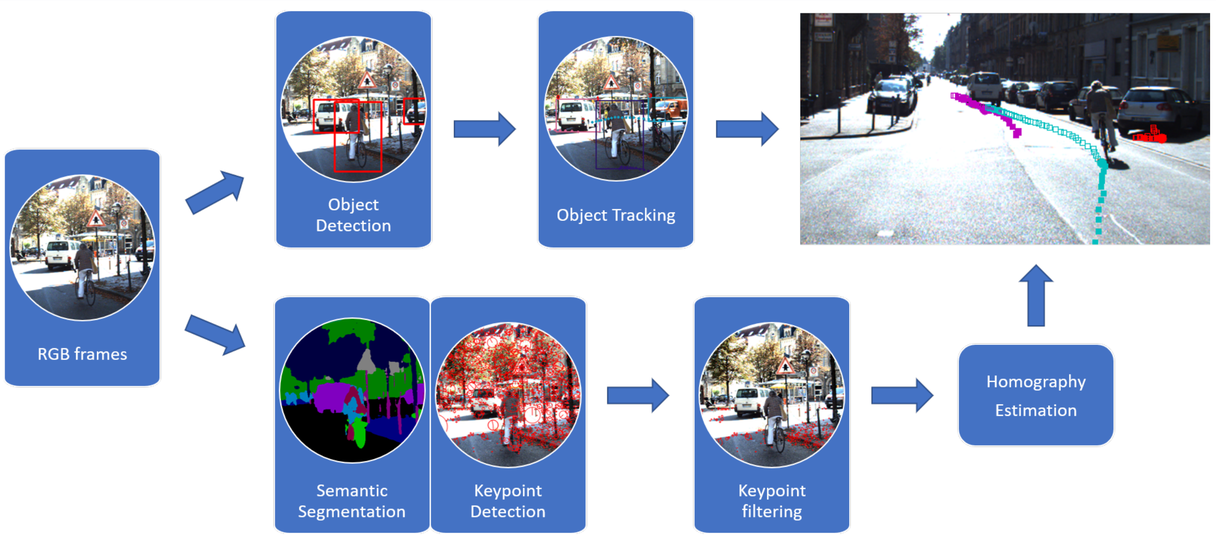
Vehicle Trajectories from Unlabeled Data through Iterative Plane Registration

Federico Becattini, Lorenzo Seidenari, Lorenzo Berlincioni, Leonardo Galteri, and Alberto Del Bimbo
MICC, University of Florence, Italy
{name.surname}@unifi.it

Introduction

- Trajectory prediction is a vital task to ensure safety in automotive.
- Data collection is a slow and expensive process.
- We automatically collect vehicle trajectories from videos without annotations.
- Iterative Plane Registration: meta-algorithm to obtain holistic trajectories (past-present-future) paired with a context.
- Based on ground plane tracking to map object positions across frames.

Iterative Plane Registration

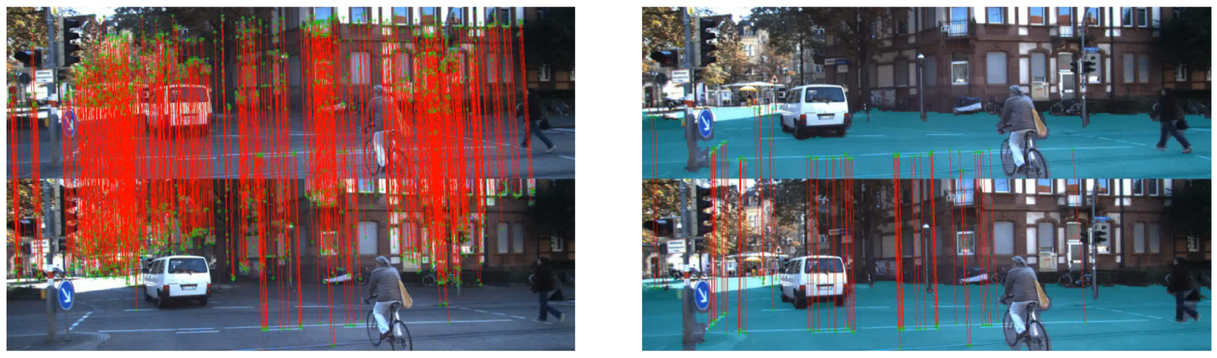


- Meta-algorithm to track the ground plane in a video, obtaining a series of homographies.
- Objects are detected and tracked.
- The ground plane is segmented and tracked by estimating homographies using RANSAC [5].
- Trajectories are obtained by mapping object positions across frames.

Object Detection and Tracking

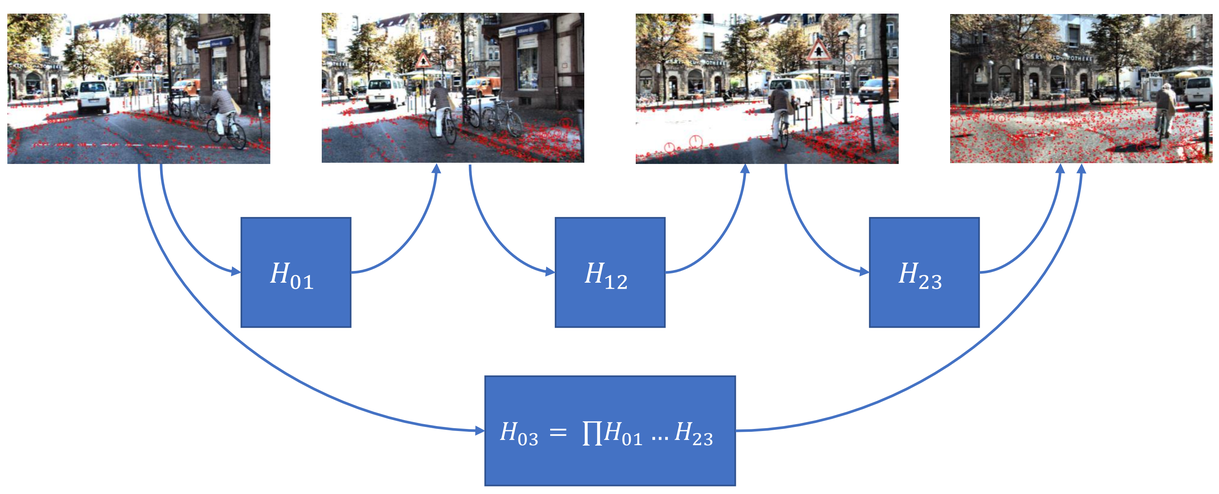
- Objects are detected using Mask-RCNN [1].
- Objects are tracked using a multiple target tracker based on bounding-box association [2].
- We detect and track objects belonging to the following classes: *car*, *person*, *bicycle*, *motorbike*, *truck*, *train*.

Ground Tracking



- The image is segmented using DeepLab v3+ [3] to identify pixels belonging to the ground plane.
- Local keypoints are extracted from the ground plane region. SIFT [4] is used as local descriptor.
- Thanks to the semantic segmentation other objects or buildings are not considered.

Homography Estimation and Trajectory projection



- By matching road keypoints across frames with RANSAC [5], a set of homographies is obtained to map points across adjacent frames.
- The ground is assumed locally planar.
- Homographies are chained together to be able to map back and forth points on the ground across the whole video.
- The position of each object is mapped onto the current frame to obtain a holistic trajectory (past-present-future).
- Combining homographies may lead to instabilities. We ensure the validity of a transformation by checking its determinant [6]: $\det(\mathbf{H}) > 0$.

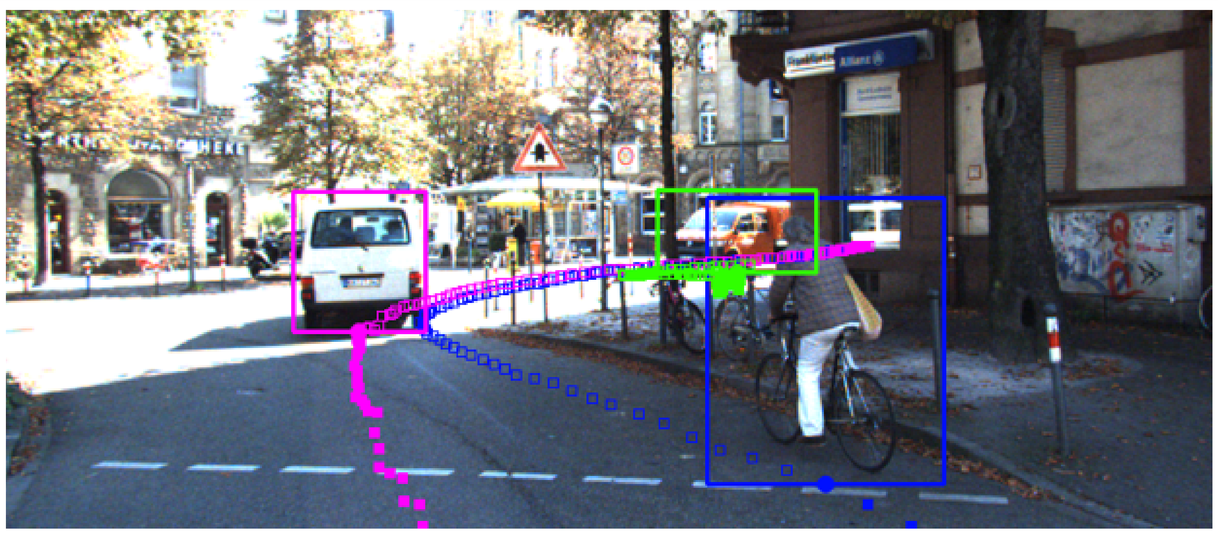
Iterative Plane Registration Algorithm

Algorithm 1 Iterative Plane Registration

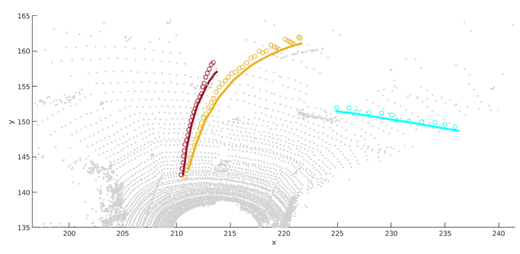
Input: RGB video sequence $F_{t_i}, t_i \in [t_0, t_{end}]$
Output: Homography set

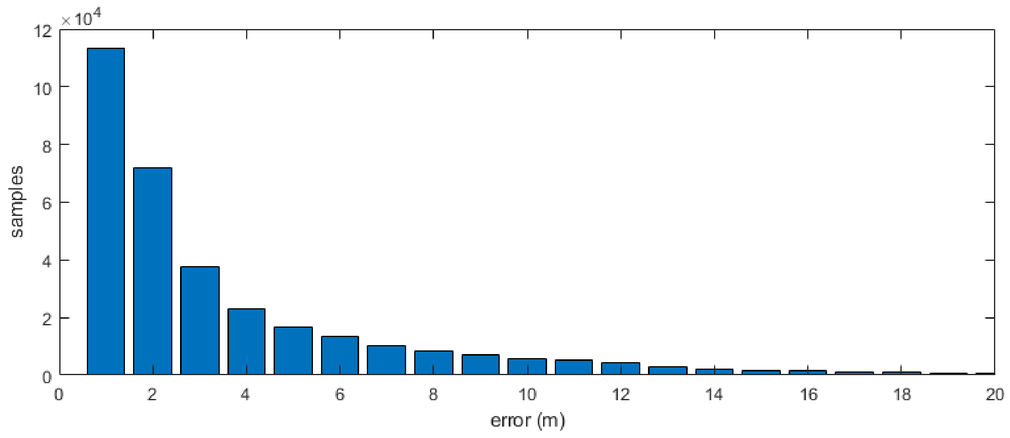
1: Initial timestep t_0.
2: **while** $t_i < t_{end}$ **do**
3: Apply semantic segmentation algorithm (e.g. DeepLab [3]) to frame F_{t_i}, obtaining a pixel-wise labeling $S^c_{t_i}, c \in \{\texttt{'road'}, \texttt{'car'}, \texttt{'sidewalk'} \dots\}$.
4: Extract local keypoints L_{t_i} (e.g. SIFT [4]) from F_{t_i}.
5: Discard keypoints not laying on the ground plane based on the semantic segmentation: $L'_{t_i} = \{k \in L_{t_i} \ \texttt{s.t.} \ S_{t_i}[k_x, k_y] \in \{\texttt{'road'}, \texttt{'sidewalk'}\}\}$
6: Estimate homography to map the ground between frames $F_{t_{i-1}}$ and F_{t_i}: $\mathbf{H}_{t_{i-1}t_i} = \texttt{RANSAC}(L'_{t_{i-1}}, L'_{t_i})$
7: $t_i = t_{i+1}$;
8: **end while**
9: **return** $\{\mathbf{H}_{t_i}\}$

Holistic Trajectories



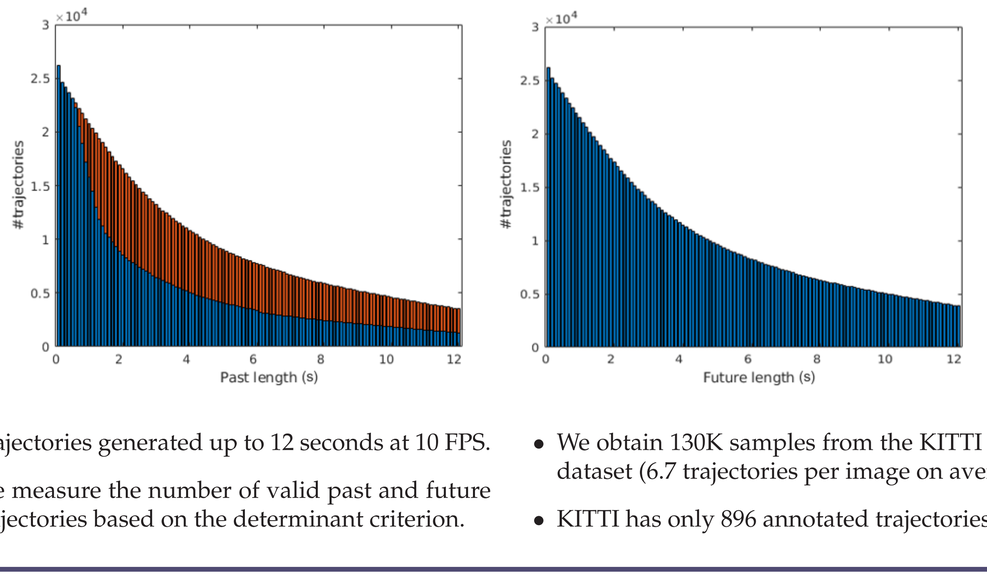
Experimental evaluation





- We evaluate the Iterative Plane Registration algorithm on the KITTI dataset [7].
- Trajectories are generated for all annotated tracks in the dataset and compared with the ground truth trajectories acquired by LiDAR + GPS + IMU.
- To establish a comparison we map the obtained trajectories in the LiDAR metric reference system and measure an error in meters.

Statistics



- Trajectories generated up to 12 seconds at 10 FPS.
- We measure the number of valid past and future trajectories based on the determinant criterion.
- We obtain 130K samples from the KITTI tracking dataset (6.7 trajectories per image on average).
- KITTI has only 896 annotated trajectories.

Bibliography

[1] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *Proc. of ICCV*, 2017.
[2] G. Cuffaro, F. Becattini, C. Baecchi, L. Seidenari, and A. Del Bimbo, "Segmentation free object discovery in video," in *European Conference on Computer Vision*, pp. 25–31, Springer, 2016.
[3] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," *arXiv preprint arXiv:1706.05587*, 2017.
[4] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.
[5] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.
[6] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*. Cambridge university press, 2003.
[7] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the kitti vision benchmark suite," in *Proc. of CVPR*, 2012.