

# Planar Object Tracking via Weighted Optical Flow



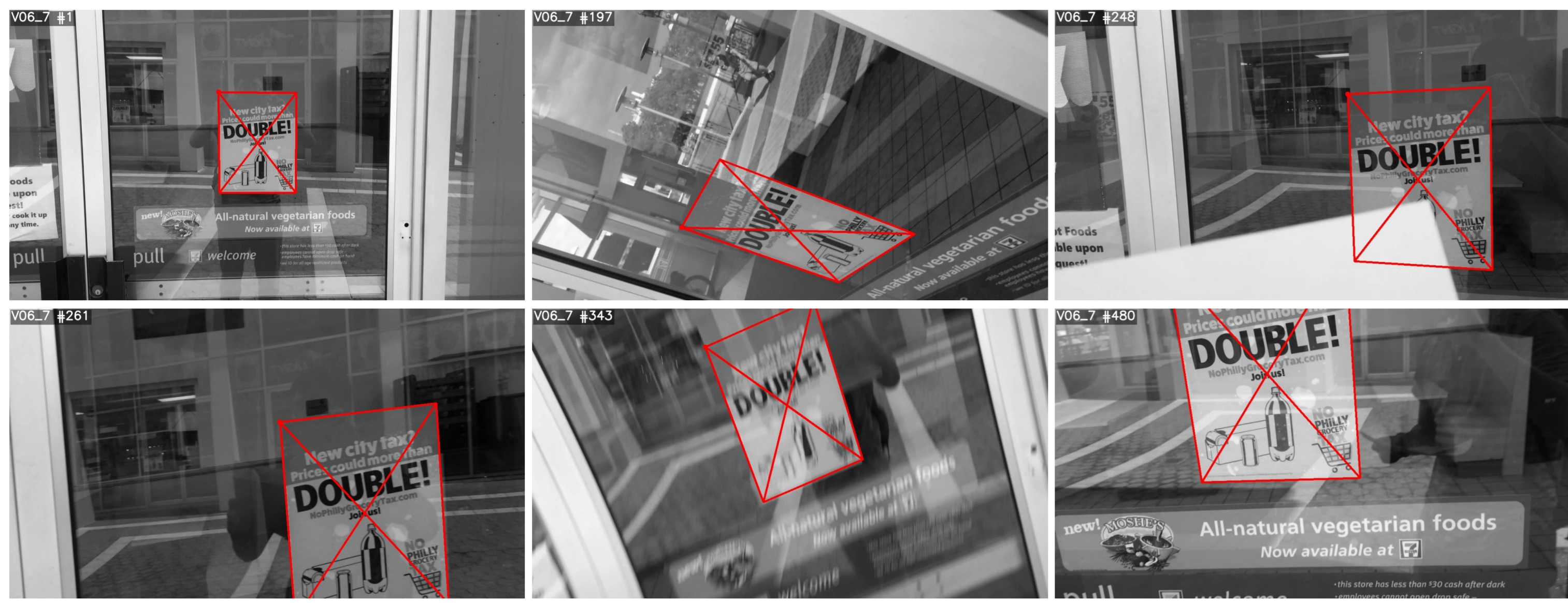
Jonáš Šerých, Jiří Matas

Center for Machine Perception, Czech Technical University in Prague, FEE



## Planar Object Tracking

= Homography  $\mathbf{H} \in \mathbb{R}^{3 \times 3}$  estimation in video



### Contributions

- A:** Fully differentiable Homography estimation method
- B:** Planar tracker achieving state-of-the-art results
- C:** Improved ground-truth on standard benchmark POT-210 [23]

## A: Homography estimation

**Idea: Estimate homography by weighted least squares from optical flow correspondences**

### Optical flow pros

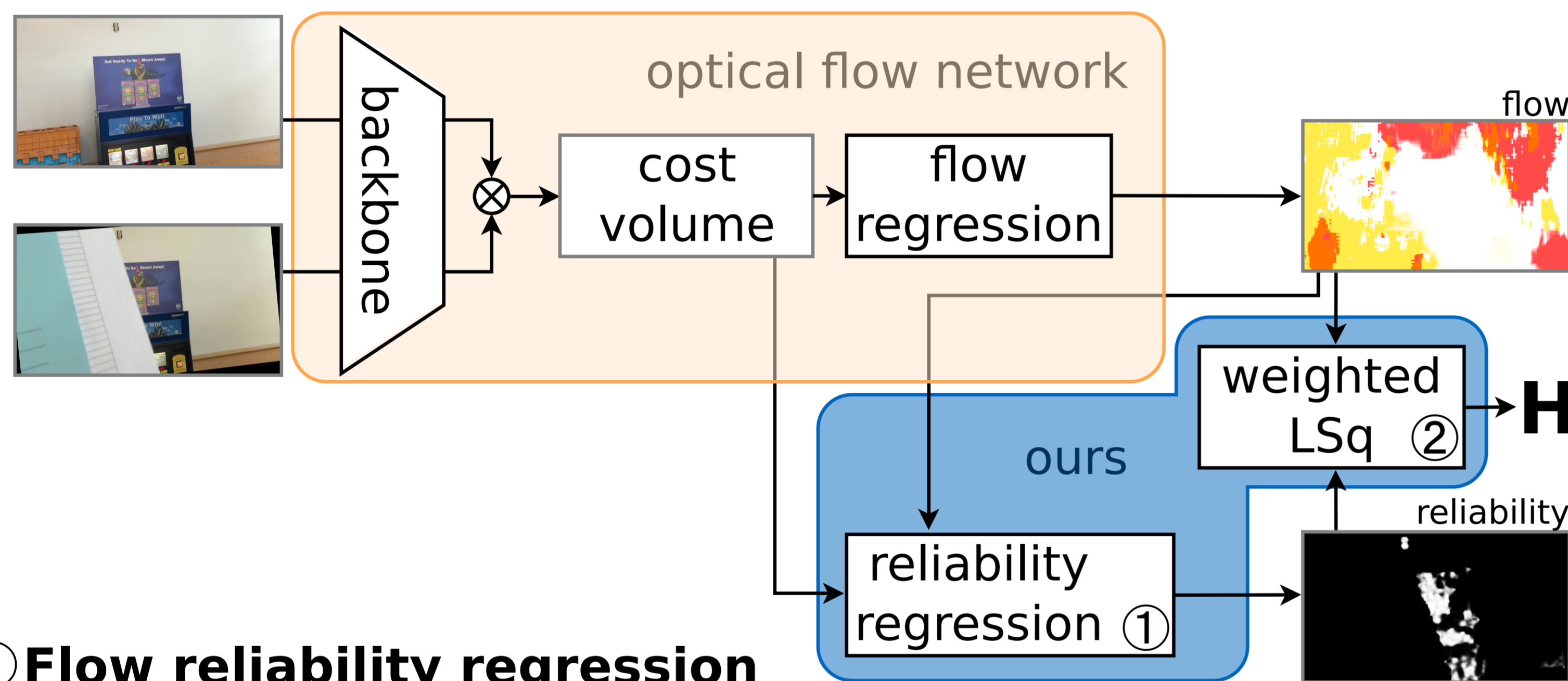
- + dense correspondences
- + mature algorithms available
- + near real-time
- + perform well on many scenes

### Optical flow cons

- not reliable everywhere
- gives no reliability estimate

→ **Standard solution**  
robust estimator (RANSAC)  
- not differentiable

### Our approach



#### ① Flow reliability regression

3-layer CNN, flow and costvolume as input

#### ② Weighted least squares (LSq): reliability scores as weights each correspondence gives:

$$\begin{bmatrix} 0 & 0 & 0 & -x_i & -y_i & -1 & y'_i x_i & y'_i y_i \\ x_i & y_i & 1 & 0 & 0 & 0 & -x'_i x_i & -x'_i y_i \end{bmatrix} \mathbf{h} = \begin{bmatrix} -y_i \\ x'_i \end{bmatrix}$$

$$\mathbf{A} \mathbf{h} = \mathbf{b}$$

$$\min_{\mathbf{h}} \sum_{j=1}^{2N} w_j \|\mathbf{A}_j \cdot \mathbf{h} - \mathbf{b}_j\|_2^2 \rightarrow \min_{\mathbf{h}} \sum_{j=1}^{2N} \|\bar{\mathbf{A}}_j \cdot \mathbf{h} - \bar{\mathbf{b}}_j\|_2^2$$

$$\rightarrow \bar{\mathbf{A}} = \mathbf{Q} \mathbf{R} \rightarrow \mathbf{R} \mathbf{h} = \mathbf{Q}^T \bar{\mathbf{b}}$$

QR decomposition, triangular system solver → differentiable

Code, results, improved GT:



### Our method is differentiable

→ we train both reliability and flow network

#### Training dataset



MS Coco + random H warps

#### Loss function

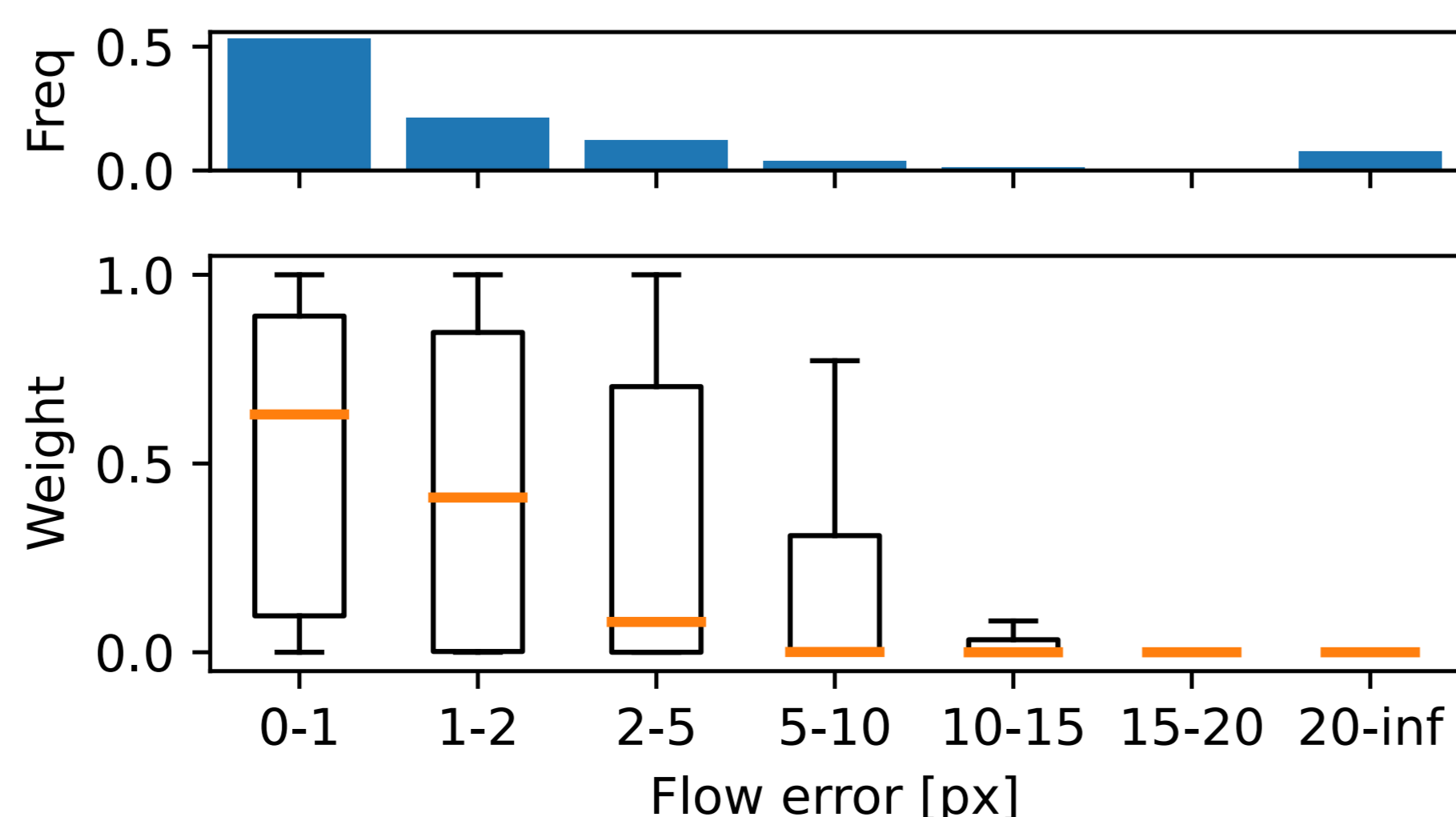
mean reprojection error

$$L(\mathbf{H}) = \frac{1}{N} \sum_{i=1}^N \|p_i - \mathcal{W}(\mathbf{H}^{-1} \mathbf{H}_{GT}, p_i)\|_2$$

Homography warp      position

### Trained reliability

- bad flow → low reliability
- good flow → often high reliability

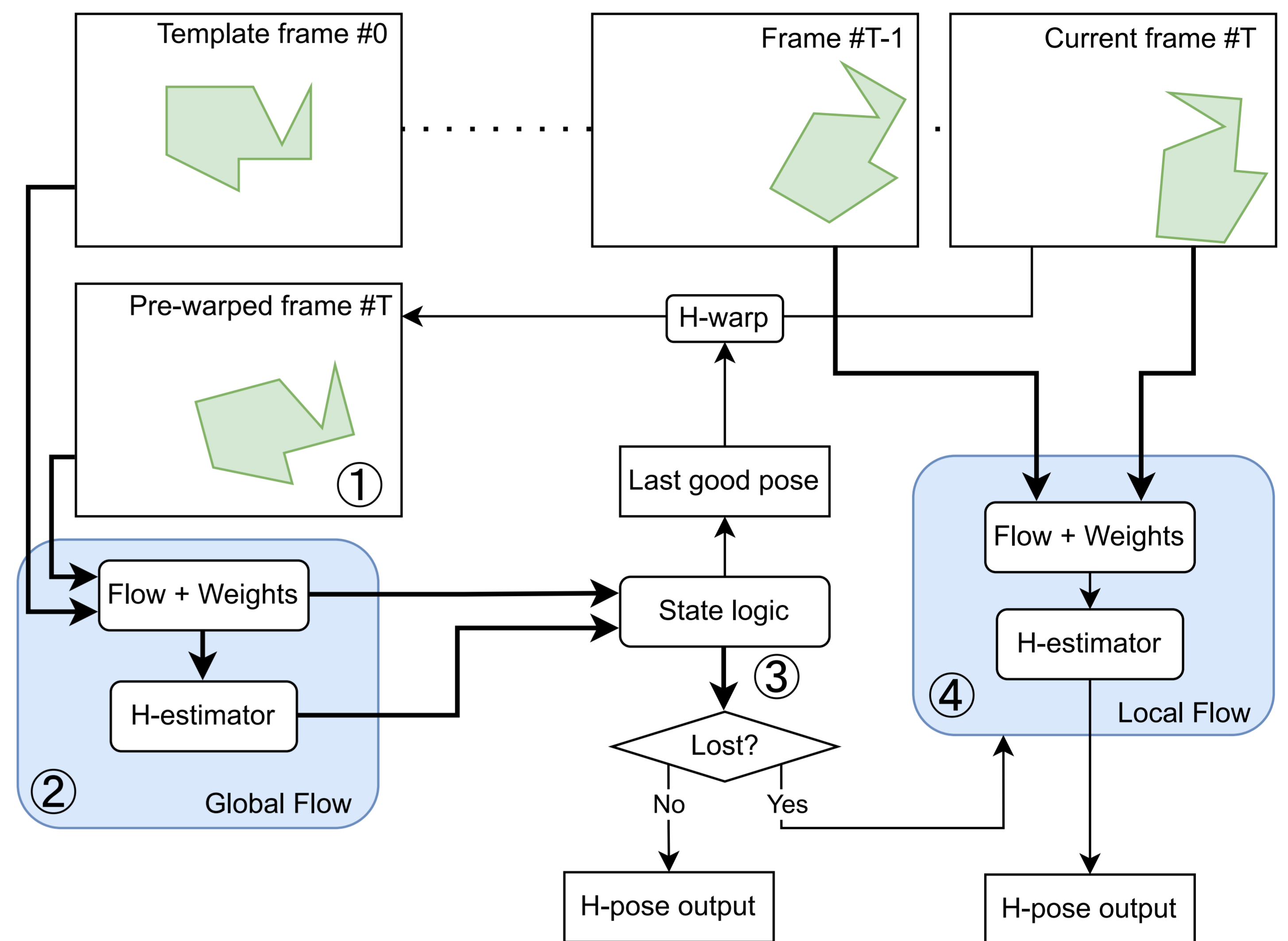


## B: WOFT Planar Tracker

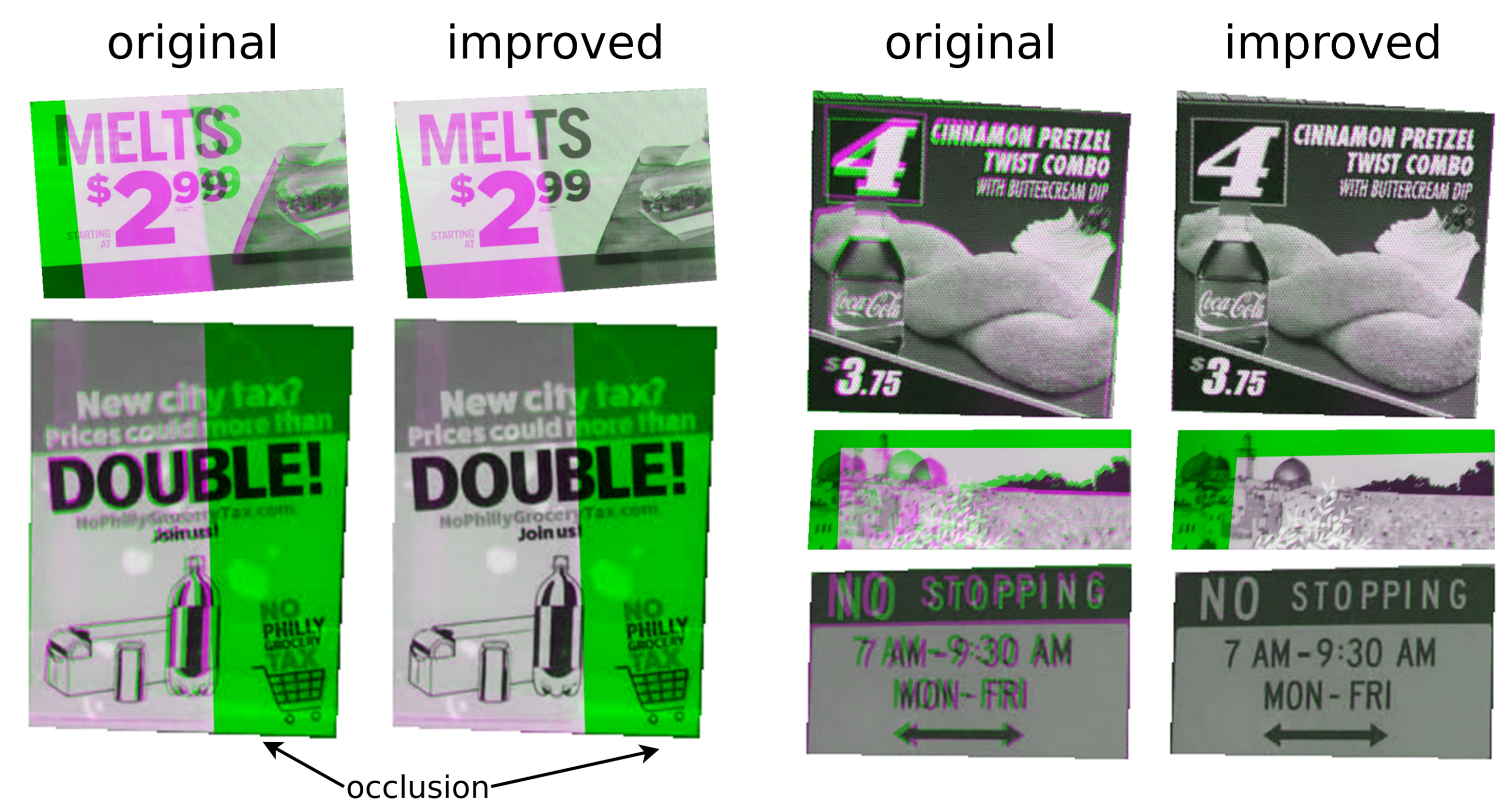
Chaining H estimations drifts, direct H estimation between the template and the current frame fails.

### WOFT Tracker

- ① Pre-warp the current frame using previous pose
- ② Estimate H between the template and the pre-warped frame
- ③ If that fails, ④ estimate H between the last two frames



## C: Improved POT-210 Ground Truth



## SOTA results on POT-210 dataset

P@N = fraction of frames with error under N pixels

method	year	FPS	improved GT		orig		rean	
			P@5	P@15	orig	rean	orig	rean
GOP-ESM [7]	2019	4.95 <sup>*</sup>	42.9	—	49.7	—	—	—
SuperGlue [38, 22]	2020	3.7 <sup>*</sup>	39.1	42.1	58.0	55.7	—	—
Gracker [45]	2017	4.8 <sup>*</sup>	39.2	—	63.2	—	—	—
SiamESM [6]	2019	—	58.7	—	66.2	—	—	—
SOSNet [43, 22]	2019	1.5 <sup>*</sup>	56.6	60.9	69.9	67.0	—	—
SIFT [27, 22]	2004	0.8 <sup>*</sup>	62.2	65.8	71.3	69.6	—	—
OBD [29]	2021	30 <sup>*</sup>	48.4	54.3	79.3	79.2	—	—
LISRD [32, 22]	2020	7 <sup>*</sup>	61.6	68.3	79.6	79.2	—	—
HDN [49]	2022	10.6 <sup>*</sup>	61.3	70.9	91.5	92.4	—	—
WOFT <sub>13</sub> (ours)	19.2	68.9	80.5	91.2	92.3	—	—	—
WOFT (ours)	3.5	80.6	90.4	93.9	95.6	—	—	—

half of the error caused by inaccurate GT

## References

- [6] Lin Chen, et al. Learning robust features for planar object tracking. IEEE Access 2019.
- [7] Lin Chen, et al. Robust visual tracking for planar objects using gradient orientation pyramid. JEI 2019.
- [22] Pengpeng Liang, et al. Planar object tracking benchmark in the wild. Neurocomputing 2021.
- [23] Pengpeng Liang, et al. Planar object tracking in the wild: A benchmark. ICRA 2018.
- [27] David G Lowe. Distinctive image features from scale-invariant keypoints. IJCV 2004.
- [29] Dmitrii Matveichev, Daw-Tung Lin. Mobile augmented reality: Fast, precise, and smooth planar object tracking. ICPR 2020
- [32] Rémi Pautrat, et al. Online invariance selection for local feature descriptors. ECCV 2020.
- [38] Paul-Edouard Sarlin, et al. Superglue: Learning feature matching with graph neural networks. CVPR 2020.
- [43] Yurun Tian, et al. SOSNet: Second order similarity regularization for local descriptor learning. CVPR 2019.
- [45] Tao Wang, Haibin Ling. Gracker: A graph-based planar object tracker. PAMI 2017.
- [49] Xinrui Zhan, et al. Homography decomposition networks for planar object tracking. AAAI 2022.

## Acknowledgments

This work was supported by Toyota Motor Europe, by CTU student grant SGS20/171/OHK3/3T/13, and by the Research Center for Informatics project CZ.02.1.01/0.0/0.0/16\_019/0000765 funded by OP VVV.