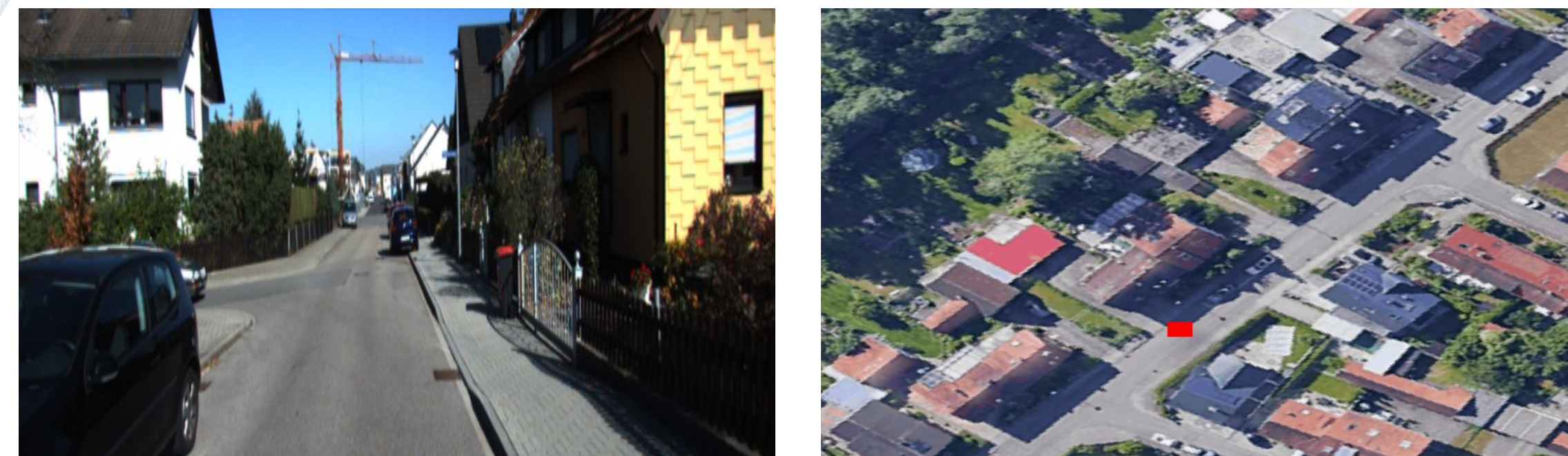


LAYS: Cross-View Yaw Estimation in Location Uncertainty with Line-Aligning Yaw Scoring

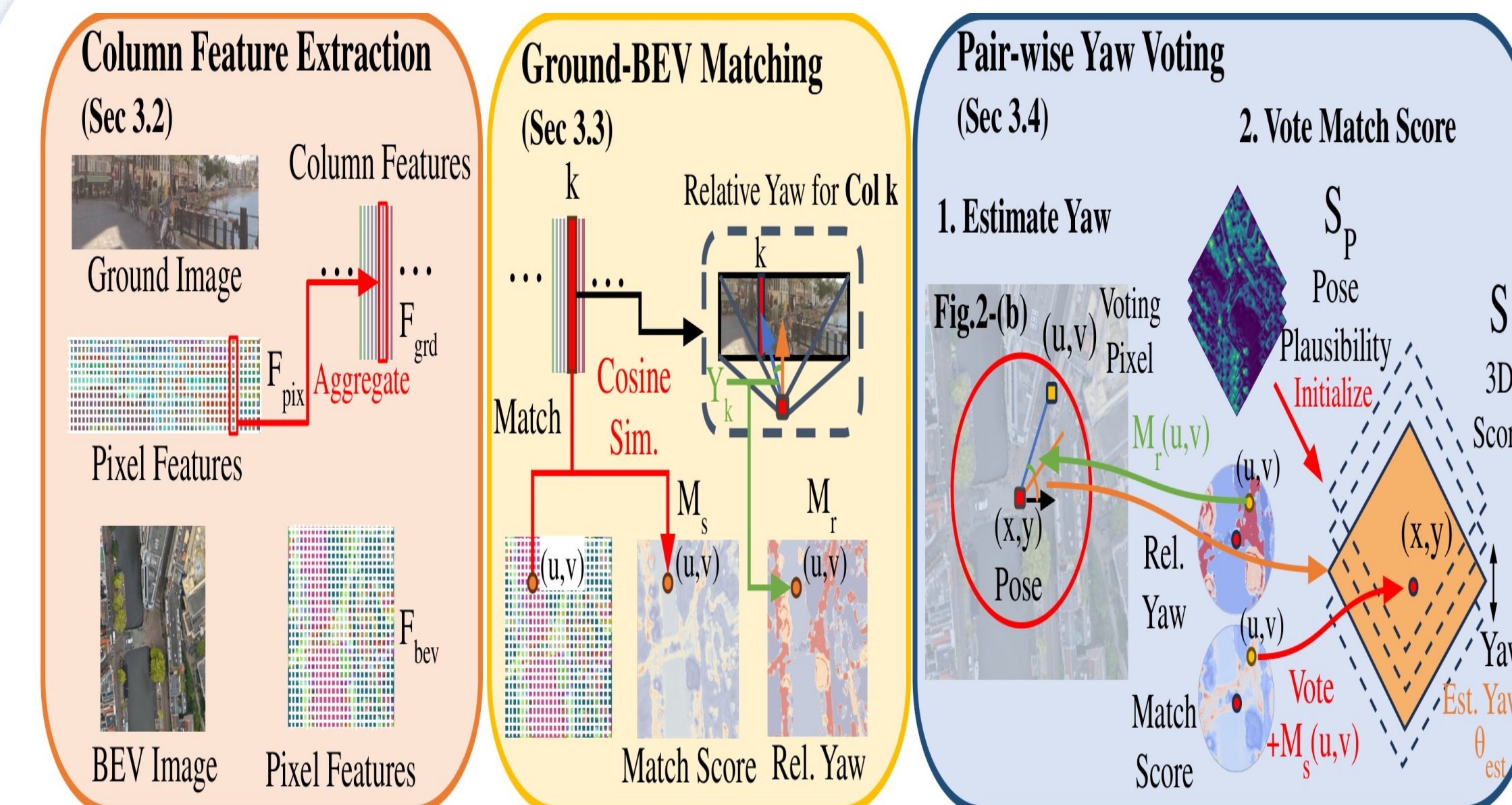
TaeHo Kang¹ · Nairan Zhang¹ · Yelin Kim¹ · Yujiao Shi² · Youngki Lee¹
¹ Seoul National University (HCS Lab) ² Nanyang Technological University · ECCV 2026 Oral Presentation

1. Problem Setup & Motivation



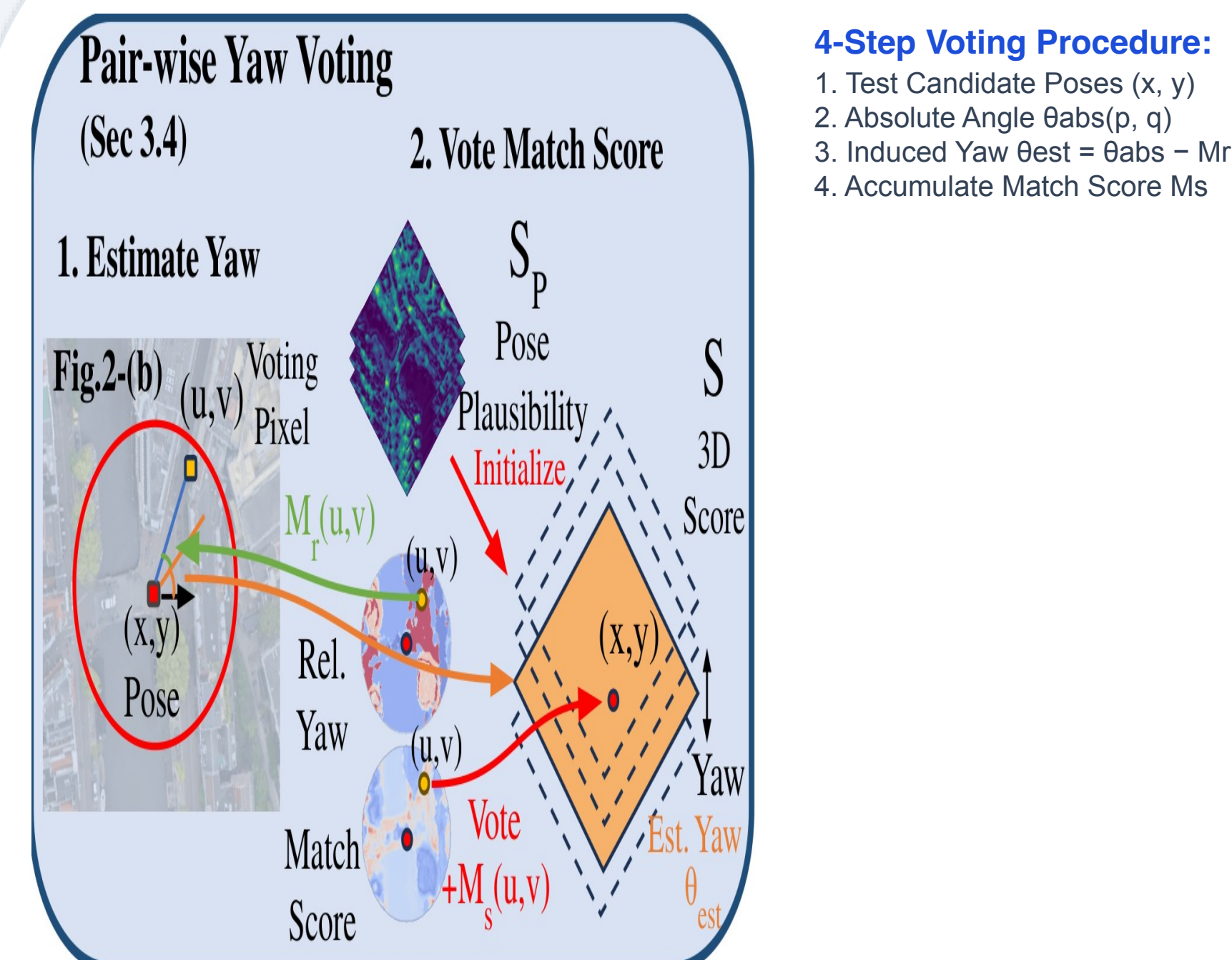
- Real-World Challenge: Standard GPS supplies only a coarse 2D position fix (± 20 m noise), while camera yaw is completely unknown ($\pm 180^\circ$).
- Breakdown of 3-DoF Search: Existing cross-view localizers optimize (x, y, θ) jointly. Severe location error causes rotational optimization to get trapped in bad local minima.
- ★ **LAYS Strategy:** Decouple yaw estimation into an independent, position-invariant line consensus task to achieve strict sub-degree ($< 1^\circ$) yaw accuracy.

3. LAYS 3-Stage Architecture



- End-to-End Neural Architecture:
 1. Column Feature Extraction: Bearing coordinate channels + Vertical attention softmax.
 2. Dense Ground-BEV Matching: Pairwise cosine similarity + Soft training sampling.
 3. Pair-Wise 3D Consensus Voting: Discrete vote accumulation + Multi-resolution pyramid.

5. Stage 3: Pair-Wise 3D Consensus Voting



- Combined Score Tensor:** $S(\theta, x, y) = SM(\theta, x, y) + SP(\theta, x, y)$
- SM: Geometric line-matching consensus volume.
 - SP: Learned BEV pose prior (road drivability).
 - 3-Level Pyramid: $4^\circ \rightarrow 2^\circ \rightarrow 1^\circ$ coarse-to-fine refinement.
 - Final Decision: $\theta_{\text{best}} = \arg \max_{\theta} \max_{(x,y)} S(\theta, x, y)$.

7. Quantitative Benchmark Results ($< 1^\circ$)

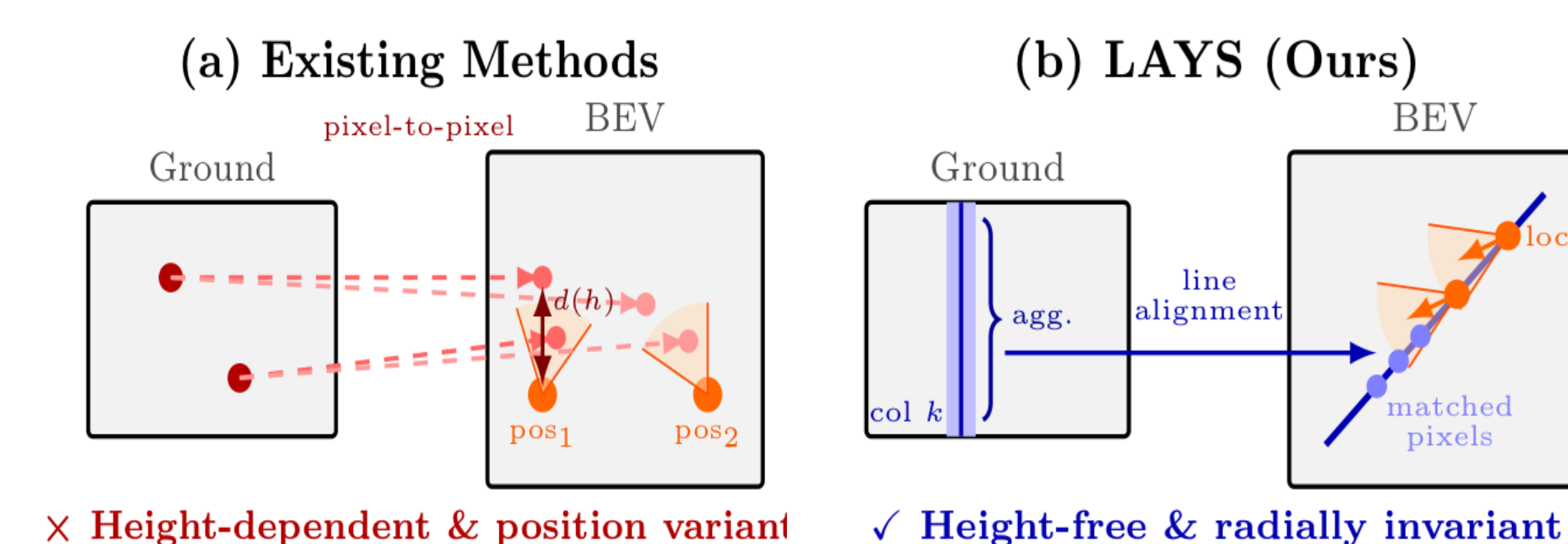
MGL (Urban Canyon · $\pm 180^\circ$)
 LAYS: 34.81% vs CCVPE: 6.55%
 Gain: +28.3 pp ($> 5\times$ Gain)

KITTI (Autonomous Driving · $\pm 180^\circ$)
 LAYS: 49.59% vs CCVPE: 3.14%
 Gain: +46.5 pp ($\sim 15\times$ Gain)

Ford Multi-AV (Highway · $\pm 45^\circ$)
 LAYS: 67.05% vs CCVPE: 41.36%
 Gain: +25.7 pp Gain

VIGOR (Panoramas · $\pm 180^\circ$)
 LAYS: 34.16% vs FG2†: 20.78%
 Gain: +13.4 pp Gain

2. Core Insight: Radial Line Invariance

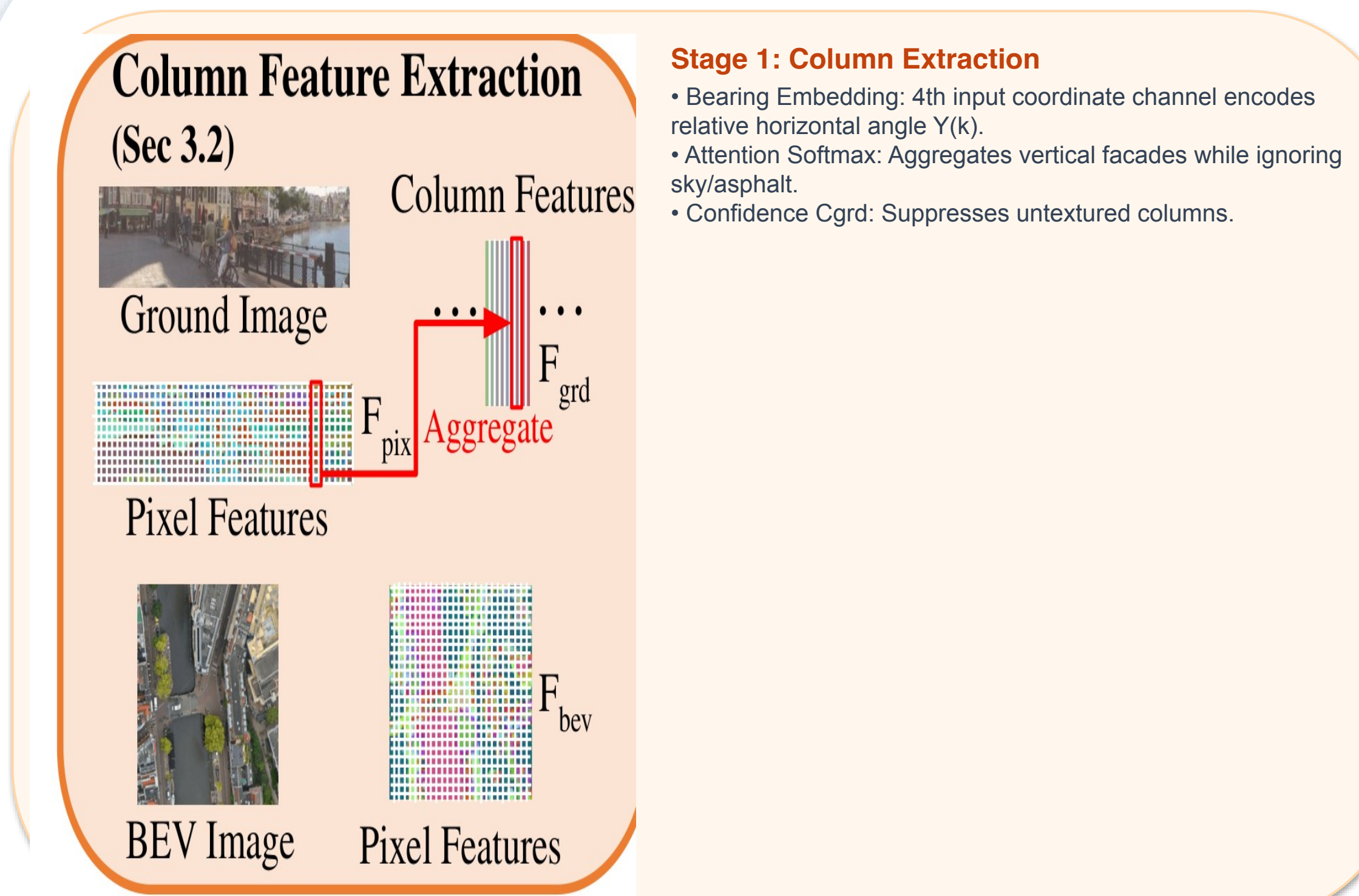


Proposition 1 (Position-Invariant Yaw Along Radial Rays)

For any ground column k matched to BEV pixel (u, v) , any two candidate camera poses $p1 = (x1, y1)$ and $p2 = (x2, y2)$ collinear with (u, v) satisfy:

$$\theta_{\text{best}}(p1, q) = \theta_{\text{best}}(p2, q) = \arctan((v - y) / (u - x)) - Y(k)$$

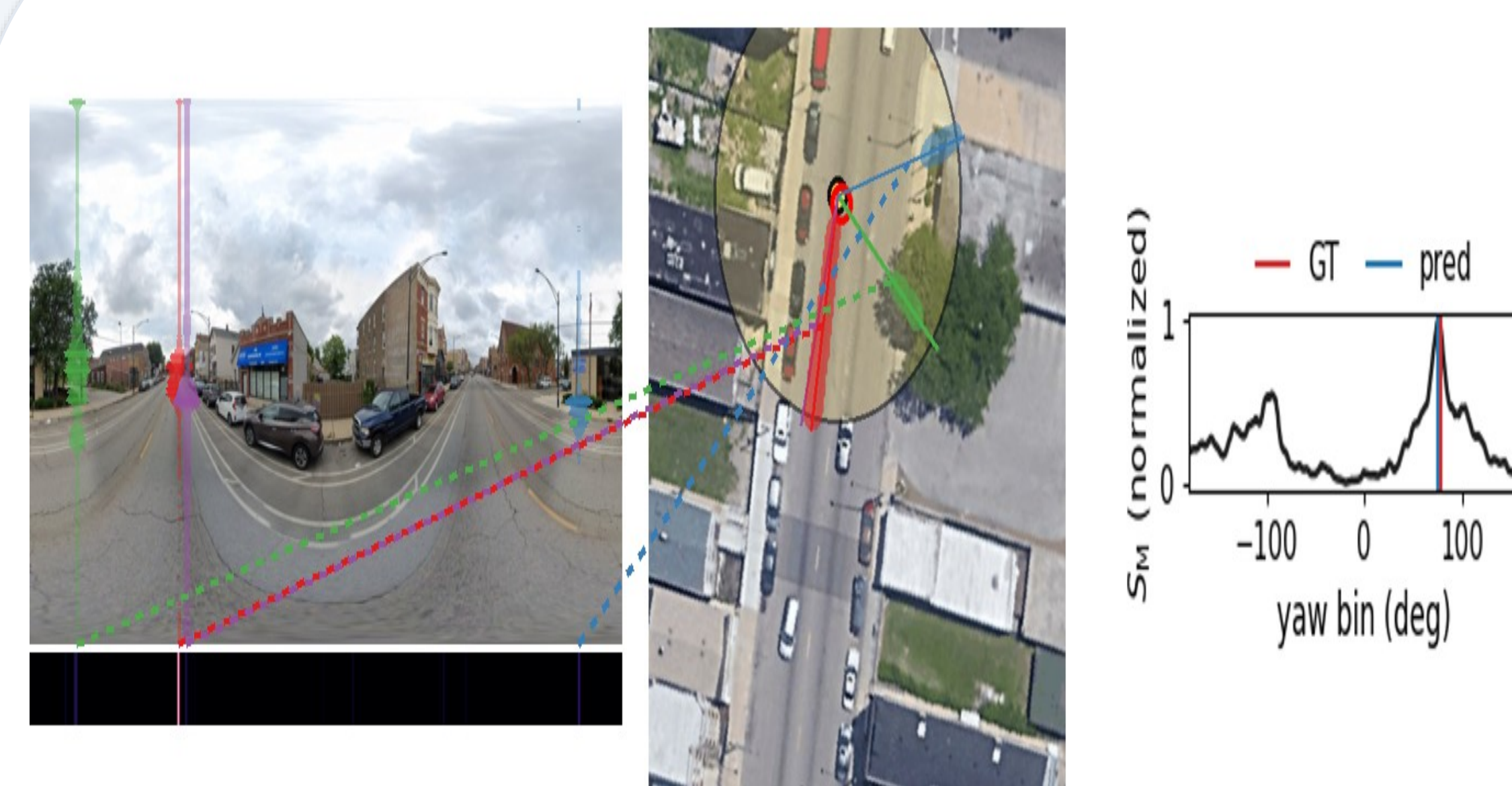
4. Stage 1 & 2: Features & Matching



- Stage 1: Column Extraction**
- Bearing Embedding: 4th input coordinate channel encodes relative horizontal angle $Y(k)$.
 - Attention Softmax: Aggregates vertical facades while ignoring sky/asphalt.
 - Confidence Cgrd: Suppresses untextured columns.

- Stage 2: Cross-View Matching**
- Cosine Match Score: Dense similarity $P((u, v), k) = |F_{\text{bev}} \cdot F_{\text{grd}}| \cdot C_{\text{grd}}$.
 - Probabilistic Training: Soft sampling explores diverse multi-cue matches.
 - Fast Inference: Deterministic argmax produces score map M_s and relative yaw map M_r .

6. Qualitative Consensus Evidence



Real Scene Evidence (VIGOR Panorama Benchmark):

- Diverse Multi-Cue Rays: Road edges, building corners, and tree lines cast independent radial rays into BEV.
- Sharp GT Consensus Peak: All valid rays intersect and reinforce each other, yielding a sharp spike at true yaw ($< 0.5^\circ$ error).
- Symmetry Disambiguation: Learned pose prior SP resolves 180° road symmetry ambiguity.

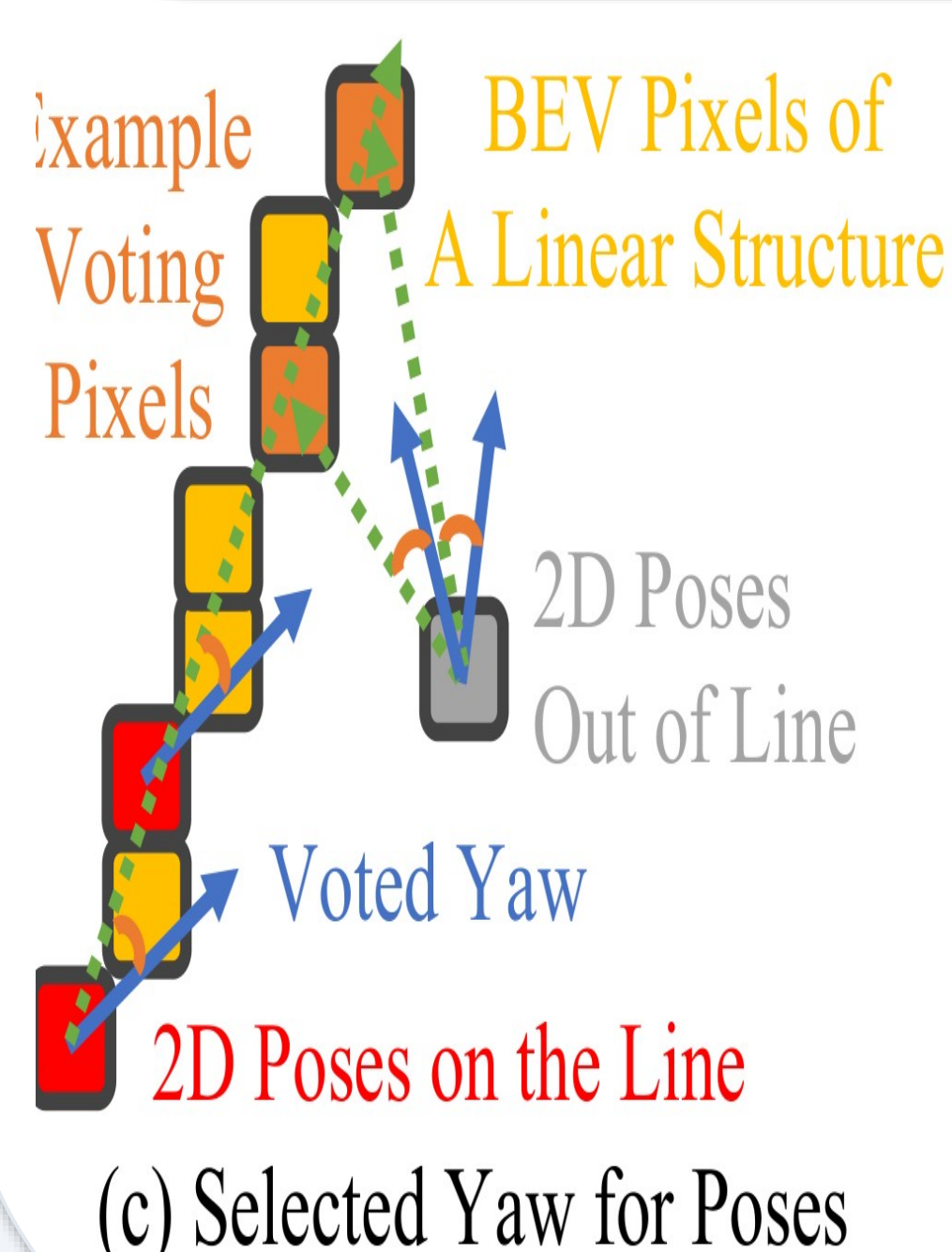
8. Downstream Impact (3-DoF $\rightarrow$ 2-DoF)

Collapsing Search: $(x, y, \theta) \rightarrow (x, y)$

- KITTI Median Position Error: 12.26 m $\rightarrow$ 1.48 m (🔥 88% Error Reduction!)
- MGL Median Position Error: 8.95 m $\rightarrow$ 3.99 m (🔥 55% Error Reduction!)
- Evaluated via FG2 (Xia et al., CVPR 2025). Supplying LAYS yaw eliminates local rotational minima.

Key Takeaways:

1. Decoupled yaw via radially invariant line consensus.
2. Height-free vertical feature aggregation.
3. State-of-the-art sub-degree accuracy across 4 benchmarks.
4. Enables high-precision 2-DoF visual localization.



- Consensus Properties:**
- On-Line Consensus: All collinear candidate poses vote for the EXACT same true yaw, reinforcing a massive peak.
 - Off-Line Noise: Divergent angles cancel into flat background noise.
 - Height-Free: Zero dependence on camera elevation or flat-ground assumptions.