Deep Vanishing Point Detection: Geometric priors make dataset variations vanish

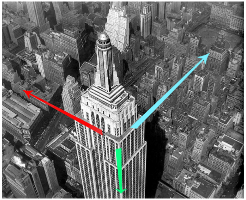
Yancong Lin, Ruben Wiersma, Silvia Laura Pintea, Klaus Hildebrandt, Elmar Eisemann and Jan C. van Gemert

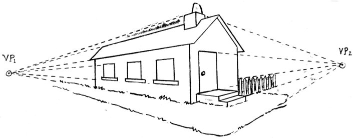
Computer Vision Lab, TUDelft, The Netherlands
Computer Graphics and Visualization, TUDelft, The Netherlands

https://github.com/yanconglin/VanishingPoint_HoughTransform_GaussianSphere/

CVPR, 2022

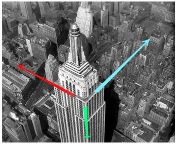
Vanishing Point Detection

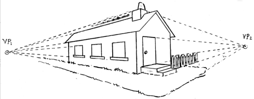




A vanishing point is the intersection of a set of image lines projected from **3D parallel lines**.

Vanishing Point Detection



- SLAM [1];
- Camera calibration [2];
- Scene understanding [3];
- ...

---

[1] A. J. Davison, I. D. Reid, N. D. Molton, and O. Stasse (2007). “MonoSLAM: Real-time single camera SLAM”. In:

Current approaches

- Knowledge-based
  - Two-stage: Line segment detector + clustering [4]
  - ◦ pros: unsupervised or little annotation
  - ◦ cons: outlier/noisy lines, performance
- learning-based
  - Conic convolutions [5]
  - ◦ pros: end2end, intuitive, SOTA results
  - ◦ cons: massive labelled data
  - ◦ cons: Manhattan assumption
  - ◦ cons: vulnerable to domain shifts, hampering generalization

[4] H. Li, J. Zhao, J.-C. Bazin, and Y.-H. Liu (2020). "Quasi-globally Optimal and Near/True Real-time Vanishing Point Estimation in Manhattan World". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence*

[5] Y. Zhou, H. Qi, J. Huang, and Y. Ma (2019). "NeurVPS: neural vanishing point scanning via conic convolution". In: *arXiv preprint arXiv:1910.06316*

Contribution: add two geometric priors into CNNs

- Prior 1: *Hough transform* for identifying lines;
- Prior 2: *Gaussian sphere* for localizing vps on a unit hemisphere;
- an *end-to-end* learning framework;
- improved *data efficiency*;
- superior performance *w/o Manhattan assumption*;
- generalization from synthetic to real-world;

Overview of the proposed model

Adding two priors into end2end learning.

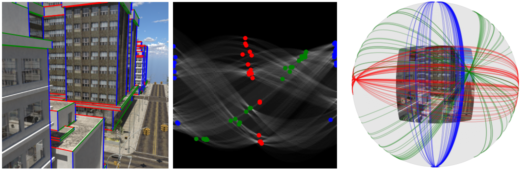
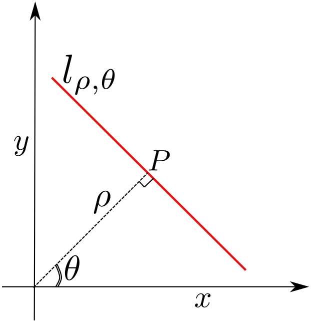


Image domain (i) Hough Transform (ii) Gaussian sphere

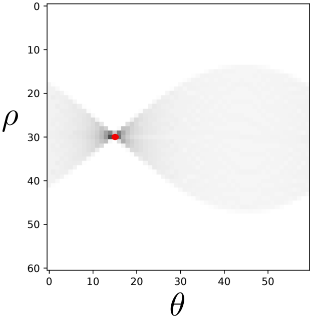
- CNNs for semantic feature learning;
- Hough Transform for line detection;
- Gaussian sphere for VPs detection;

Prior 1: HT (Hough Transform) line priors [6]

- Maps lines to bins in Hough space: a 2D angle-offset histogram;
- Each pixel votes for a number of bins by $\rho = x_i \cos\theta + y_i \sin\theta$;
- Local maxima in Hough space correspond to lines in images



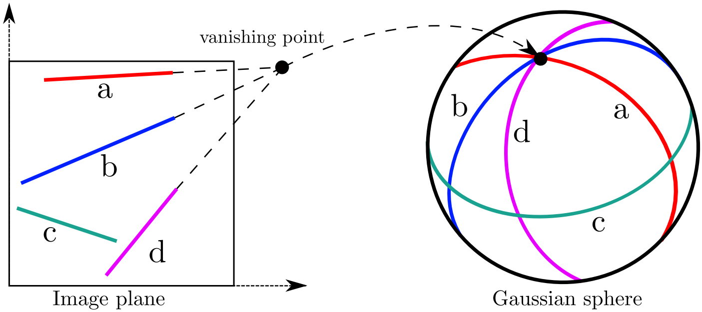
Line



$\mathcal{HT}$ space

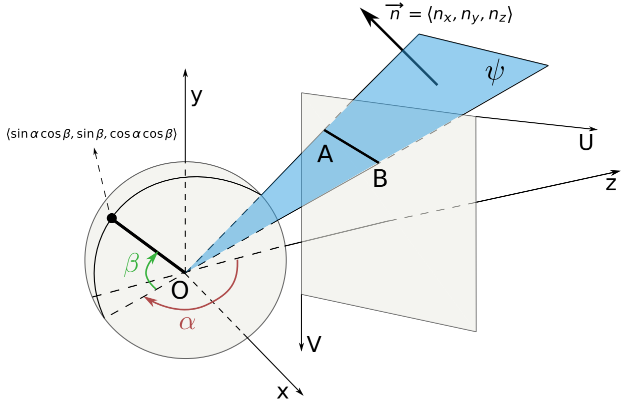
[6] Y. Lin, S. L. Pintea, and J. C. van Gemert (2020). "Deep Hough-Transform Line Priors". In

Prior 2: Gaussian sphere mapping



- VPs are intersections of multiple great circles;
- Spherical convolutions;
- Clustering algorithms to detect multiple unorthoginal VPs.

Prior 2: Gaussian sphere mapping

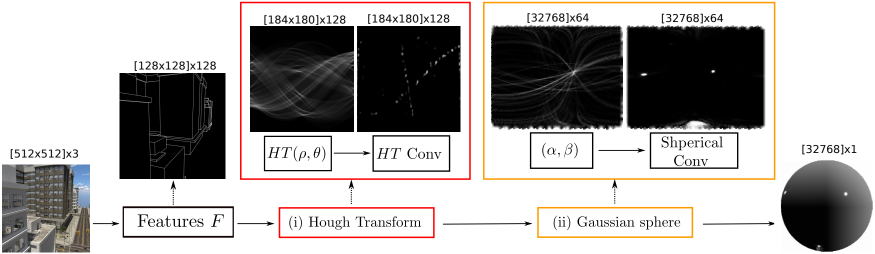


$$\vec{n} = (n_x, n_y, n_z) = \frac{\overrightarrow{OA} \times \overrightarrow{OB}}{\left\| \overrightarrow{OA} \times \overrightarrow{OB} \right\|}. \quad (1)$$

$$\langle \sin\alpha\cos\beta, \sin\beta, \cos\alpha\cos\beta \rangle \langle n_x, n_y, n_z \rangle = 0 \quad (2)$$

Overview of the proposed model

Add two priors into end2end learning.



- CNNs for semantic feature learning;
- Hough Transform for line detection;
- Gaussian sphere for VPs detection;
- Clustering algorithms for detecting unorthoginal VPs (optional).

Experiments: datasets and evaluation

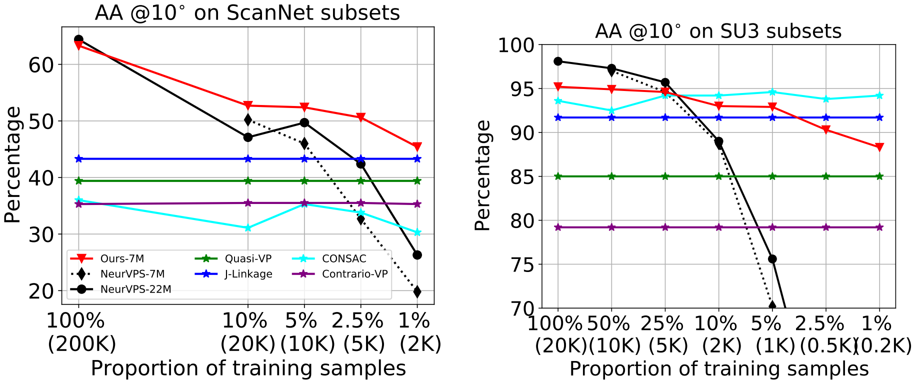
Dataset: Synthetic, Manhattan, non-Manhattan.
Evaluation: angular accuracy (AA) in camera space.
Higher is better.

Visual examples.

Experiments: Improved data efficiency

Comparison with SOTA [7].

Evaluation: angular accuracy (AA), higher is better.

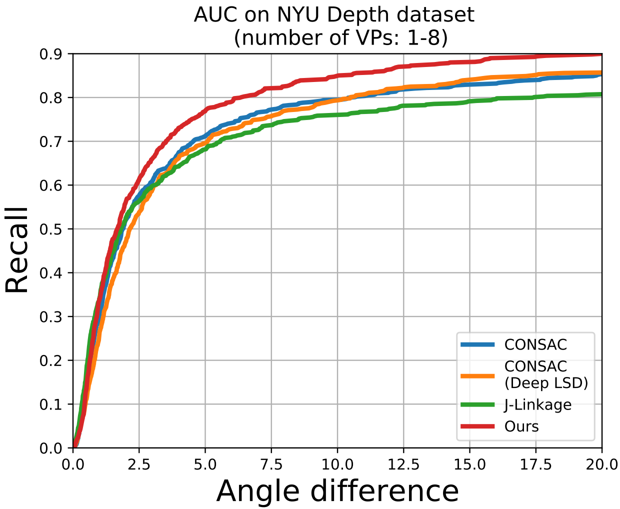


(a) ScanNet (real-world) (b) SU3 (synthetic)

ScanNet: top results on various subsets;

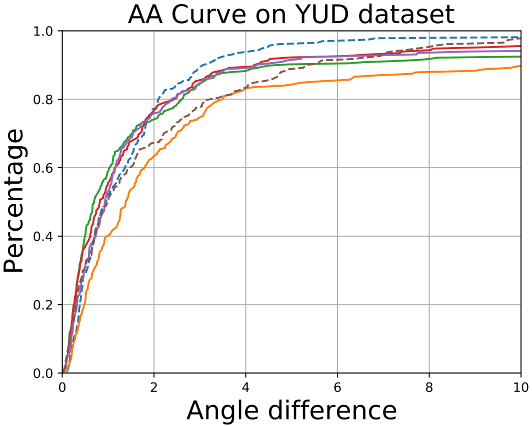
SU3: minor performance decrease when decreasing from 20K to 0.2K.

Experiments: Comparison in *non-Manhattan* world



NYU: consistently outperforms SOTA on detecting multiple unorthogonal VPs.

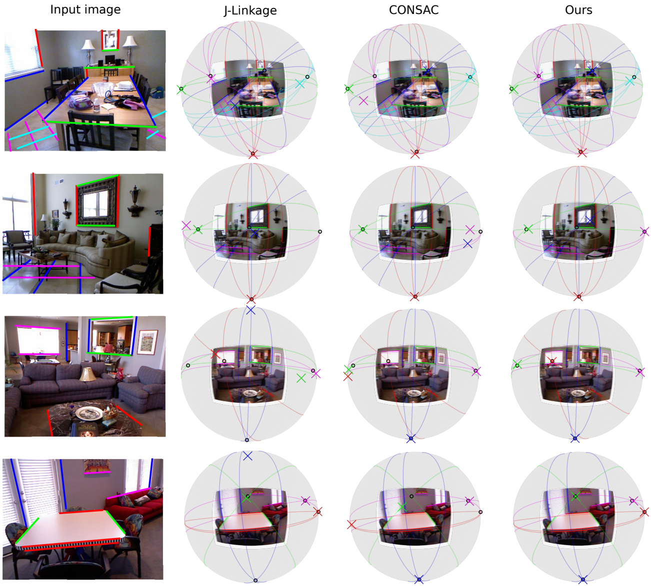
Experiments: Comparison in *Manhattan* world



AA curves on on the ScanNet and YUD.

- On YUD (102 images in total), our model outperforms state-of-the-art *w/o finte-tuning, trained on SU3 (synthetic only)*.

Visualization



Limitations

- Quantization errors (both HT and sphere mapping);
- Dense sampling on the sphere leads to more computation;
- Comparing to classic methods: still need hundreds of training samples.

Conclusions

- Adding two geometric priors:
  Hough Transform and *Gaussian sphere mapping*;
- Improved data efficiency-data efficient learning;
- Applicable in non-Manhattan world;
- Generalizable from synthetic to real-world.

Thank you!

https://github.com/yanconglin/VanishingPoint_HoughTransform_GaussianSphere/