# Robust Multi-Target Tracking in Outdoor Traffic Scenarios via Persistence Topology based Robust Motion Segmentation

Somrita Chattopadhyay, Qian Ge, Chunpeng Wei, Edgar Lobaton*
Electrical and Computer Engineering
North Carolina State University, NCSU
Raleigh, North Carolina, USA
schatto2@ncsu.edu, qge2@ncsu.edu, cwei2@ncsu.edu, edgar.lobaton@ncsu.edu

***Abstract*—In this paper, we present a motion segmentation based robust multi-target tracking technique for on-road obstacles. Our approach uses depth imaging information, and integrates persistence topology for segmentation and min-max network flow for tracking. To reduce time as well as computational complexity, the max flow problem is solved using a dynamic programming algorithm. We classify the sensor reading into regions of stationary and moving parts by aligning occupancy maps obtained from the disparity images and then, incorporate Kalman filter in the network flow algorithm to track the moving objects robustly. Our algorithm has been tested on several real-life stereo datasets and the results show that there is an improvement by a factor of three on robustness when comparing performance with and without the topological persistent detections. We also perform measurement accuracy of our algorithm using popular evaluation metrics for segmentation and tracking, and the results look promising.**

***Index Terms*—image motion analysis, image sequence analysis, object detection, image segmentation, autonomous agents**

## I. Introduction

In the past few decades, many researchers have explored the area of intelligent vehicles and tried to make vehicles perceive and analyze their surrounding environment in order to enhance on-road safety. However until now, reliable detection and tracking of on-road obstacles remains one of the most complex tasks for driver assistance and autonomous navigation systems. This issue is challenging due to variable illumination conditions, changing weather conditions, and highly dynamic background. In this paper, we focus on the identification of moving and stationary obstacles on the road, and the robust tracking of mobile agents using visual sensors by applying a segmentation method based on topological persistence. We describe how image sequences taken by a stereo camera can be processed to detect and track multiple moving objects against a moving background.

Traditionally, the problem of multi-object tracking in dynamic environments has been performed either by detecting and classifying the targets in one frame, and tracking those detections in the consecutive frames [1], [2]; or by background subtraction methods i.e. tracking by segmenting the moving objects from the static background [3], [4]. We proceed by segmenting the obstacles present in field of view of the ego vehicle, then classify the obstacles as static and dynamic, and then track only the moving obstacles over time. Previously, the classical approaches have addressed the problem of obstacle segmentation by simply thresholding some likelihood function. These methods are heavily reliant on the choice of threshold values and small changes in these thresholds lead to huge variations in the segmentation results. In our work, we propose a novel approach to dynamic multi-target tracking from a mobile platform utilizing our previous work on robust obstacle segmentation based on persistent topology [5]. Owing to its hierarchical nature, the persistence based method does not rely on a single threshold value; instead, it keeps track of all the detection results, corresponding to different thresholds.

Fig. 1 provides an overview of our methodology. This work builds on our recent results on robust obstacle segmentation based on topological persistence [5]. We start with a UV-disparity approach [6] to separate the ground from the obstacles. A visibility based occupancy map [7] is then computed to segment the obstacles in the occupancy domain. We perform occupancy grid mapping over consecutive frames [8] to compensate the motion of the ego vehicle and label only the obstacles which are in motion. To make our segmentation results further robust, a topological persistence technique is employed. For the tracking module, our method extends the works of Zhang et al. [9] and Pirsiavash et al. [10], and incorporates a dynamic programming approach followed by Kalman filtering [11]. Tracking of the moving obstacles is performed by data association between consecutive frames using network flows. Combining the dynamic programming with Kalman filtering for position and velocity estimation reduces the dependence on color histogram to find correct matches in consecutive frames yielding better results in scenarios of variable illumination conditions.

The remainder of this paper is organized as follows. An overview of our robust motion segmentation approach is presented in sec. II. A description of our proposed obstacle tracking methodology is introduced in sec. III. The metrics used for performance evaluation are explained in sec. IV. Results are discussed in sec. V. Finally, sec. VI summarizes our findings and discusses future scope.

## II. Robust Motion Segmentation

Obstacles are robustly segmented from an occupancy map via topological persistence. The occupancy grid map is obtained from a stereo pair of images using a disparity computation approach [5], [7]. Fig. 2 illustrates the methodology. A fast and easy way of segmentation is simply thresholding the occupancy map. However, the ideal threshold value may change between images even in the

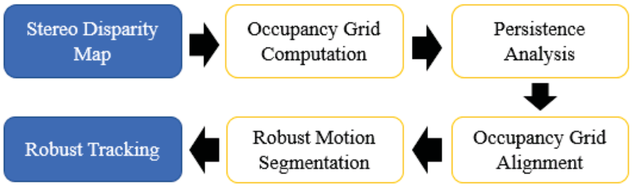
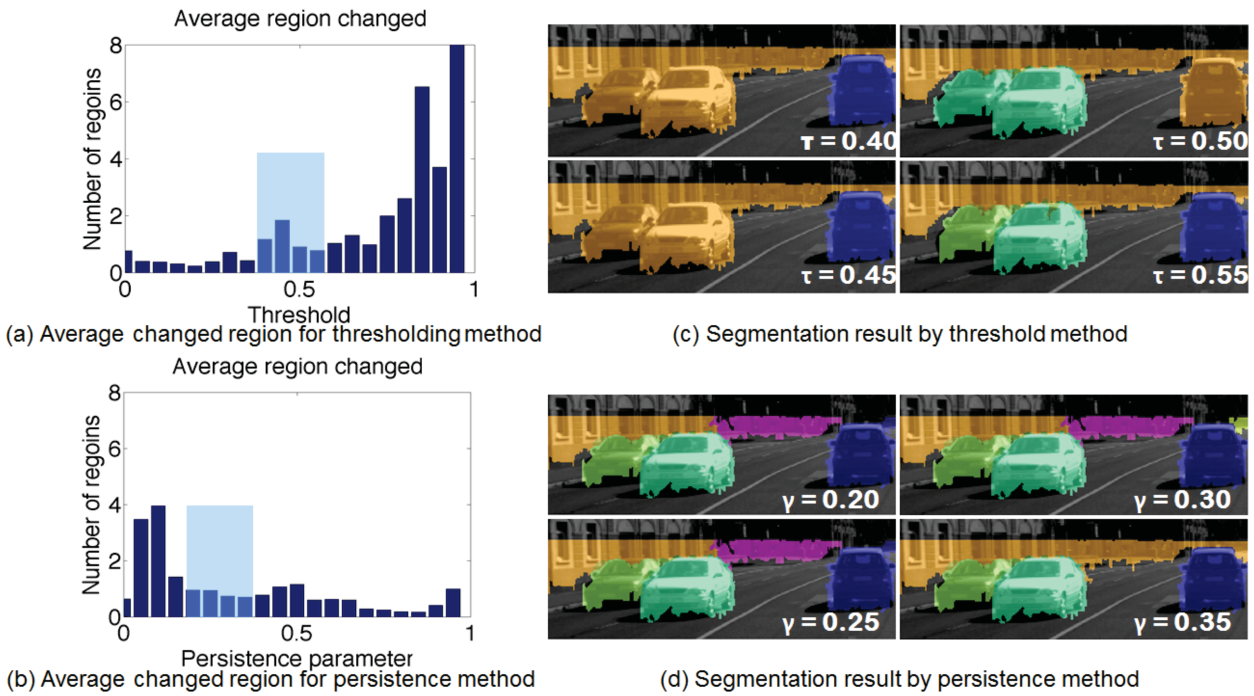


Fig. 1: Flowchart of the proposed methodology

[*] This work was supported by the National Science Foundation under award CNS-1239323.



(a) Average changed region for thresholding method

(c) Segmentation result by threshold method

(b) Average changed region for persistence method

(d) Segmentation result by persistence method

Fig. 3: (a) Average number of regions changed for thresholding approach. (b) Average number of regions changed for persistence approach. (c) Segmentation of thresholding approach for threshold from $0.4$ to $0.55$. (d) Segmentation of persistence approach for threshold parameter from $0.2$ to $0.35$.

same sequence due to variations in the occupancy map attributed to the quality of the disparity map. Furthermore, there may not exist a single threshold that provides all the obstacles corresponding to the peaks of the occupancy map without merging or dividing obstacle regions which are not supposed to. In order to address all of these issues, a topological persistence based approach, introduced in [5], is implemented to generate a more robust segmentation. The main advantage of the persistence diagram is its stability property [12], which translates into the following: we can obtain a segmentation result that is robust to parameter value changes and small variations in the disparity map.

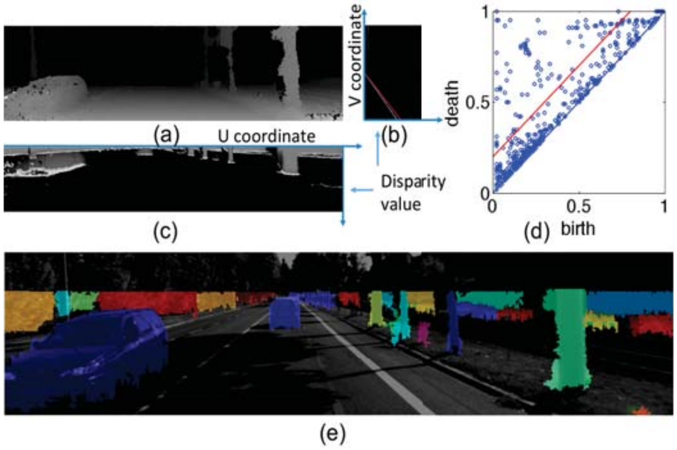


Fig. 2: Robust segmentation processing. (a) Disparity map. (b) Road segmentation in $v$-disparity map. (c) Occupancy grid map in $u$-disparity map. (d) Persistence diagram. (e) Robust segmentation result.

Fig. 3 illustrates the robustness of the segmentation result, the sensitivity is measured by the number of added and removed regions as the parameter changes by $0.05$ for the simple thresholding and persistence approaches. Fig. 3 (a) and (b) show how visible regions change, averaged from $100$ frames using both approaches. On average, the persistence method gives fewer number of regions. In order to quantify this statement, we select parameter values $\tau$ from $0.4$ to $0.55$ for the thresholding method and persistence parameter $\gamma$ from $0.2$ to $0.35$ for the persistence method. These ranges are picked because in average both methods obtain acceptable results. In this range, the persistence method has $0.82$ region changes on an average and the thresholding method has $1.14$. That is a reduction of $28\%$ when using the persistence approach. Fig. 3 (c) and (d) show one example of the segmentation results using both methods over the compared parameter ranges. The thresholding method has a lot of changes, especially for the two cars on the left. On the contrary, the persistence segmentation results are very consistent.

After extracting the persistent regions from the image frames, we work towards motion compensation of the ego vehicle. We extend the occupancy grid mapping scheme to a dynamic occupancy grid mapping framework, which is able to label the stationary and moving objects in the local map [8], [13]. We create a static probability map by accumulating the individual occupancy grids of consecutive image frames. First, we compute the rotational and translational motion between the successive image frames using Scale-invariant feature matching technique (SIFT) and then, using that homography information, an accumulated probability map is constructed by employing a Bayesian filter approach. The probabilities of the visible regions in the ud-domain of the static and the original occupancy map of each frame can then be compared to distinguish between the moving and stationary objects. The regions with higher probability in the

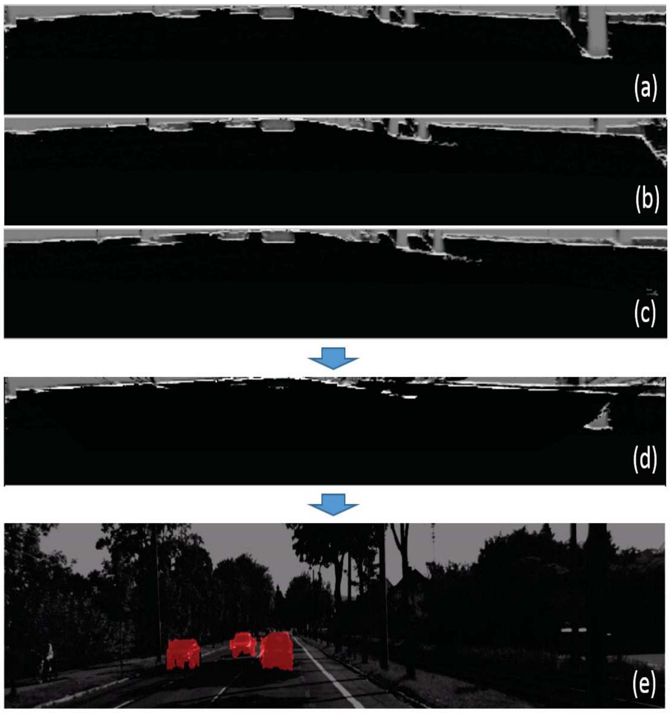


Fig. 4: Results of occupancy grid alignment (a)-(c) Occupancy Grids of 3 successive frames (d) Accumulated Occupancy Grid (e) Motion Segmentation Result.

accumulated occupancy map represent the static objects present in the scene, whereas the regions with low probability are the objects in motion. Fig. 4 displays the results of our segmentation method. Fig. 4 (d) illustrates the accumulated occupancy grid showing high probabilities only for static obstacles.

## III. Tracking Methodology

Data association is done using network flows and then the maximum flow problem is solved using push relabeling [9] and global greedy [10] algorithms. We first construct a Hidden Markov Model based flow network using prior works of Zhang et al [9]. In each frame, the objects segmented as moving are represented as vertices of the flow network. Edges between every pair of vertices represent the cost for considering two objects belonging to the same trajectory. To calculate the shortest, i.e. the minimum cost path in a network, we modify the works of Pirsiavash et el. [10], in which a Discriminatively Trained Part-Based Model (DPM) object detector [14] is used to detect objects in each frame and each object is assigned a unique score. In our work, we have detected the objects in each frame from our persistence based segmentation method and modified the scores associated with the objects using their birth and death times from the persistence diagram. Once the whole network is constructed, instead of using a push-relabeling method [9], we use an iterative dynamic programming method [10] combined with a Kalman filter. This helps us to get an improved multi-target tracking algorithm by incorporating location information in the transition costs. Use of Kalman filter also reduces dependency on the color histogram of the images, which improves the performance in scenarios with variable lighting conditions.

## IV. Evaluation Metrics

To quantitatively measure the performance of our motion segmentation and multi-target tracking algorithm, we have used several popular evaluation metrics from the literature [15]–[17]. The following metrics are used to quantify the performance of our motion segmentation approach:

- **Precision** = correct matches / total groundtruth objects
- **Recall** = correct matches / output objects.
- **FA/Frm** = No. of false alarms per frame.

For tracking performance we use:

- **GT** = No. of groundtruth trajectories.
- **Mostly tracked (MT%)** = Percentage of GT trajectories covered by tracker output for more than 80% in length.
- **Mostly lost (ML%)** = Percentage of GT trajectories covered by tracker output for less than 20% in length.
- **Fragments (Fr)** = The total of No. of times that a groundtruth trajectory is interrupted in tracking result.
- **ID switches (IDS)** = The total of No. of times that a tracked trajectory changes its matched GT identity.
- **Multiple Object Tracking Accuracy (MOTA)** measures the discrete number of errors that occur during tracking.

$$MOTA = 1 - \frac{\Sigma_t(FP(t) + FN(t) + ID(t))}{\Sigma_t N_{GT}(t)} \qquad (1)$$

  where, $FP(t)$, $FN(t)$ and $ID(t)$ denote the number of false positives, missed targets and identity switches at time $t$, respectively. $N_{GT}(t)$ denotes the total number of annotated groundtruth targets at time $t$.
- **Multiple Object Tracking Precision (MOTP)** assesses the trackers precision, i.e. its ability to localize the target in the image.

$$MOTP = \frac{\sum_{t,i} \bar{d}(GT_i^t H_{g(i)}^t)}{\sum_t m_t} \qquad (2)$$

  where $GT_i^t$ and $H_{g(i)}^t$ are the target and its associated hypothesis, respectively, and $m_t$ is the number of matches at time $t$. Intuitively, it provides the average distance over all matched pairs.

## V. Experiments

All the experiments and simulations are implemented in MATLAB on a 2.4 GHz dual-core laptop with 16 GB RAM. We tested and analyzed our proposed method qualitatively and quantitatively using several stereo image sequences representing real road environments from KITTI Vision Benchmark Suite [18]–[20]. In this paper, we have mainly focused on two different datasets, A and B, consisting of 200 and 120 frames representing inner city and residential traffic respectively. Each image is of size $1242 \times 375$. Persistence based segmentation takes about 13.76 seconds per image, and tracking takes about 25.57 seconds for dataset A and 16.23 seconds for dataset B. The persistence parameter for both the datasets is $\gamma = 0.15$. The groundtruth for our experiments is generated by manual annotations. Our method is heavily reliant on depth imaging information. In this paper, we do not focus on generating our own disparity maps and use the ones available in KITTI website. Sometimes, the results suffer due to bad disparity results. We, thus, generate the groundtruth assuming we have good disparity result and see that our motion segmentation and tracker output gives good robust results for disparity values within 7 and 40. The source code of our implementation and datasets used for validation (including manual annotations) will be made available upon acceptance of the paper for publication.

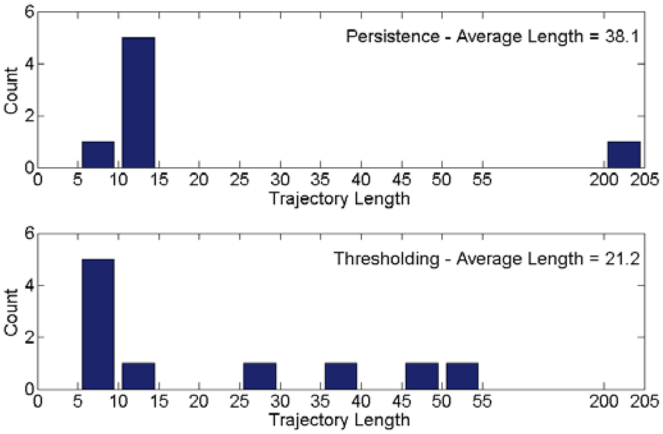


Fig. 5: Comparison result between persistence and threshold methods on dataset A. (Top) Trajectory length distribution for Persistence based tracking method. (Bottom) Trajectory length distribution for Threshold based tracking method.

In this experiment, we use image sequences from datasets, A and B, and apply both traditional threshold method and persistence based segmentation method on those image sequences. The threshold value for the traditional segmentation method is $\tau = 0.5$, which gives a relatively constant segmentation result.

Fig. 5 shows the statistical analysis of the same tracking method for different inputs. The total number of trajectories obtained from the threshold method is 10, which is higher than the number of trajectories computed through persistence method, which equals the number of groundtruth trajectories i.e. 7. But if we focus on the bottom plot in Fig. 5, we can clearly see that the number of trajectories with smaller lengths are much more than the number of trajectories of longer lengths. The segmentation based on thresholding is not robust to the change of scene and location of the object. As a consequence, we can not always get consistent segmentation results for the same object in different frames. If the segmentation results vary hugely between frames, the transition cost between correct tracking pairs increases and leads to gaps in the trajectories, which results in a large number of trajectories having short lengths. On the other hand, our persistence based method gives a high measurement accuracy and higher average trajectory length. This demonstrates the effectiveness of our approach in tracking moving objects over longer periods of time. Fig. 6 shows a consistent tracking result of our algorithm over 10 consecutive frames of dataset A.

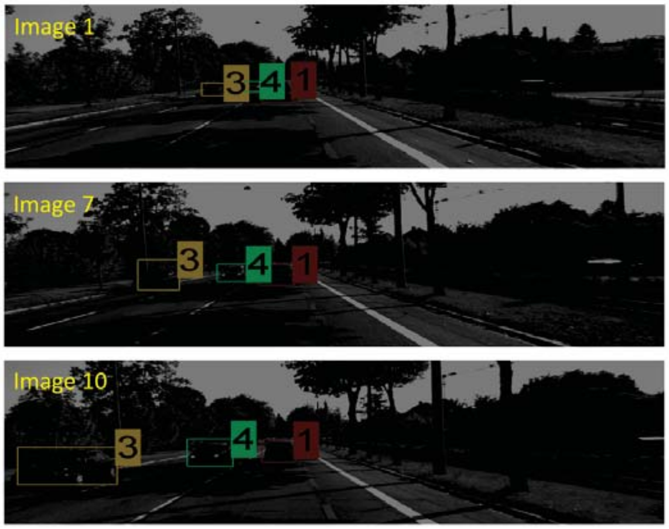


Fig. 6: Tracking Results over 10 consecutive frames from Dataset A

| Dataset | Precision | Recall | FA/Frm |
|---|---|---|---|
| A | 0.95 | 0.97 | 0.25 |
| B | 0.91 | 0.96 | 0.42 |

TABLE I: Analysis of our Motion Segmentation Method

| Dataset | GT | MT% | ML% | Fr | IDS | MOTA | MOTP |
|---|---|---|---|---|---|---|---|
| A | 7 | 0.80 | 0 | 0 | 0 | 1 | 0.83 |
| B | 7 | 0.86 | 0 | 0 | 3 | 0.88 | 0.81 |

TABLE II: Quantitative Analysis of our Tracking Method

To demonstrate the effectiveness of our proposed method, sample evaluation runs are made on several datasets from KITTI and the results are shown in Table I and Table II.

Our evaluation on Datasets A and B depicts that the proposed algorithm works pretty accurately and precisely. The rates of false positives, true positives and false negatives for set A are 0, 1 and 0 respectively and 0.1, 0.87 and 0.03 respectively using the CLEARMOT metrics. We only consider a match to be correct if the overlap is more than 50%.

## VI. Conclusion

In this paper, we propose a robust tracking technique based on motion segmentation [5]. We extend the prior works of Zhang et al. [9] and Pirsiavash et al. [10] by incorporating a Kalman filter, and apply tools from persistent topology to refine their approaches. The quantitative analysis of our method shows that the tracking performance is more robust when we use persistent topology for segmentation instead of the traditional thresholding method. Our analysis also depicts that incorporating a Kalman filter to estimate the position and velocity of the target objects in consecutive frames further improves our results.

In the future, we plan to continue our work to further improve the robustness of our proposed approach to handle errors in disparity computations as well as extending our methodology to incorporate classification of obstacles.

## References

[1] A. Geiger, M. Lauer, C. Wojek, C. Stiller, and R. Urtasun, “3d traffic scene understanding from movable platforms,” *Pattern Analysis and Machine Intelligence (PAMI)*, 2014.

[2] H. Zhang, A. Geiger, and R. Urtasun, “Understanding high-level semantics by modeling traffic patterns,” in *International Conference on Computer Vision (ICCV)*, 2013.

[3] F. Erbs, A. Barth, and U. Franke, “Moving vehicle detection by optimal segmentation of the dynamic stixel world,” in *Intelligent Vehicles Symposium (IV), 2011 IEEE*. IEEE, 2011, pp. 951–956.

[4] C. Wang and Z. Song, “Vehicle detection based on spatial-temporal connection background subtraction,” in *Information and Automation (ICIA), 2011 IEEE International Conference on*. IEEE, 2011, pp. 320–323.

[5] C. Wei, Q. Ge, S. Chattopadhyay, and E. Lobaton, “Robust obstacle segmentation based on topological persistence in outdoor traffic scenes,” in *Computational Intelligence in Vehicles and Transportation Systems (CIVTS), IEEE Symposium on*, 2014.

[6] Z. Hu and K. Uchimura, "Uv-disparity: an efficient algorithm for stereovision based scene analysis," in *Intelligent Vehicles Symposium, 2005. Proceedings. IEEE*. IEEE, 2005, pp. 48–54.

[7] M. Perrollaz, J.-D. Yoder, A. Nègre, A. Spalanzani, and C. Laugier, "A visibility-based approach for occupancy grid computation in disparity space," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 13, no. 3, pp. 1383–1393, 2012.

[8] Y. Li and Y. Ruichek, "Occupancy grid mapping in urban environments from a moving on-board stereo-vision system," *Sensors*, vol. 14, no. 6, pp. 10 454–10 478, 2014.

[9] L. Zhang, Y. Li, and R. Nevatia, "Global data association for multi-object tracking using network flows," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*. IEEE, 2008, pp. 1–8.

[10] H. Pirsiavash, D. Ramanan, and C. C. Fowlkes, "Globally-optimal greedy algorithms for tracking a variable number of objects," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. IEEE, 2011, pp. 1201–1208.

[11] U. Franke, C. Rabe, H. Badino, and S. Gehrig, "6d-vision: Fusion of stereo and motion for robust environment perception," in *Pattern Recognition*. Springer, 2005, pp. 216–223.

[12] D. Cohen-Steiner, H. Edelsbrunner, and J. Harer, "Stability of persistence diagrams," *Discrete Comput. Geom.*, vol. 37, no. 1, pp. 103–120, Jan. 2007. [Online]. Available: http://dx.doi.org/10.1007/s00454-006-1276-5

[13] T. Gindele, S. Brechtel, J. Schroder, and R. Dillmann, "Bayesian occupancy grid filter for dynamic environments using prior map knowledge," in *Intelligent Vehicles Symposium, 2009 IEEE*. IEEE, 2009, pp. 669–676.

[14] P. F. Felzenszwalb, R. B. Girshick, and D. McAllester, "Cascade object detection with deformable part models," in *Computer vision and pattern recognition (CVPR), 2010 IEEE conference on*. IEEE, 2010, pp. 2241–2248.

[15] A. Milan, K. Schindler, and S. Roth, "Challenges of ground truth evaluation of multi-target tracking," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2013 IEEE Conference on*. IEEE, 2013, pp. 735–742.

[16] B. Keni and S. Rainer, "Evaluating multiple object tracking performance: the clear mot metrics," *EURASIP Journal on Image and Video Processing*, vol. 2008, 2008.

[17] Y. Li, C. Huang, and R. Nevatia, "Learning to associate: Hybridboosted multi-target tracker for crowded scene," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. IEEE, 2009, pp. 2953–2960.

[18] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the kitti vision benchmark suite," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012.

[19] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The kitti dataset," *International Journal of Robotics Research (IJRR)*, 2013.

[20] J. Fritsch, T. Kuehnl, and A. Geiger, "A new performance measure and evaluation benchmark for road detection algorithms," in *International Conference on Intelligent Transportation Systems (ITSC)*, 2013.