IC 6

Visual Localization with Environment Outline Prior

Vojtěch Pánek

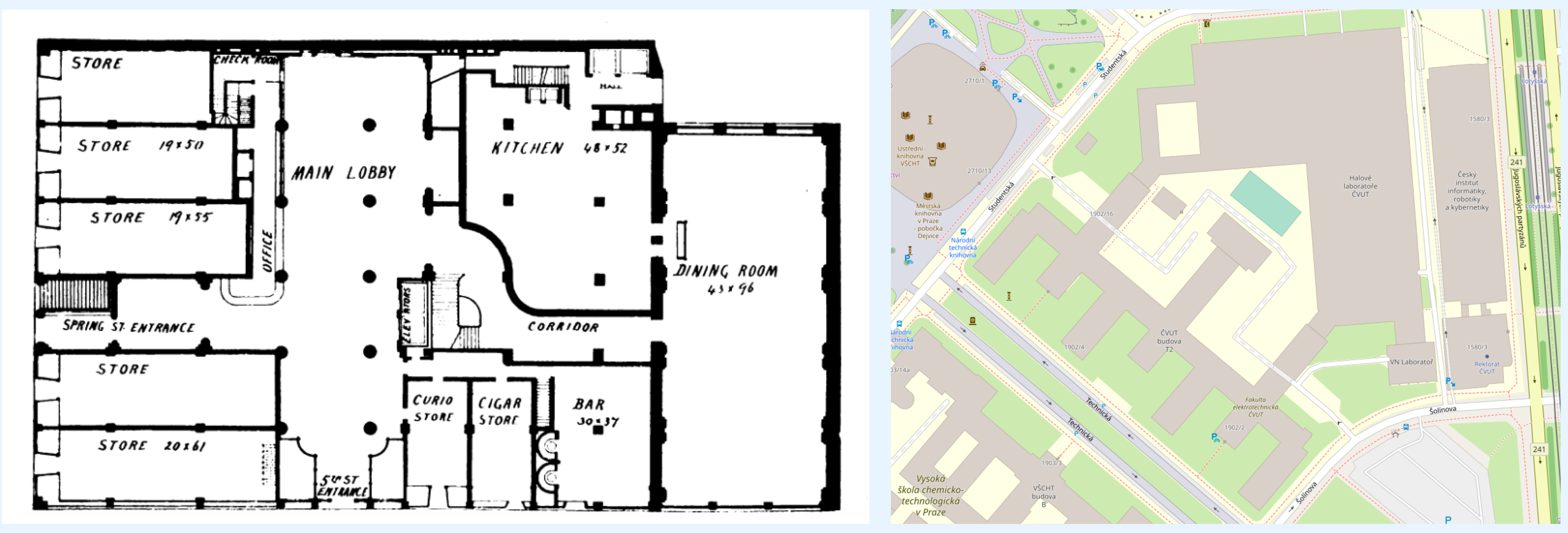
Czech Technical University in Prague, Faculty of Electrical Engineering

CTU CZECH TECHNICAL UNIVERSITY IN PRAGUE

Visual localization task

Visual localization deals with estimation of camera position and orientation within given environment representation (map), based on images coming from the camera.

We want to enable **visual localization on 2D environment layouts** (e.g. floor plans or human readable maps), which are often readily available and have very low memory footprint.

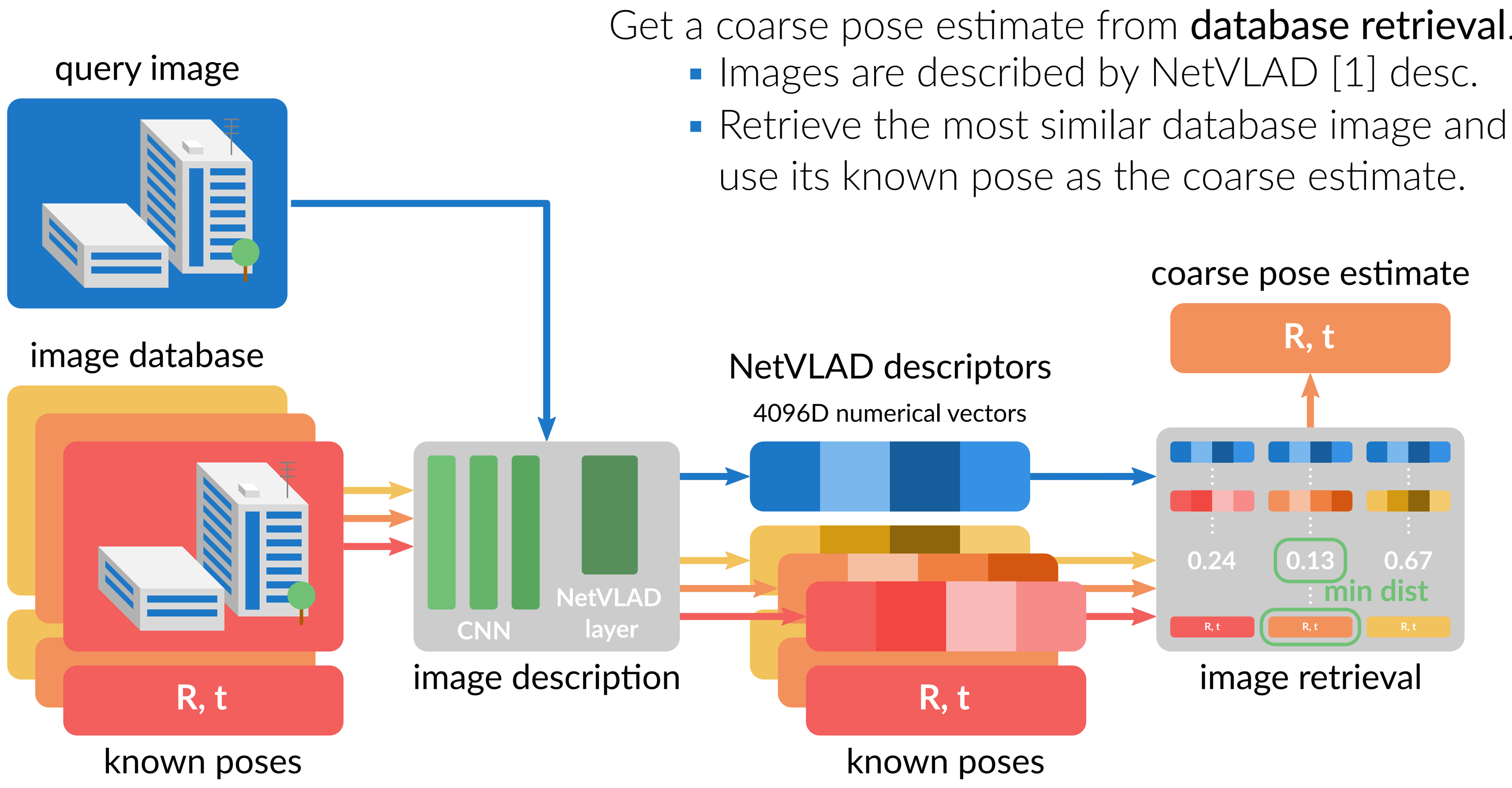


Localization pipeline for environment layouts

Our pipeline is using a **hierarchical localization** approach:

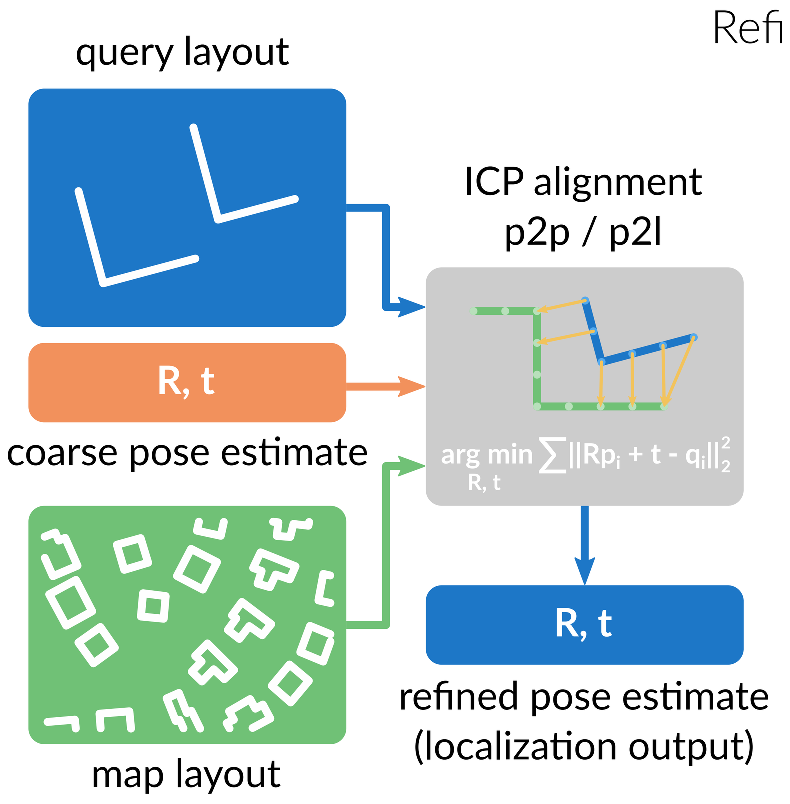
Get a coarse pose estimate from **database retrieval**.

- Images are described by NetVLAD [1] desc.
- Retrieve the most similar database image and use its known pose as the coarse estimate.

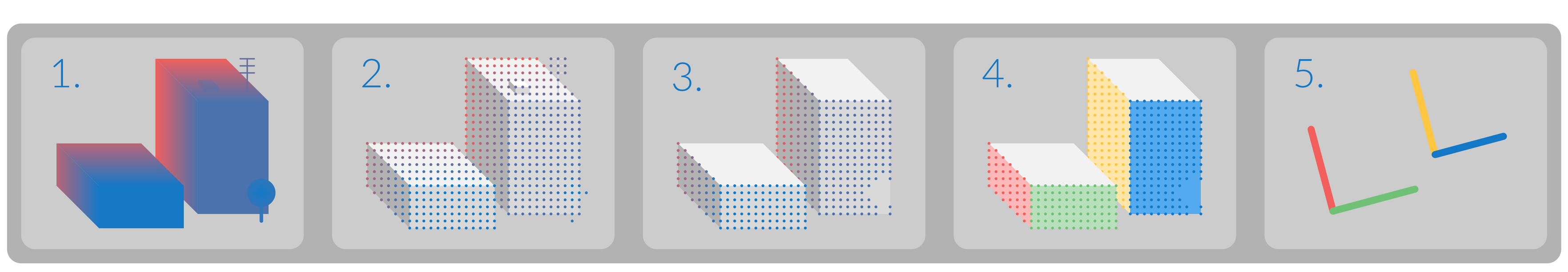


Refine the coarse estimate by **ICP algorithm [2]**.

- Try to align layout from query image on the map layout while keeping track of the applied transformations.
- Use the coarse pose estimate as initial transformation for the ICP.
- **ICP** (**Iterative Closest Point**) algorithm [2]
  1. find the closest map point to each query point (form the closest pairs)
  2. apply optimal transformation (rotation and translation) to minimize the distance between the pairs of the closest points
  3. repeat until convergence
- We tested the standard **point-point** (**p2p**) and also **point-line** (**p2l**) ICP variant.



Layout extraction from images



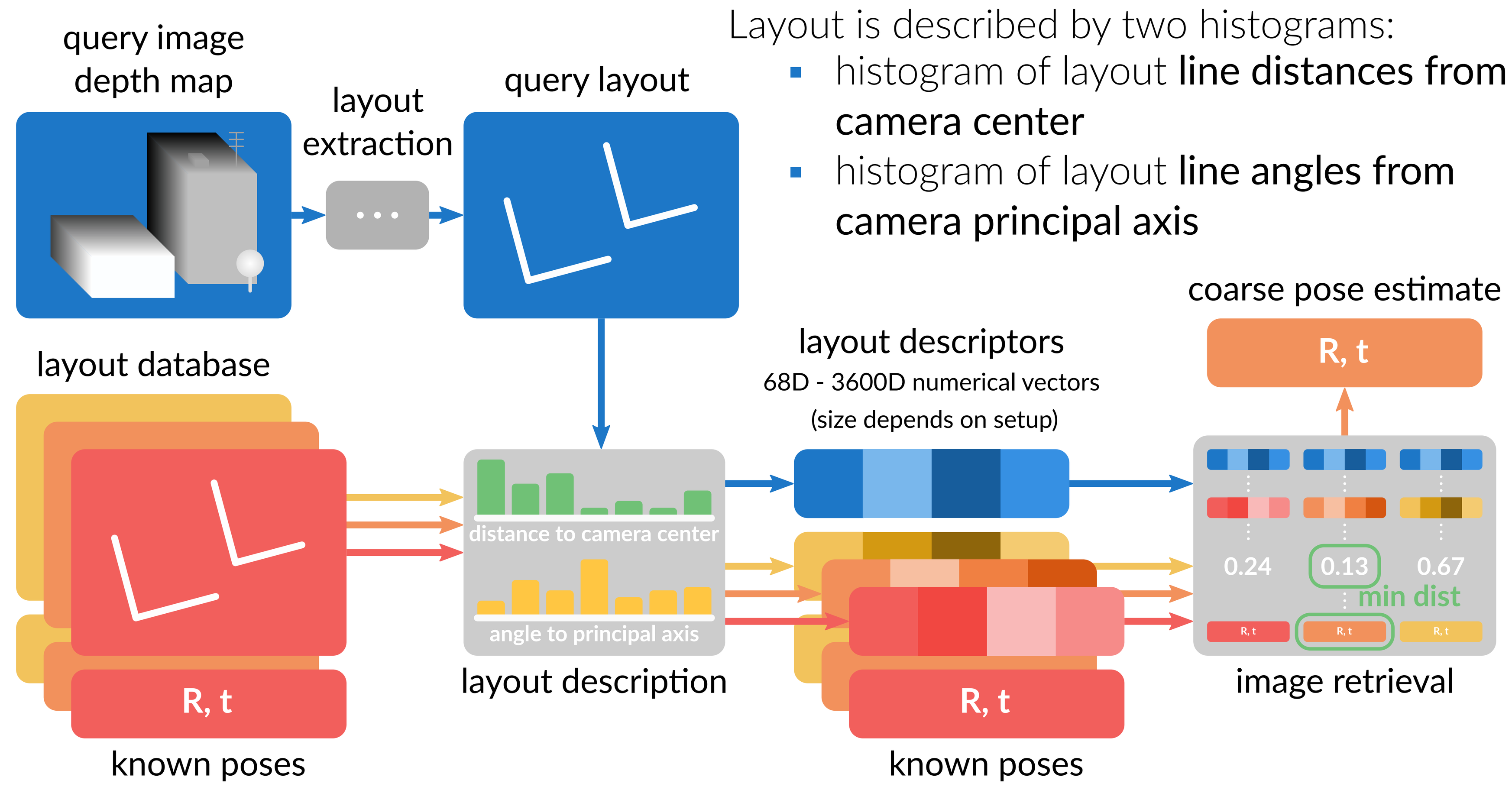
1. Take depth map as input (from RGB-D sensor or neural monocular depth estimator).
2. Reproject the depth map to 3D point cloud (knowing camera calibration).
3. Filter out points belonging to horizontal planes.
4. Use DBSCAN [4] algorithm to divide point cloud to individual plane segments.
5. Intersect the plane segments with a ground plane to get the 2D line layout.

Layout retrieval

We adapted image retrieval paradigm to layouts and developed a numerical layout descriptor, allowing easy NN (Nearest Neighbor) search in local layouts database.

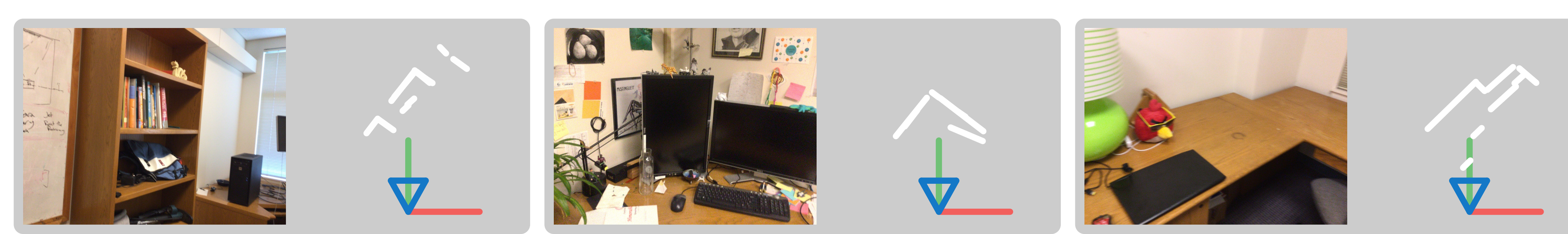
Layout is described by two histograms:

- histogram of layout **line distances from camera center**
- histogram of layout **line angles from camera principal axis**

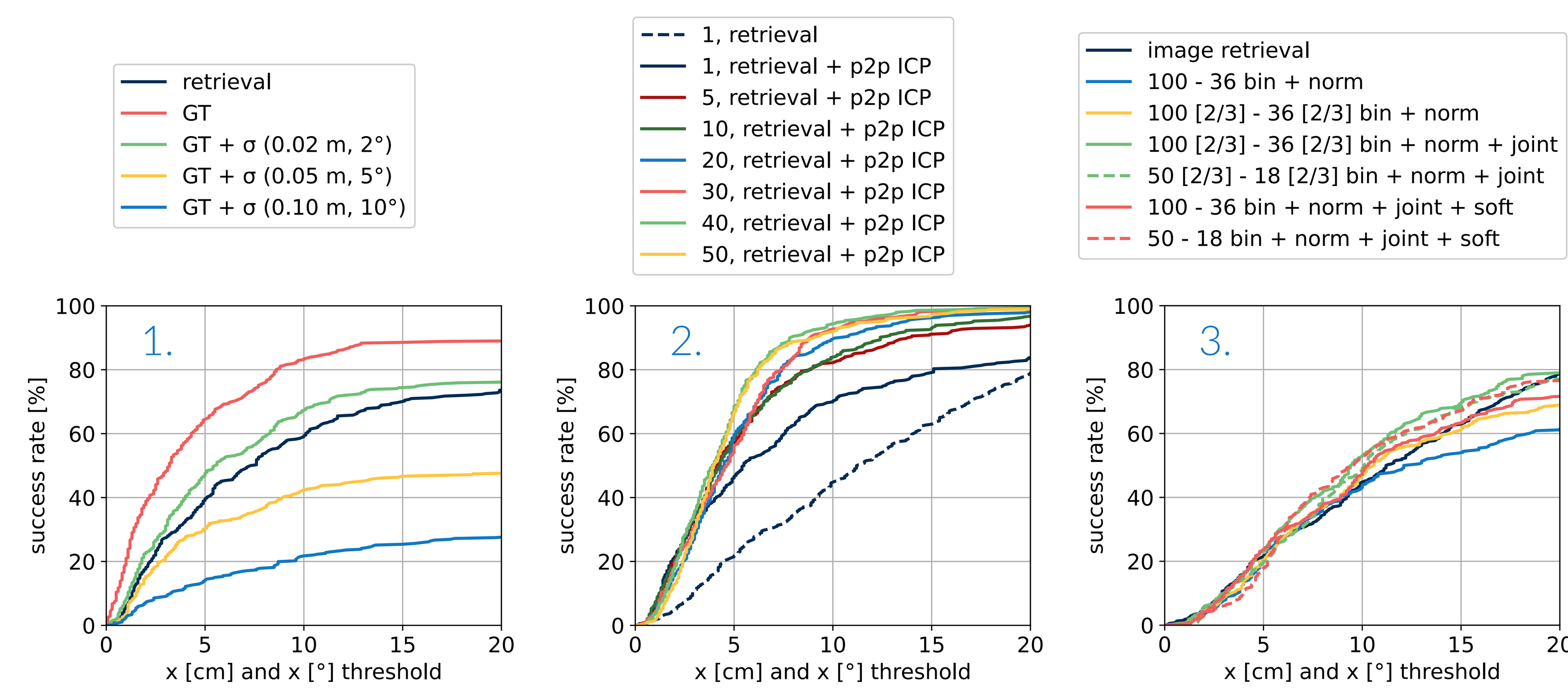


Experiments

Tested on **publicly available datasets** (12 Scenes [6]) for easy comparability.



1. influence of ICP initialization on localization performance
2. artificial FoV (Field of View) expansion by processing image sequences
3. performance of layout retrieval based on descriptor design



* success rate = percentage of images localized within given distance and orientation threshold from the ground truth pose

Acknowledgements

Research was supervised by Dr. rer. nat. Torsten Sattler, CIIRC CTU in Prague.

The work was supported by the Grant Agency of the Czech Technical University in Prague, grant No. SGS21/119/OHK3/2T/13.

References

[1] Relja Arandjelović, Petr Gronát, Akihiko Torii, Tomás Pajdla, and Josef Sivic. NetVLAD: CNN Architecture for Weakly Supervised Place Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40:1437–1451, 2018.

[2] K. S. Arun, T. S. Huang, and Steven D. Blostein. Least-Squares Fitting of Two 3-D Point Sets. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-9:698–700, 1987.

[3] Ainaz Eftekhar, Alexander Sax, Roman Bachmann, Jitendra Malik, and Amir Roshan Zamir. Omnidata: A Scalable Pipeline for Making Multi-Task Mid-Level Vision Datasets from 3D Scans. *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 10766–10776, 2021.

[4] Martin Ester, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In *KDD*, 1996.

[5] René Ranftl, Katrin Lasinger, David Hafner, Konrad Schindler, and Vladlen Koltun. Towards Robust Monocular Depth Estimation: Mixing Datasets for Zero-shot Cross-dataset Transfer. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2020.

[6] Julien P. C. Valentin, Angela Dai, Matthias Nießner, Pushmeet Kohli, Philip H. S. Torr, Shahram Izadi, and Cem Keskin. Learning to Navigate the Energy Landscape. *2016 Fourth International Conference on 3D Vision (3DV)*, pages 323–332, 2016.