Towards Unconstrained Cross-View Pose Estimation

Alexander Wollam[1], Kyle Ashley[2], Maxim Shugaev[2], Oliver Arend[2],
Ilya Semenov[2], Sumved Ravi[2], Hadis Dashtestani[2], Nathan Jacobs[1]
[1]Washington University in St. Louis, [2]Blue Halo

WACV 2026 TUCSON, AZ 3/6 – 3/10

Introduction

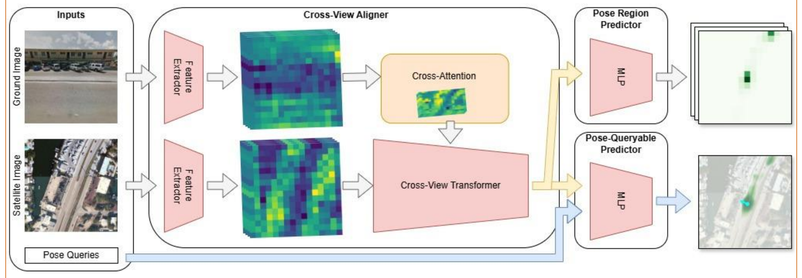
Background

- Cross-View Pose estimation has seen increasing interest in enabling refined localization/orientation of ground images within aerial views for autonomous vehicles, robotics, and within geo-localization.
- Existing work relies on constrained settings/known camera parameters to perform pixel/latent matching to derive the relative camera pose, limiting application with varied imagery

Proposal

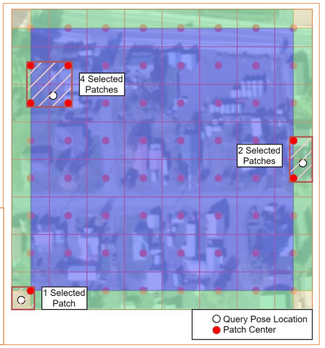
We propose leveraging an attention-based model to implicitly learn ground/aerial –view matching. This extends these models to such less-constrained settings by enabling operation with varied and unknown meta-data—acting as a step towards unconstrained cross-view pose estimation.

Proposed Architecture



Architecture Components

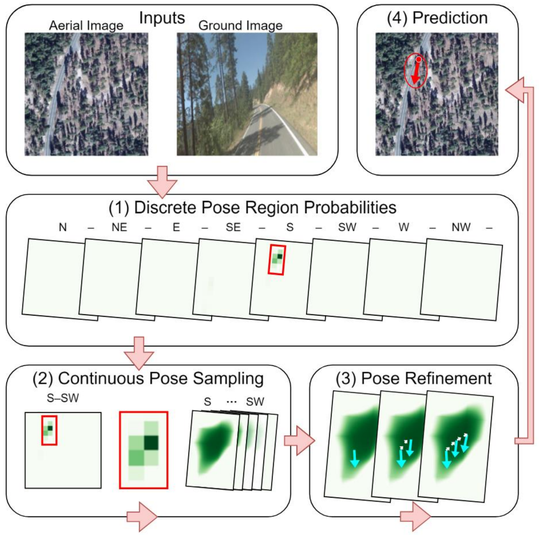
- **Cross-View Aligner**
  Merge cross-view features
- **Pose Region Predictor**
  Generate coarse pose prediction
- **Pose-Queryable Predictor**
  Enable fine-grained prediction



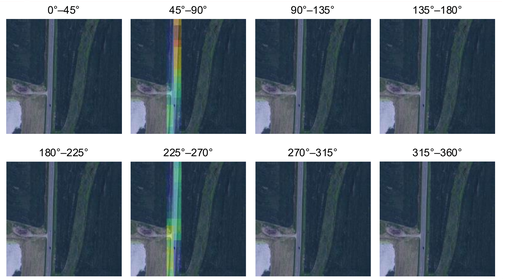
Pose-Queryable Predictor Feature Selection

- Each patch's feature-vector is used for predictions within a local region only
- Select & predict for 'surrounding' features per query—then merge predictions

Proposed Approach



Highlighted Example

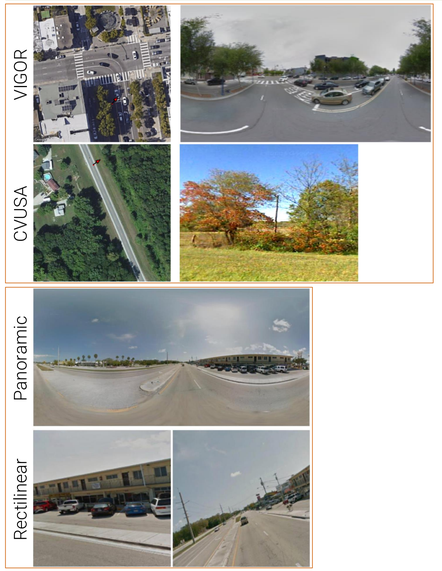


Model predicts strongly (and well!) despite few pixel matches. ***How?***

Observed data biases:

- *Model learns non-matching biases such as terrain likelihood (e.g. road vs non-road locations)*
- *Model learns sun-alignment biases (i.e. Daytime Northern hemisphere imagery → sun roughly in Southern half of sky (even without known/consistent times))*

Results



VIGOR vs CVUSA

Dataset	Method	↓Localization (m) Mean	↓Localization (m) Median	↓Orientation (°) Mean	↓Orientation (°) Median
VIGOR	CCVPE*	8.01	**1.67**	13.22	5.33
	Ours	**5.40**	1.76	**7.87**	**1.68**
CVUSA	CCVPE*	16.56	3.68	51.43	45.68
	Ours	**7.57**	**3.13**	**4.51**	**1.55**

Proj. and H-FoV

	H-FoV	Method	↓Localization (m) Mean	↓Localization (m) Median	↓Orientation (°) Mean	↓Orientation (°) Median
Panoramic	360	CCVPE*	8.01	**1.67**	13.22	5.33
		Ours	**5.40**	1.76	**7.87**	**1.68**
	180	CCVPE*	13.53	3.73	24.03	7.57
		Ours	**5.86**	**1.82**	**8.72**	**1.73**
	90	CCVPE*	21.12	14.11	43.16	16.14
		Ours	**6.04**	**1.88**	**8.89**	**1.76**
Rectilinear	90	CCVPE*	15.36	4.74	37.72	8.95
		Ours	**9.12**	**2.81**	**13.40**	**2.63**
	60	CCVPE*	16.86	7.62	50.83	14.48
		Ours	**9.87**	**3.70**	**17.07**	**2.62**
	30	CCVPE*	18.14	9.82	60.72	24.52
		Ours	**10.70**	**5.20**	**18.09**	**2.79**

Pitch and Roll Variability

Pitch	Method	↓Localization (m) Mean	↓Localization (m) Median	↓Orientation (°) Mean	↓Orientation (°) Median
±0°	CCVPE*	15.36	4.74	37.72	8.95
	Ours	**9.12**	**2.81**	**13.40**	**2.63**
±10°	CCVPE*	16.98	7.93	49.27	14.58
	Ours	**9.62**	**3.53**	**16.10**	**2.76**

Roll	Method	↓Localization (m) Mean	↓Localization (m) Median	↓Orientation (°) Mean	↓Orientation (°) Median
±0°	CCVPE*	15.36	4.74	37.72	8.95
	Ours	**9.12**	**2.81**	**13.40**	**2.63**
±20°	CCVPE*	17.27	8.19	51.55	15.61
	Ours	**9.18**	**3.22**	**16.16**	**2.63**

Acknowledgements

This material is based upon work supported by the National Geospatial Intelligence Agency under Contract No. HM0476-21-C-0004. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of NGA, DoD, or the US government. Approved for public release, NGA-U-2025-00132.