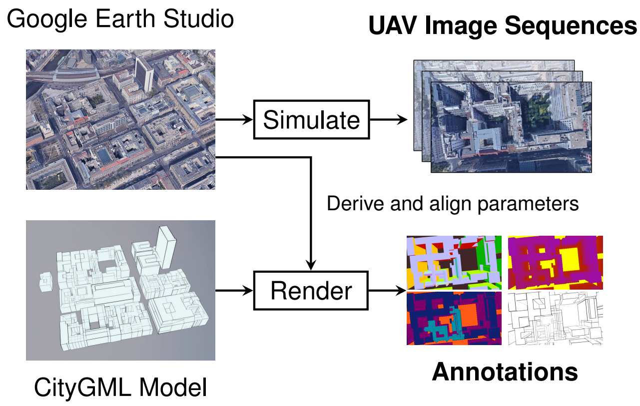
Chair of Remote Sensing Technology
Department of Aerospace and Geodesy
Technical University of Munich

Derivation of Geometrically and Semantically Annotated UAV Datasets at Large Scales from 3D City Models

Sidi Wu, Lukas Liebel, Marco Körner

While in high demand for the development of deep learning approaches, extensive datasets of annotated unmanned aerial vehicle (UAV) imagery are still scarce today. Manual annotation, however, is time-consuming and, thus, has limited the potential for creating large-scale datasets. We tackle this challenge by presenting a procedure for the automatic creation of simulated UAV image sequences in urban areas and pixel-level annotations from publicly available data sources. We synthesize photo-realistic UAV imagery from Google Earth Studio and derive annotations from an open CityGML model that not only provides geometric but also semantic information.



Code and Dataset: `https://github.com/sian1995/largescaleuavdataset`

Motivation

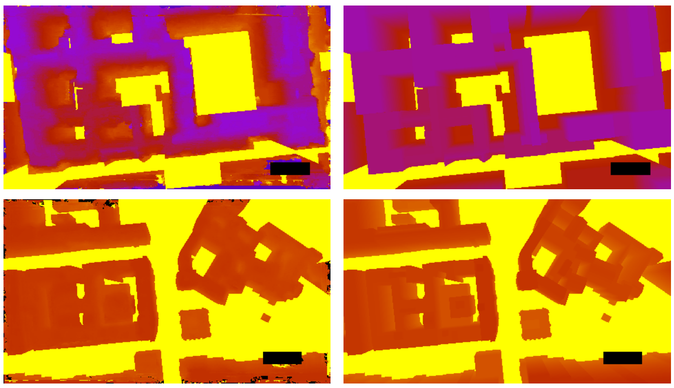
Three major tasks in UAV applications are image understanding, multi-view 3D reconstruction and geo-localization. However, current UAV datasets like VisDrone [3], Semantic Drone Dataset [2], Urban Drone Dataset [1] have limited volumes, and none of them provides large-scale annotations. We tackle this challenge by creating a large-scale UAV dataset, enriched with automatically generated annotations.

Approach

We synthesize **photo-realistic UAV imagery from Google Earth Studio** under different camera perspectives and along different trajectories and align the extrinsic parameters of the virtual camera to a CityGML model horizontally, vertically and in terms of its rotation. This allows us to generate ground truth annotations from the CityGML model under the aligned camera perspectives using Blender. We not only generate **geometric annotations (depth maps, surface normals, edge maps)** but also **semantic labels** by assigning classes to surfaces in the semantic 3D model.

Dataset

Our final dataset contains **144000 annotated UAV images** together with the extrinsic parameters of the virtual camera. We evaluated the geometric accuracy of the simulated dataset with depth maps obtained by applying Structure from Motion (SfM). Qualitatively, our rendered depth maps compare very well to their SfM-derived counterparts, and even overcome the weakness of SfM in areas of no texture, repetitive patterns and thin structures.



Depth from SfM — Ours (from city model)

References

[1] Y. Chen, Y. Wang, P. Lu, Y. Chen, and G. Wang, "Large-scale structure from motion with semantic constraints of aerial images," in *Chinese Conference on Pattern Recognition and Computer Vision (PRCV)*, 2018, pp. 347–359.

[2] *Semantic drone dataset*, Graz University of Technology. [Online]. Available: `https://dronedataset.icg.tugraz.at` (visited on 03/04/2020).

[3] P. Zhu, L. Wen, X. Bian, H. Ling, and Q. Hu, "Vision meets drones: A challenge," 2018. arXiv: `1804.07437 [cs.CV]`.

Contact: Sidi Wu (`sidiwu@ethz.ch`), Lukas Liebel (`lukas.liebel@tum.de`), Marco Körner (`marco.koerner@tum.de`)