

# Segmentation-Guided Homography Estimation for Long-Term Planar Tracking

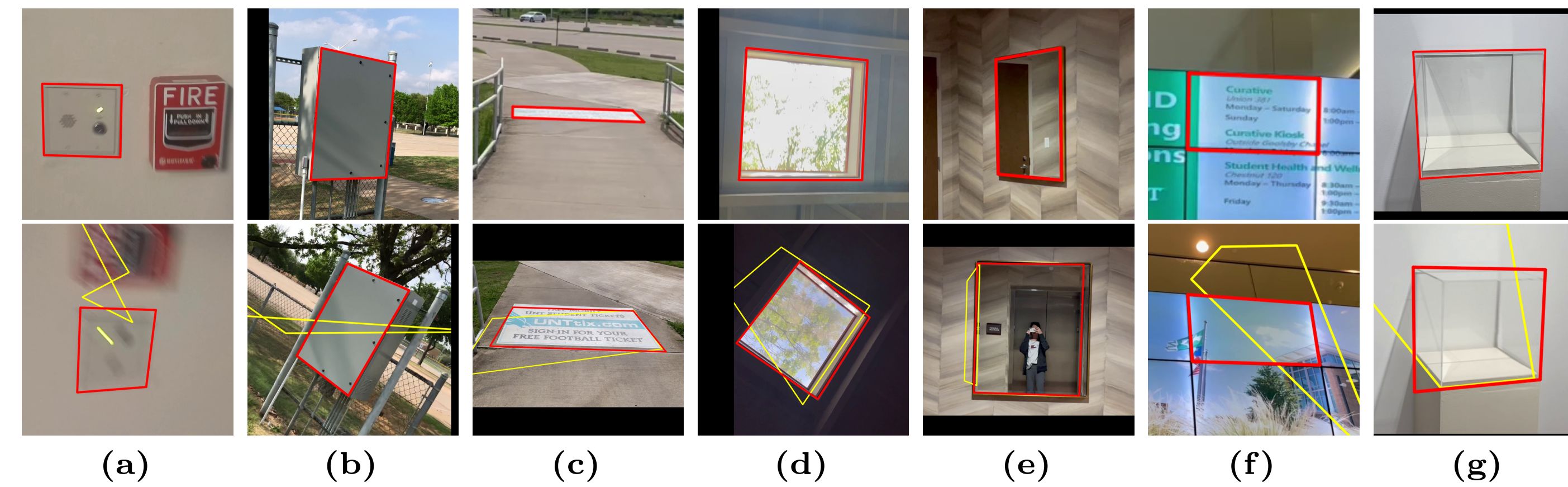
Jonas Serych, Jiri Matas

Visual Recognition Group, CTU in Prague



## 1 Planar Tracking Task

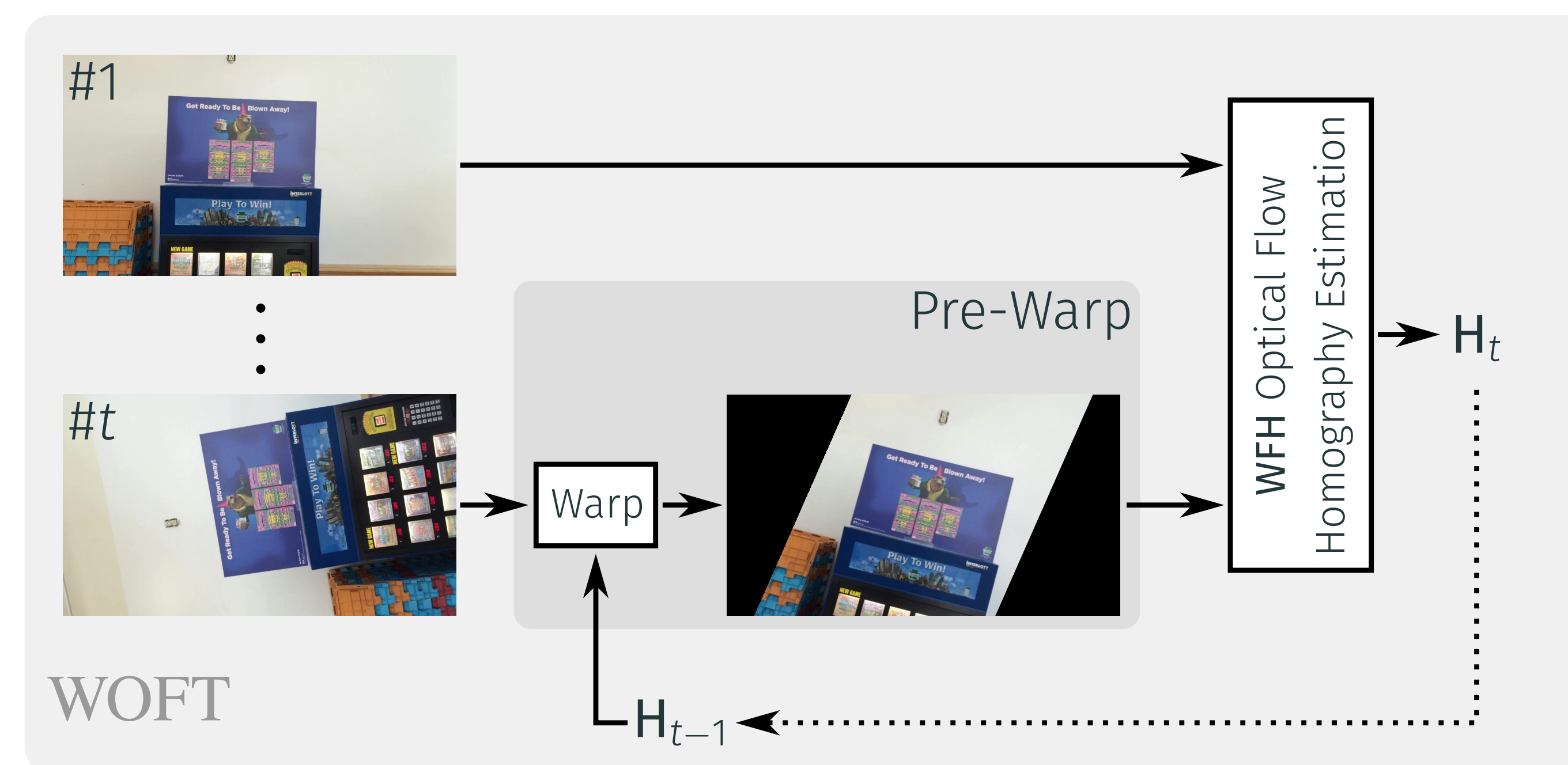
Estimate the target's 8-DoF homography in every frame.



$$\lambda \begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \underbrace{\begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix}}_{\text{8 DoF homography } H} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

## 2 Prior SOTA: Precise but Fragile

WOFT aligns optical flow to the pre-warped first frame: precise, but it struggles to recover once the target is lost.

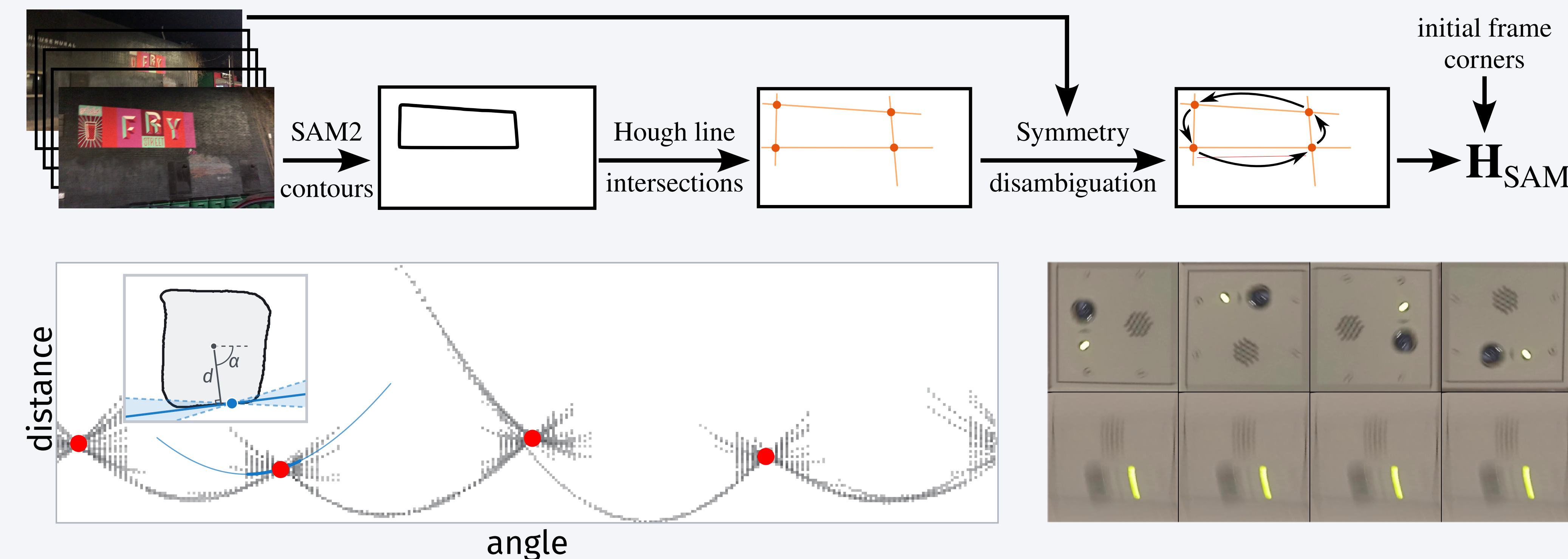


Blur, weak texture or target loss break WOFT, but SAM 2 masks survive.

**Problem** — blur, weak texture, and occlusions break homography estimation.  
**Insight** — segmentation (SAM 2) survives all of it, but a mask has no pose.  
**Method** — SAM-H extracts an 8-DoF homography from the mask boundary.  
**Result** — new planar tracking state of the art, by a large margin.

## 3 SAM-H: From Mask to Homography

Training-free: four Hough lines on the mask contour give the corners.

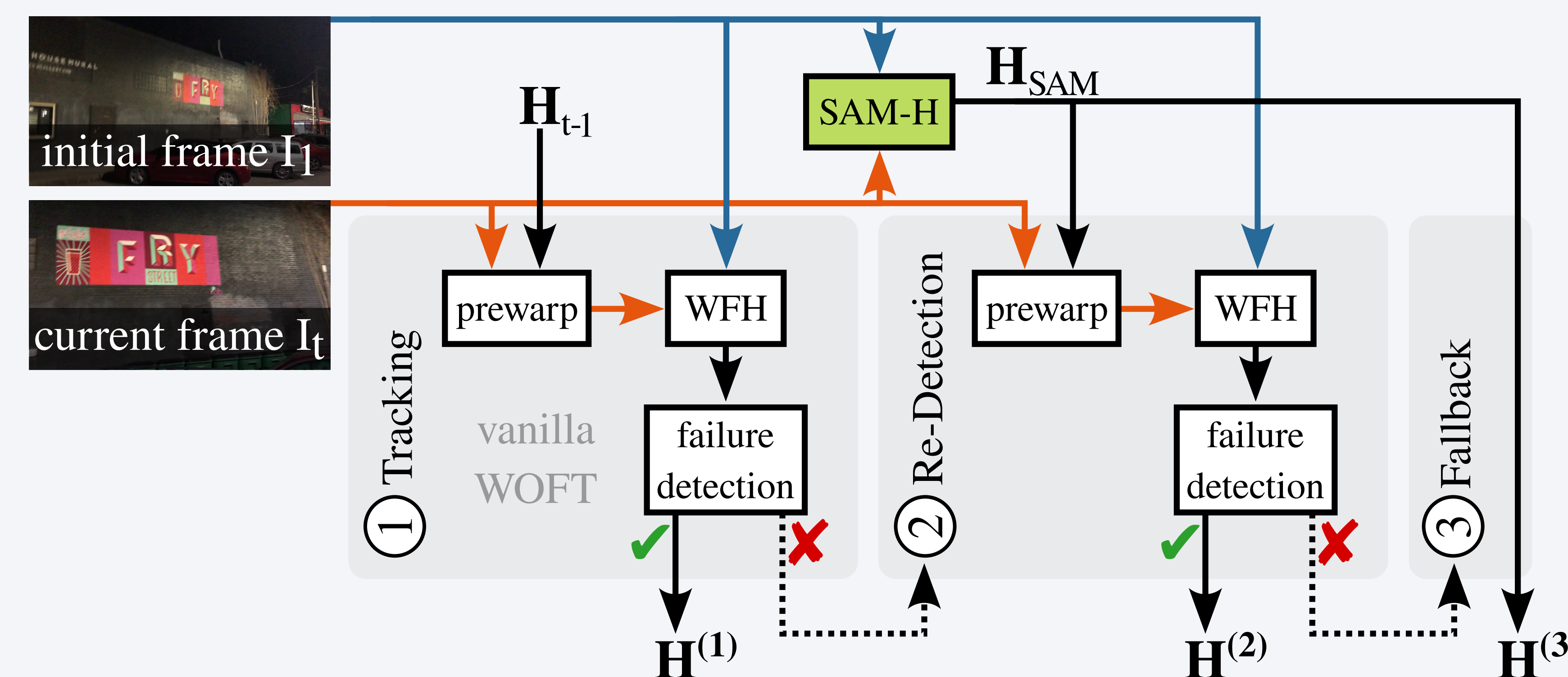


Each boundary point votes into the Hough accumulator; Accumulator peaks = the four target edges.

Four corner orderings fit the same mask; DINOv2 features pick the right one.

## 4 WOFTSAM: Track with WOFT, Re-Detect with SAM-H

Too few flow inliers → pre-warp with SAM-H and retry; still failing → output SAM-H.



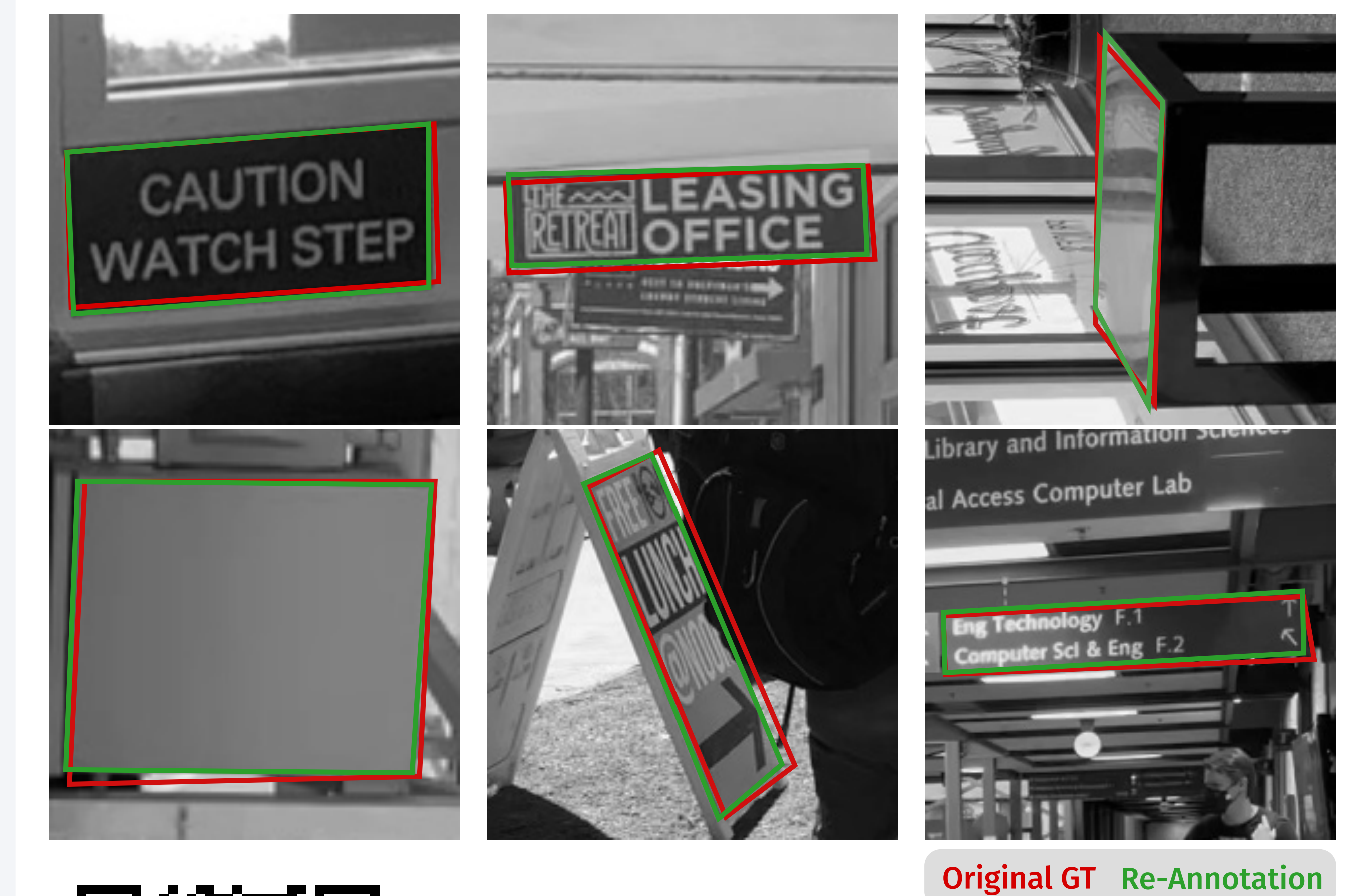
## 5 State-of-the-Art Results

WOFTSAM outperforms prior trackers on both benchmarks.  
 SAM-H alone: +15.2 pp p@15 on PlanarTrack.

| method         | POT-210     |             | PT <sub>TST</sub> |             | PT <sub>TST</sub> re-annot |                    |
|----------------|-------------|-------------|-------------------|-------------|----------------------------|--------------------|
|                | p@5         | p@15        | p@5               | p@15        | p@5                        | p@15               |
| SIFT           | 65.8        | 69.6        | 17.6              | 26.7        | —                          | —                  |
| LISRD          | 68.3        | 79.2        | 22.0              | 37.2        | —                          | —                  |
| HDN            | 70.9        | 92.4        | 26.6              | 56.5        | —                          | —                  |
| CGN            | 78.5        | 93.3        | —                 | —           | —                          | —                  |
| HVC-Net        | 91.4        | 95.8        | 42.0              | 63.1        | 48.8 (+6.8)                | 64.3 (+1.2)        |
| WOFT           | 90.0        | 95.4        | 43.6              | 64.8        | 48.9 (+5.3)                | 65.9 (+1.1)        |
| <b>WOFTSAM</b> | <b>91.9</b> | <b>97.5</b> | <b>51.5</b>       | <b>77.2</b> | <b>57.4 (+5.9)</b>         | <b>77.6 (+0.4)</b> |
| <b>SAM-H</b>   | 64.4        | 89.0        | <b>62.0</b>       | <b>80.0</b> | <b>62.0 (+0.0)</b>         | <b>80.1 (+0.1)</b> |

## 6 Precise PlanarTrack Re-Annotation

Init-frame GT error amplifies as the target grows, explaining over half of the residual p@5 gap.



Code & PlanarTrack init frame re-annotations  
[github.com/serycjon/WOFTSAM](https://github.com/serycjon/WOFTSAM)