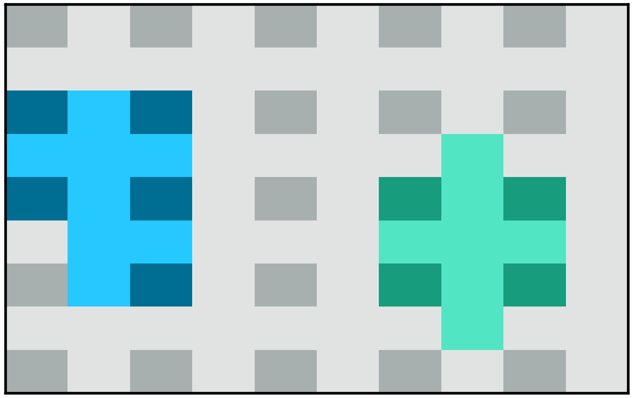
- Valid optical flow positions of instance i
  - $F_{t,i}^{(v)} = S_{t,i} \cap F_t^{(v)}$



$F_t^{(v)}$

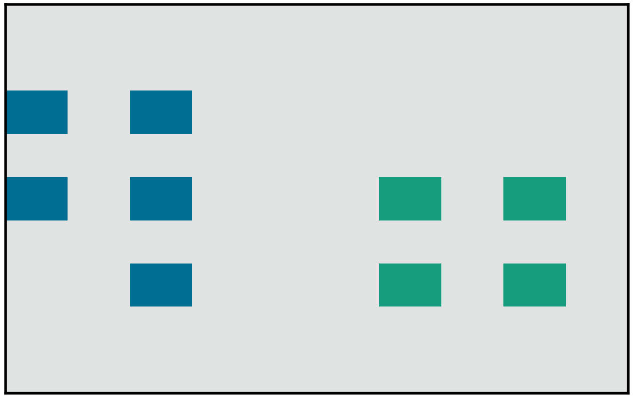
III. Prediction of Object Segmentation Instances

- Valid optical flow positions of instance i
  - $F_{t,i}^{(v)} = S_{t,i} \cap F_t^{(v)}$



III. Prediction of Object Segmentation Instances

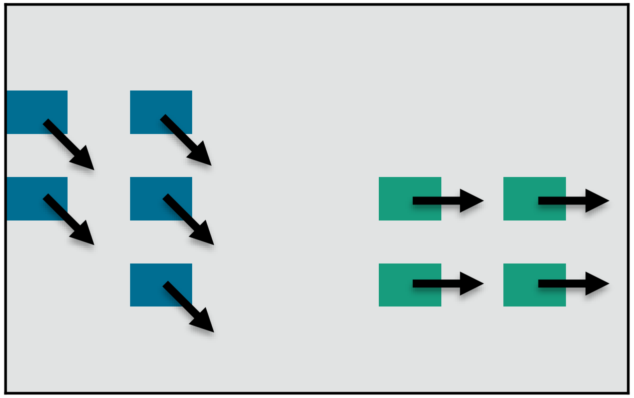
- Valid optical flow positions of instance i
  - $F_{t,i}^{(v)} = S_{t,i} \cap F_t^{(v)}$



Fraunhofer IOSB

III. Prediction of Object Segmentation Instances

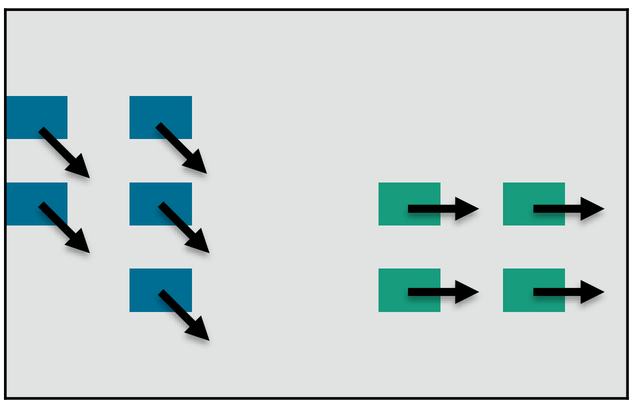
- Valid optical flow positions of instance i
  - $F_{t,i}^{(v)} = S_{t,i} \cap F_t^{(v)}$



$F_{t,1}^{(v)}$ $\qquad$ $F_{t,2}^{(v)}$

III. Prediction of Object Segmentation Instances

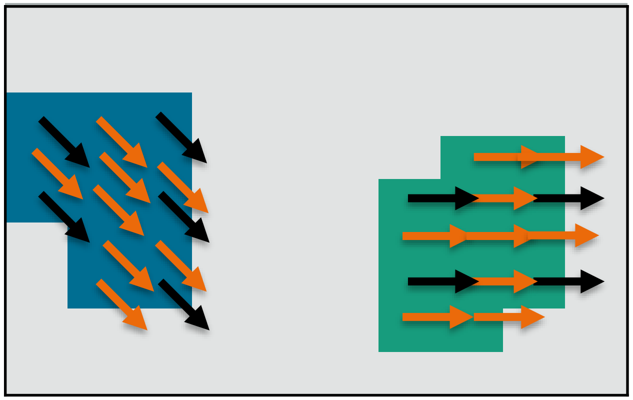
- Interpolation of optical flow at positions where no flow information is available
  - Linear interpolation inside the convex hull of $F_t^{(v)}$
  - Nearest neighbor interpolation outside of the convex hull of $F_t^{(v)}$



$F_{t,1}^{(v)}$ $F_{t,2}^{(v)}$

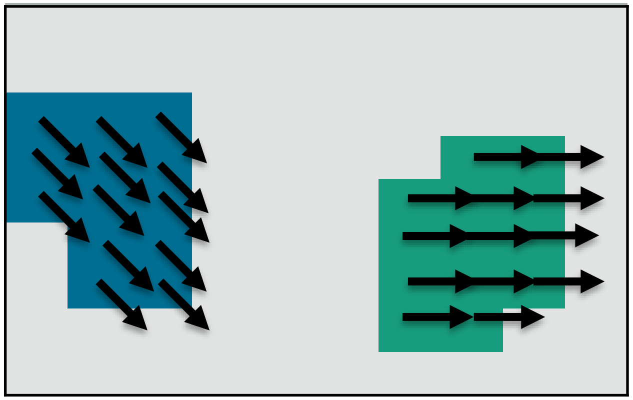
III. Prediction of Object Segmentation Instances

- Interpolation of optical flow at positions where no flow information is available
  - Linear interpolation inside the convex hull of $F_t^{(v)}$
  - Nearest neighbor interpolation outside of the convex hull of $F_t^{(v)}$



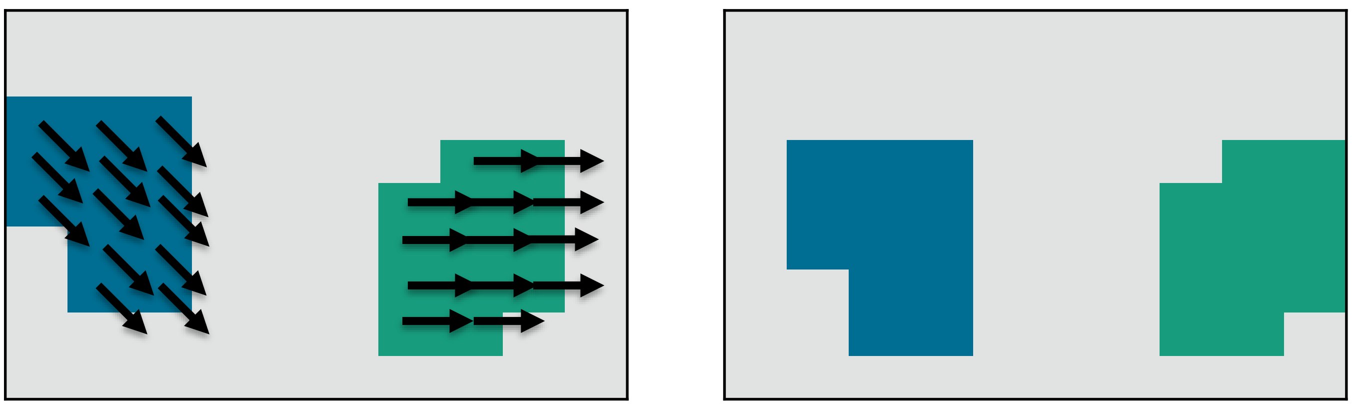
III. Prediction of Object Segmentation Instances

- Interpolation of optical flow at positions where no flow information is available
  - Linear interpolation inside the convex hull of $F_t^{(v)}$
  - Nearest neighbor interpolation outside of the convex hull of $F_t^{(v)}$



III. Prediction of Object Segmentation Instances

- Prediction of pixel positions of instance i
  - $P_{t\rightarrow t+o,i} = \left\{\left(x_p, y_p\right) \mid \left(x_p, y_p\right) = (x,y) + F_{t\rightarrow t+o}(x,y), (x,y) \in F_{t,i}^{(v)}\right\}$
  - Apply a morphological closing operation as post processing



$P_{t\rightarrow t+o,1}$ $P_{t\rightarrow t+o,2}$

IV. Affinity and Correspondence of Objects in Subsequent Frames

- Similarity between object i in frame I_t and object j in frame I_{t+o}

$P_{t \to t+o,i}$

$S_{t+o,j}$

  - Overlap $O_{i,j}$ of $P_{t \to t+o,i}$ and $S_{t+o,j}$
    - $O_{i,j} = \#(P_{t \to t+o,i} \cap S_{t+o,j})$
  - Usage of overlap O as similarity / affinity measure
- Similarity measure $O_{i,j}$ reflects
  - Locality
  - Visual similarity

Fraunhofer IOSB

IV. Affinity and Correspondence of Objects in Subsequent Frames

- Overlap $O_{i,j}$ of $P_{t\rightarrow t+o,i}$ and $S_{t+o,j}$
  - $O_{i,j} = \#(P_{t\rightarrow t+o,i} \cap S_{t+o,j})$
- Affinity / Similarity matrix
  - $A = \begin{bmatrix} O_{1,1} & \cdots & O_{1,j} & \cdots & O_{1,n_j} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ O_{i,1} & \cdots & O_{i,j} & \cdots & O_{i,n_j} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ O_{n_i,1} & \cdots & O_{n_i,j} & \cdots & O_{n_i,n_j} \end{bmatrix}$
- Computation of corresponding objects using the Kuhn-Munkres algorithm

Instance Flow Based Online Multiple Object Tracking: Tracker State

- Tracker state T_t at time t
  - consists of a a set of segmentation instances $S_{t,k}$
    - With a unique identifier $id_{t,k}$
    - With a counter for the number of successive missing detections $m_{t,k}$
  - $T_t = \{(S_{t,k}, id_{t,k}, m_{t,k} \mid k \in \{1, \ldots, n_t\})\}$
    - With the number of tracks n_t at time t

Fraunhofer IOSB

Instance Flow Based Online Multiple Object Tracking: Algorithm

- Initialize tracker state with segmentation instances of frame 0
- For subsequent frames
  - Compute instance-aware semantic segmentation (I)
  - Compute optical flow for current and previous image (II)
  - Predict segmentations $S_{t,k}$ using (III)
  - Solve associations (IV)
    - Building affinity matrix of predicted segmentations $S_{t,k}$ of the current tracker state and segmentations $S_{t+1,j}$ found in the current frame
  - Update matching tracker segmentations $S_{t,k}$ with corresponding segmentations $S_{t+1,j}$
  - Remove dead tracks
  - Update non-matching tracker segmentations $S_{t,k}$ with corresponding predictions
  - Create new tracks for unmatched segmentations $S_{t+1,j}$

Evaluation: References

- Instance-aware semantic segmentation
  - MNC: *Instance-Aware Semantic Segmentation via Multi-Task Network Cascades*, Dai et al., CVPR 2016
- Optical Flow / Dense Matching
  - PolyExp: *Two-frame motion estimation based on polynomial expansion*, Farnebäck, SCIA 2003
  - DeepMatch: *Deepmatching: Hierarchical deformable dense matching*, Revaud et al., IJCV 2016
  - CPM: *Efficient Coarse-to-Fine PatchMatch for Large Displacement Optical Flow*, Hu et al., CVPR 2016
- Multiple Object Tracking
  - SORT: Simple Online and Realtime Tracking, Alex Bewley et al., ICIP 2016

Evaluation

- MOT 2D 2015 Benchmark Test Set Evaluation

Method	md *	MOTA ↑	MOTP ↑
FasterRNN + SORT	-	33.4	72.1
MNC + SORT	-	27.5	70.5
MNC + CPM (ours)	0	30.6	71.3
MNC + CPM (ours)	1	32.1	70.9

- FasterRNN outperforms MNC in terms of detection rates
- Our approach outperforms SORT using same detection

*md = missing detections

Evaluation

- MOT 2015 Benchmark KITTI-13 Evaluation

Method	md	MOTA ↑	MOTP ↑
MNC + SORT	-	12.9	65.2
MNC + CPM (ours)	0	18.6	67.2
MNC + CPM (ours)	1	19.2	66.7
MNC + DeepMatch (ours)	0	18.6	67.2
MNC + DeepMatch (ours)	1	16.9	66.8
MNC + PolyExp (ours)	0	16.8	67.3
MNC + PolyExp (ours)	1	11.7	66.8

- Our method handles high object and camera motion better than SORT
  - Quality of optical flow matters!

Evaluation

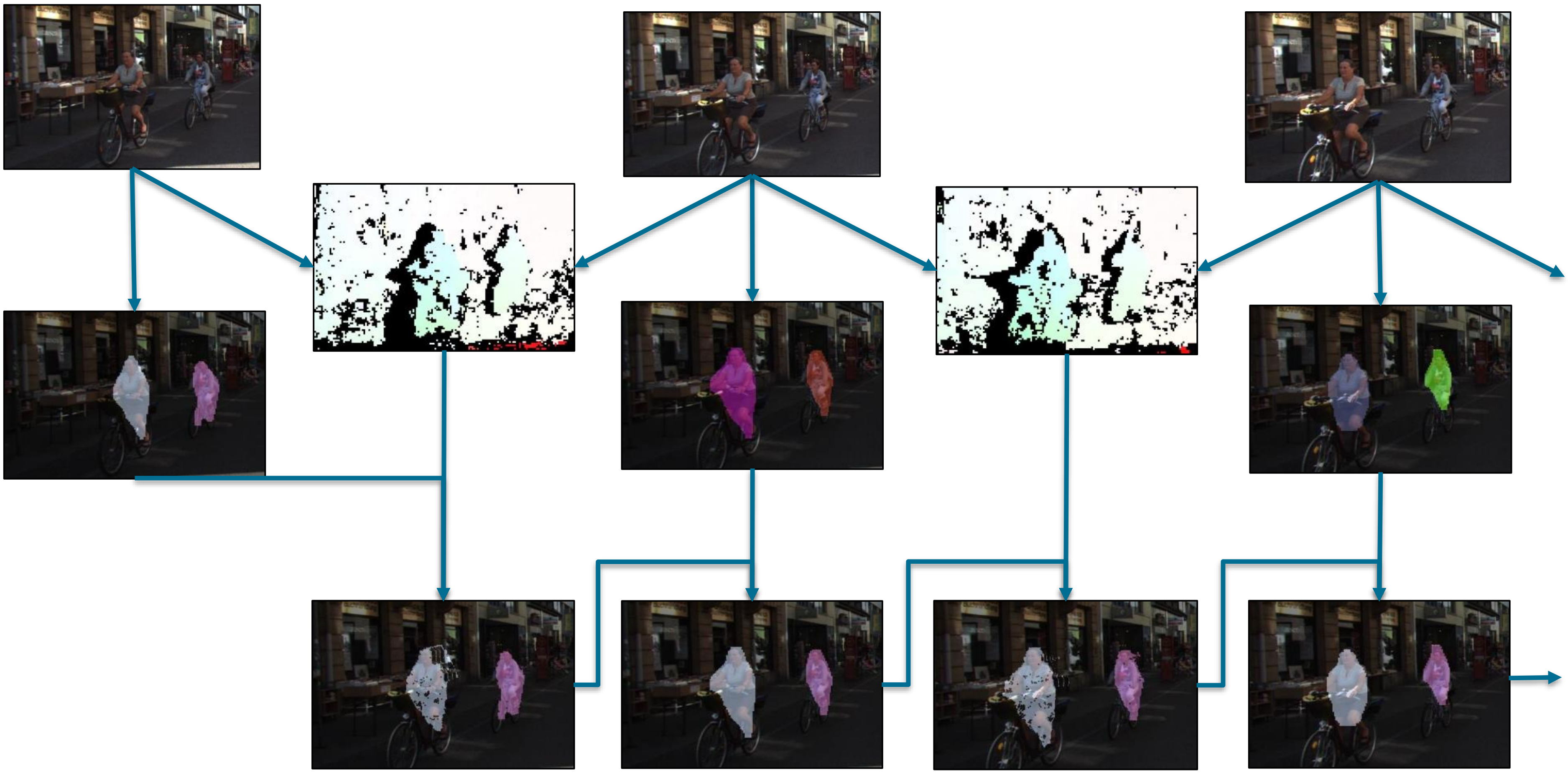
Prediction using CPM

Prediction using PolyExp

Conclusion

- We presented a MOT approach
  - Which combines optical flow and semantic segmentation to compute object correspondences in subsequent frames
  - Which is able to track the two-dimensional shape of objects
- The quality of the optical flow algorithm shows its importance for sequences with high object translation
- Detection vs. Tracking
  - Difficult to solely evaluate tracking performance
  - In many benchmark settings the overall performance is stronger influenced by the detector than the tracker

Instance Flow Based Online Multiple Object Tracking



Fraunhofer IOSB