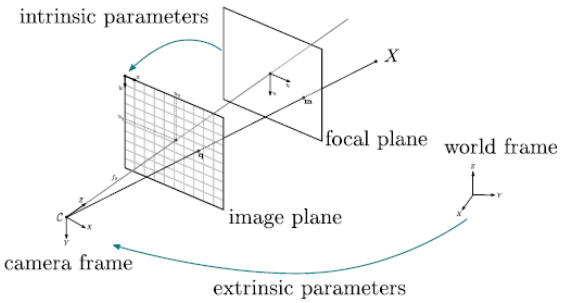
SOFI: Multi-Scale Deformable Transformer for Camera Calibration with Enhanced Line Queries

Sebastian Janampa and Marios Pattichis

Introduction

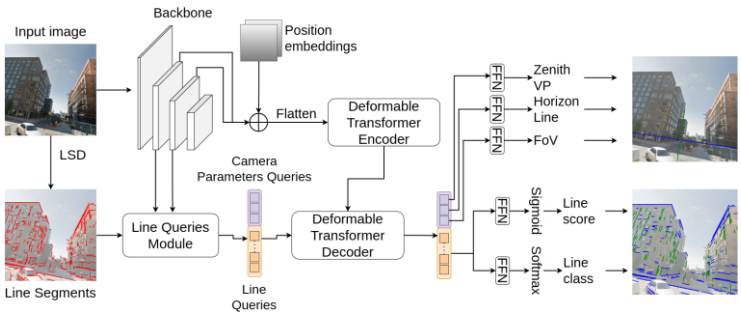
- Camera calibration consists of estimating the internal and external camera parameters.
- Camera calibration enables higher-level applications such as 3D scene tenderization, image rectification, metrology, 3D pose estimation, and depth estimation.
- There are many approaches to estimating the camera parameters from a single image.

Traditional methods	CNN	Transformers
Pros ✓		
• No need to be trained. • Fast performance.	• Promote cross-scale interaction. • Low time and memory complexities	• Promote intra-scale interaction. • Long receptive fields (all the image).
Cons ✗		
• Multiple pictures where the known pattern is visible. • Manhattan assumption for vanishing points.	• Small receptive fields (kernel area). • Complex architectures for integrating the line information into the deep learning model.	• High complexity. • Extract only the geometry information from the line segments.

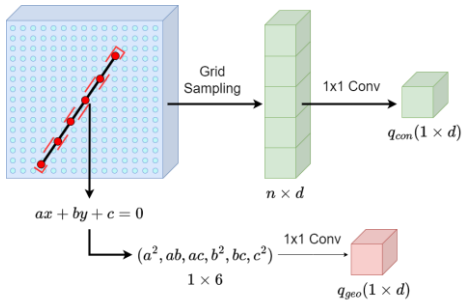


Methodology

SOFI architecture: we use deformable attention to promote cross- and intra-scale interactions



Line Queries Module



Results

Model	Up (°) ↓		Pitch (°) ↓		Roll (°) ↓		FoV (°) ↓	
	Mean	Med.	Mean	Med.	Mean	Med.	Mean	Med.
Google Street View								
Upright	3.05	1.92	2.90	1.80	6.19	0.43	9.47	4.42
DeepHorizon	3.58	3.01	2.76	2.12	1.78	1.67	-	-
Perceptual	2.73	2.13	2.39	1.78	0.96	0.66	4.61	3.89
UprightNet	28.20	26.10	26.56	24.56	6.22	4.33	-	-
GPNet	2.12	1.61	1.92	1.38	0.75	0.47	3.59	2.72
CTRL-C	1.71	1.43	1.52	**1.20**	0.57	0.46	3.38	2.64
MSCC	1.75	**1.42**	1.56	1.24	0.58	0.46	**3.04**	**2.29**
SOFI (ours)	**1.64**	1.44	**1.51**	1.28	**0.54**	**0.43**	3.09	2.79
Holicity								
DeepHorizon*	7.82	3.99	6.10	2.73	3.97	2.67	-	-
Perceptual*	7.37	3.29	6.32	2.86	3.10	1.82	-	-
GPNet*	4.17	**1.73**	**1.46**	**0.74**	1.36	0.95	-	-
CTRL-C	2.66	2.19	2.26	1.78	1.09	0.77	12.41	11.59
MSCC	2.28	1.88	1.87	1.43	**1.08**	**0.81**	13.60	12.20
SOFI (ours)	**2.23**	1.82	1.75	1.31	1.16	0.85	**11.47**	**11.25**

Results of camera calibration parameters on testing datasets. The ⋆ is used to mark models that were trained using the SUN360 dataset.

Model	Google Street View			Horizon Line in the Wild			Holicity		
	@ 0.10	@ 0.15	@ 0.25	@ 0.10	@ 0.15	@ 0.25	@ 0.10	@ 0.15	@ 0.25
DeepHorizon*	-	-	74.25	-	-	45.63	-	-	70.13
Perceptual*	-	-	80.40	-	-	38.29	-	-	70.80
GPNet*	-	-	83.12	-	-	48.90	-	-	81.72
CTRL-C	69.49	78.92	87.16	24.04	33.56	46.37	38.84	55.13	72.31
MSCC	**70.39**	79.59	87.63	24.85	34.44	47.28	49.71	63.60	77.43
SOFI (ours)	70.32	**79.84**	**87.87**	**27.93**	**37.55**	**49.69**	**59.83**	**72.05**	**82.96**

AUC percentages (out of 100%) for horizon line errors on testing datasets. The ⋆ is used to mark models that were trained using the SUN360 dataset. @ 0.10, @ 0.20, and @ 0.25 refer to the area under the curve from zero to Error=0.10, 0.20, and 0.25.

Training dataset: Google Street View.
Testing datasets: Google Street View, Horizon Line in the Wild, and Holicity.
SOFI has a significant improvement in unseen datasets

Model	Google Street View	Horizon Line in the Wild	Holicity
CTRL-C	**22.6**	**19.0**	**25.9**
MSCC	18.0	13.2	18.4
SOFI (ours)	21.6	17.4	23.6

Inference speed comparisons for transformers-based models with batch size of 1. Results are shown in terms of frames per second.

Conclusions & Future Work

SOFI produces state-of-the-art results while keeping a competitive inference speed.

For future work, we will consider a deeper study of the encoder since increasing the number of sampling offsets boosts the performance but requires a longer inference time.