Exploration of Structure-Aware Line Segment Classification

Deep Learning HS22 Project Proposal

Yifan Yu Shinjeong Kim Alexis Tabin Marco Voegeli

1. Introduction

In computer vision, line segments are useful in a wide range of tasks such as SLAM (Zuo et al., 2017; Pumarola et al., 2017; Gomez-Ojeda et al., 2019), pose estimation (Xu et al., 2017; Gao et al., 2021), and 3D reconstruction (Hofer et al., 2017; Zhou et al., 2019). Compared to feature points, line segments provide more information on the high-level structural layout of the scene, and are more robust to viewpoint and illumination changes.

Conceptually, structural lines (e.g. edges of walls, ceilings, floor) provide more structural cues than repetitive textural lines (e.g. grid lines of floor tiles) in man-made scenes. Therefore it could be beneficial for downstream tasks, especially reconstruction and scene understanding if such classification of line segments is recovered. Existing line segment detectors work well in recovering 2D line segments from images. They however do not provide information on whether a line segment is more structural or textural. In this project, we aim to novelistically design a structure-aware line segment classifier using neural networks, to supplement line segment detections with this information. As there is no prior work on this task, we are also interested in exploring ways to create/rework datasets for this task, as well as the effect of having such structural classification online-based downstream applications.

2. Related Work

Line Detection. Conventional line segment detectors such as LSD (Von Gioi et al., 2008) and EDLines (Akinlar & Topal, 2011) rely on grouping image gradients. With more computational cost, deep learning-based detectors offer better results, in particular for specialized tasks like wireframe parsing (Huang et al., 2018; Zhang et al., 2019; Xue et al., 2020; Huang et al., 2020; Pautrat et al., 2021; Yoon & Kim, 2021). We plan to make our classification detector-independent, while the classification could also be used to evaluate the characteristics of the detection results of different detectors.

Image Datasets. Wireframe (Huang et al., 2018) and York Urban (Denis et al., 2008) are widely-used datasets for line segment detection and wireframe parsing. However, they do not carry information about the nature of the detected lines, nor do they provide necessary scene information that could be used to retrieve such information, such as depth or 3D geometry. For our proposed task, RGB-D or synthetic image datasets enable us to estimate scene structure with depth and/or geometry information.

3. Method and Evaluation

Data Preparation. For the scope of this project we will focus on indoor scenes as image datasets with depth and/or scene geometry are mainly indoor datasets. We will start with Hypersim (Roberts et al., 2021), which is a photorealistic synthetic dataset for holistic indoor scene understanding. It contains accurate per-pixel ground-truth depth and scene geometry, which we will use to generate per-pixel data (e.g. depth variance, surface normal discontinuities) to train our classifier.

Method. First, we plan to design and train a CNN-based pixel-wise binary classifier that takes as input the raw image, and outputs pixel-wise ”structure score” which will be formulated based on preprocessed quantities such as depth variance and/or surface normal discontinuities from the training data. Such score will be subsequently averaged along each detected line segment by some existing detector to give a classification of line segments. Then, we will try to come up with a more sophisticated network to combine the two steps: a network that directly takes the image as well as line segment(s) as input and outputs a classification score per segment.

Goal and Evaluation Our fundamental goal (80%) is to build the pixel-wise binary classifier and evaluate it quantitatively (the regression result of per-pixel structural information such as depth variance/surface normal discontinuity) and qualitatively (by observing to what extent the results align with the scene structures). Built upon this model, the combined one-step classifier will be compared to this baseline (100%). The effect of our classification on line-based downstream applications (reconstruction, pose estimation, etc.) could be further tested (120%). We will also study different parameter settings for the models. Finally, if possible, we could propose a new deep learning-based line segment detector based on existing method(s) integrated with our classification target, which provides structural information simultaneously with the detected segments (200%).

References

Akinlar, C. and Topal, C. Edlines: Real-time line segment detection by edge drawing (ed). In *IEEE International Conference on Image Processing*, 2011.

Denis, P., Elder, J. H., and Estrada, F. J. Efficient edge-based methods for estimating manhattan frames in urban imagery. In *European Conference on Computer Vision (ECCV)*, 2008.

Gao, S., Wan, J., Ping, Y., Zhang, X., Dong, S., Li, J., and Guo, Y. Pose refinement with joint optimization of visual points and lines. *arXiv preprint arXiv:2110.03940*, 2021.

Gomez-Ojeda, R., Moreno, F.-A., Zuñiga-Noël, D., Scaramuzza, D., and Gonzalez-Jimenez, J. PL-SLAM: A stereo SLAM system through the combination of points and line segments. *IEEE Transactions on Robotics*, 35(3):734–746, 2019.

Hofer, M., Maurer, M., and Bischof, H. Efficient 3D scene abstraction using line segments. *Computer Vision and Image Understanding (CVIU)*, 157:167–178, 2017.

Huang, K., Wang, Y., Zhou, Z., Ding, T., Gao, S., and Ma, Y. Learning to parse wireframes in images of man-made environments. In *Computer Vision and Pattern Recognition (CVPR)*, June 2018.

Huang, S., Qin, F., Xiong, P., Ding, N., He, Y., and Liu, X. Tp-lsd: Tri-points based line segment detector. In *European Conference on Computer Vision (ECCV)*, 2020.

Pautrat, R., Lin, J.-T., Larsson, V., Oswald, M. R., and Pollefeys, M. Sold2: Self-supervised occlusion-aware line description and detection. In *Computer Vision and Pattern Recognition (CVPR)*, 2021.

Pumarola, A., Vakhitov, A., Agudo, A., Sanfeliu, A., and Moreno-Noguer, F. PL-SLAM: Real-time monocular visual SLAM with points and lines. In *International Conference on Robotics and Automation (ICRA)*, 2017.

Roberts, M., Ramapuram, J., Ranjan, A., Kumar, A., Bautista, M. A., Paczan, N., Webb, R., and Susskind, J. M. Hypersim: A photorealistic synthetic dataset for holistic indoor scene understanding. In *International Conference on Computer Vision (ICCV)*, 2021.

Von Gioi, R. G., Jakubowicz, J., Morel, J.-M., and Randall, G. Lsd: A fast line segment detector with a false detection control. *IEEE Trans. Pattern Analysis and Machine Intelligence (PAMI)*, 32(4):722–732, 2008.

Xu, C., Zhang, L., Cheng, L., and Koch, R. Pose estimation from line correspondences: A complete analysis and a series of solutions. *IEEE Trans. Pattern Analysis and Machine Intelligence (PAMI)*, 39(6), 2017.

Xue, N., Wu, T., Bai, S., Wang, F., Xia, G.-S., Zhang, L., and Torr, P. H. Holistically-attracted wireframe parsing. In *Computer Vision and Pattern Recognition (CVPR)*, 2020.

Yoon, S. and Kim, A. Line as a visual sentence: Context-aware line descriptor for visual localization. *IEEE Robotics and Automation Letters (RAL)*, 6(4):8726–8733, 2021.

Zhang, Z., Li, Z., Bi, N., Zheng, J., Wang, J., Huang, K., Luo, W., Xu, Y., and Gao, S. PPGNet: Learning point-pair graph for line segment detection. In *Computer Vision and Pattern Recognition (CVPR)*, 2019.

Zhou, Y., Qi, H., Zhai, Y., Sun, Q., Chen, Z., Wei, L.-Y., and Ma, Y. Learning to reconstruct 3D manhattan wireframes from a single image. In *ICCV*, 2019.

Zuo, X., Xie, X., Liu, Y., and Huang, G. Robust visual slam with point and line features. In *International Conference on Intelligent Robots and Systems (IROS)*, 2017.