Computer Vision Course Project

KITTI Benchmarks for Self-Driving Cars Related Problems

Introduction to Visual Odometry Problem

The odometry benchmark consists of 22 stereo sequences, saved in loss less png format:

1. 11 sequences (00-10) with ground truth trajectories for training
2. 11 sequences (11-21) without ground truth for evaluation.

For this benchmark you may provide results using monocular or stereo visual odometry, laser-based SLAM or algorithms that combine visual and LIDAR information. The only restriction we impose is that your method is fully automatic (e.g., no manual loop-closure tagging is allowed) and that the same parameter set is used for all sequences. A development kit provides details about the data format.

- Download odometry data set (grayscale, 22 GB)
- Download odometry data set (color, 65 GB)
- Download odometry data set (velodyne laser data, 80 GB)
- Download odometry data set (calibration files, 1 MB)
- Download odometry ground truth poses (4 MB)
- Download odometry development kit (1 MB)
- Lee Clement and his group (University of Toronto) have written some python tools for loading and parsing the KITTI raw and odometry datasets

From all test sequences, our evaluation computes **translational and rotational errors** for all possible subsequences of length (100,...,800) meters. The evaluation table below ranks methods according to the average of those values, where errors are measured in percent (for translation) and in degrees per meter (for rotation). A more detailed comparison for different trajectory lengths and driving speeds can be found in the plots underneath.

Note: On 03.10.2013 we have changed the evaluated sequence lengths from (5,10,50,100,...,400) to (100,200,...,800) due to the fact that the GPS/OXTS ground truth error for very small sub-sequences was large and hence biased the evaluation results. Now the averages below take into account longer sequences and provide a better indication of the true performance. Please consider reporting these number for all future submissions. The last leaderboard right before the changes can be found here!

Introduction to 3D Object Detection Problem

The object detection and object orientation estimation benchmark consists of 7481 training images and 7518 test images, comprising a total of 80.256 labeled objects. All images are color and saved as png. For evaluation, we compute precision-recall curves for object detection and orientation-similarity-recall curves for joint object detection and orientation estimation. In the latter case not only the object 2D bounding box has to be located correctly, but also the orientation estimate in bird's eye view is evaluated. To rank the methods we compute average precision and average orientation similarity. We require that all methods use the same parameter set for all test pairs. Our development kit provides details about the data format as well as MATLAB / C++ utility functions for reading and writing the label files.

- Download left color images of object data set (12 GB)
- Download right color images, if you want to use stereo information (12 GB)
- Download the 3 temporally preceding frames (left color) (36 GB)
- Download the 3 temporally preceding frames (right color) (36 GB)
- Download Velodyne point clouds, if you want to use laser information (29 GB)
- Download camera calibration matrices of object data set (16 MB)
- Download training labels of object data set (5 MB)
- Download object development kit (1 MB) (including 3D object detection and bird's eye view evaluation code)
- Download pre-trained LSVM baseline models (5 MB) used in Joint 3D Estimation of Objects and Scene Layout (NIPS 2011). These models are referred to as LSVM-MDPM-sv (supervised version) and LSVM-MDPM-us (unsupervised version) in the tables below.
- Download reference detections (L-SVM) for training and test set (800 MB)
- Qianli Liao (NYU) has put together code to convert from KITTI to PASCAL VOC file format (documentation included, requires Emacs).
- Karl Rosaen (U.Mich) has released code to convert between KITTI, KITTI tracking, Pascal VOC, Udacity, CrowdAI and AUTTI formats.

We evaluate object detection performance using the PASCAL criteria and object detection and orientation estimation performance using the measure discussed in our CVPR 2012 publication. For **cars** we require an **overlap of 70%**, while for pedestrians and cyclists we require an overlap of 50% for a detection. Detections in don't care areas or detections which are smaller than the minimum size do not count as false positive. Difficulties are defined as follows:

- **Easy:** Min. bounding box height: 40 Px, Max. occlusion level: Fully visible, Max. truncation: 15 %
- **Moderate:** Min. bounding box height: 25 Px, Max. occlusion level: Partly occluded, Max. truncation: 30 %

- **Hard:** Min. bounding box height: 25 Px, Max. occlusion level: Difficult to see, Max. truncation: 50 %

All methods are ranked based on the moderately difficult results. Note that for the hard evaluation ~2 % of the provided bounding boxes have not been recognized by humans, thereby upper bounding recall at 98 %. Hence, the hard evaluation is only given for reference.

Note 1: On 25.04.2017, we have fixed a bug in the object detection evaluation script. As of now, the submitted detections are filtered based on the min. bounding box height for the respective category which we have been done before only for the ground truth detections, thus leading to false positives for the category "Easy" when bounding boxes of height 25-39 Px were submitted (and to false positives for all categories if bounding boxes smaller than 25 Px were submitted). We like to thank Amy Wu, Matt Wilder, Pekka Jänis and Philippe Vandermersch for their feedback. The last leaderboards right before the changes can be found here!

Note 2: On 08.10.2019, we have followed the suggestions of the Mapillary team in their paper Disentangling Monocular 3D Object Detection and use 40 recall positions instead of the 11 recall positions proposed in the original Pascal VOC benchmark. This results in a more fair comparison of the results, please check their paper. The last leaderboards right before this change can be found here: Object Detection Evaluation, 3D Object Detection Evaluation, Bird's Eye View Evaluation.

Project Requirements

- Each team will select only one problem to work on.
- Each team is required to read and summarize at least two papers.
- To understand the problem and the results well your team must find some working implementation and run it.
- To confirm your great understanding of the problem and the available frameworks every team must convert a code from TensorFlow to PyTorch or vice versa.
    - In some cases, it will be easier for you to write from scratch except for the data loading and pre-processing.
- You must work in a group of 2-3 students

Deliverables

You are required to deliver the following:

- Your code (40%).
- Summarization of two papers.
  - Papers should be from the list of participants on KITTI or published in CVPR, ICCV, ECCV or ICRA are highly recommended(20%).
- Group Presentation (20%).
  - Overview, Method, Results, figures, analysis, failure cases, success cases, …
- Group Report (20%). More details of the presentation. Include your trials the failed and successful ones. Your decisions, …
- Group Discussion (10%).
- Individual Reports (10%).
- Future direction, criticizing and additional solutions in the project

5 References

- http://www.cvlibs.net/datasets/kitti/index.php
- http://www.cvlibs.net/datasets/kitti/eval_odometry.php
- http://www.cvlibs.net/datasets/kitti/eval_object.php?obj_benchmark