Sparse Scene Flow Segmentation for Moving Object Detection in Urban Environments

Philip Lenz, Julius Ziegler, Andreas Geiger and Martin Roser
Institute of Measurement and Control Systems
Karlsruhe Institute of Technology
D-76131 Karlsruhe, Germany
{lenz | ziegler | geiger | roser}@kit.edu

***Abstract*— Modern driver assistance systems such as collision avoidance or intersection assistance need reliable information on the current environment. Extracting such information from camera-based systems is a complex and challenging task for inner city traffic scenarios. This paper presents an approach for object detection utilizing sparse scene flow. For consecutive stereo images taken from a moving vehicle, corresponding interest points are extracted. Thus, for every interest point, disparity and optical flow values are known and consequently, scene flow can be calculated. Adjacent interest points describing a similar scene flow are considered to belong to one rigid object. The proposed method does not rely on object classes and allows for a robust detection of dynamic objects in traffic scenes. Leading vehicles are continuously detected for several frames. Oncoming objects are detected within five frames after their appearance.**

I. INTRODUCTION

Inner city traffic is a very complex and demanding scenario for modern driver assistance systems (DAS). Different applications such as collision avoidance, lane keeping or intersection assistance need reliable information on the current traffic situation.

Perception and understanding of highly dynamic traffic scenes is crucial for such systems. Urban traffic is more complex than highway traffic, thus the task for DAS is more challenging and currently an unsolved problem: Traffic scenes are crowded with many different types of traffic participants such as cars, pedestrians, cyclists or trams which must be distinguished while the surrounding scenery may differ arbitrarily. Two typical scenarios for rural roads and inner city traffic are given in Fig. 1. In these situations, the DAS itself must work reliably with low error-rates and react correctly to abruptly changing scenarios considering objects in a wide range (from close to the hood to more than $50\,\mathrm{m}$ ahead).

The goal of our approach is a class-independent detection of moving objects for inner city traffic scenarios without applying a previous training step. The proposed method provides three-dimensional (3D) position information of moving objects in a world coordinate system. We believe that this information constitutes valuable input to downstream processing steps, which enable functions like tracking, trajectory estimation, and, ultimately, a high-level description of the current environment.

In this work we apply a stereo camera for object detection. By exploiting four images at once, stereo and optical flow information are obtained. Since for every detected point the 3D position and the optical flow between two consecutive frames is known, a sparse scene flow description [27] of the scene is achieved, describing 3D motion of world points. Since interest points can be tracked over several frames, the scene flow is computed via finite differences for a track up to five 3D positions. Utilizing this information, points describing a similar scene flow are grouped together belonging to corresponding rigid objects in the scene. Adjacent points are connected using a Delaunay triangulation and the resulting edges are removed if the scene flow difference exceeds a certain threshold. As the error of the 3D position grows quadratically with the distance, error propagation of the scene flow computation is taken into account to remove edges by thresholding the calculated Mahalanobis distance.

Pedestrians and cars are detected within a reasonable range for inner city intersection scenarios. Detected objects in front of the observer vehicle are tracked for up to $15\,\mathrm{s}$ and infrequent traffic participants such as wheelchair users are detected as well.

(a) Rural Road Traffic

(b) Inner City Traffic

Fig. 1. **Typical traffic situations for rural and inner city scenarios.** The road geometry of highways and rural roads (a) is rather clear compared to arbitrary inner city streets (b) where distinctive features may be missing or misleading. Due to the high degree of diversity of traffic participants and views, class dependent object detectors will be error prone for DAS applications.

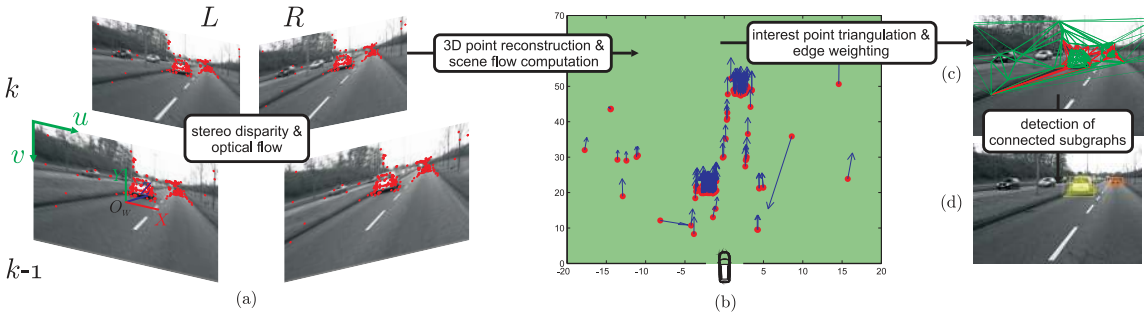


Fig. 2. **System Overview.** (a) Interest points are detected in two consecutive stereo image pairs. (b) Stereo disparity and optical flow information lead to a 3D reconstruction of the detected points which are here projected on the detected ground plane. (c) Interest points are connected in a graph-like structure and the resulting edges are removed (marked in red) if the scene flow difference exceeds a certain threshold. (d) Remaining connected components describe moving objects in the scene.

The paper is structured as follows: In the next section, related work is presented and distinguished from our work. An overview of the method is given in Section III before the object detection is discussed in more detail in Section IV. Some results are shown in Section V before the paper is concluded by a summary and an outlook.

II. Previous Work

Robust scene perception in urban environments is a current field of research [1], [8], [9], [13], [18], [21], [22], [28] and is investigated by applying different sensors. LIDAR systems offer a high geometric accuracy for 3D world points that can be used to detect multiple objects under ground plane assumption [28] or by using a model-based object detector [17]. Recent work shows, that motion estimation from range images gives promising results as well [21].

While LIDAR can directly provide important 3D information, the required data lacks of rich appearance. This is where camera based systems bear high potential. Several approaches for object recognition in the image domain exist to detect single [6], [20] or multiple [11], [26] object classes. However, this task is even harder, if the observing camera is moving and egomotion is affecting the image acquisition due to blur or changing lighting condition [5], [19]. These methods generally use an uninformed search over the whole image and discard further information on the scene geometry. Additionally, environment perception for DAS does not only require detected objects in the image domain but also demands a 3D description of the observed traffic scenario in order to perform e.g. collision avoidance. The 3D scene geometry can be estimated using only a single frame as well, but several restrictions must apply and the scene must be rich of known objects [16].

A stereo camera system allows for geometric reconstruction of the scene in a 3D world coordinate system [15]. Nedevschi [23], [24] introduced a system for inner city object detection and classification utilizing a stereo camera system. Dense stereo information is used for reconstruction and 3D data used for model selection and scale estimation. Object classification relies on pattern matching exploiting object scale, orientation and location in the scene obtained by a previous model selection step. Franke [13] and Wedel [29] utilize stereo data to estimate 3D position and 3D motion of interest points for a robust detection of moving objects. Pfeiffer [25] introduced the tracking of stixels obtained from dense stereo images as a representation for suburban traffic scenarios.

Mueller [22] evaluated the optical flow difference of adjacent interest points to group them describing independent objects in the scene. This approach utilizes interest points detected in a monocular image sequence which are connected in a graph-like structure and is most similar to the proposed one. In contrast, we are using scene flow information for detected interest points since the approach in [22] uses constraints describing motion of rigid objects in the image plane that are not strong enough for inner city scenarios.

Ess [9] uses the HOG framework [6] for object detection and consequently multiple object tracking for inner city traffic scenes. For this application, it turned out to be superior compared to [11], [20] since the shape variations of traffic participants are relatively small [9].

III. System Overview

Figure 2 outlines our system: Interest points are detected in two consecutive and rectified stereo images and checked for mutual consistency. The resulting disparity values d lead to a 3D description for every detected interest point. Due to the unknown mounting position of the stereo camera rig and the pitching of the vehicle, the ground plane is detected to obtain a meaningful description of the surrounding. Every interest point is tracked over time and its scene flow is calculated using finite difference approximation to yield derivatives. The associated covariance is obtained by linear error propagation.

A graph-like structure connecting all detected interest points in the image plane is generated using Delaunay triangulation [2]. The resulting edges are removed according to scene flow differences exceeding a certain threshold with respect to the uncertainty of the computed 3D position of every interest point. The remaining connected components of the graph describe moving objects in the scene. Detected objects are tracked over time using a global nearest neighbor (GNN) approach [3].

IV. Object Detection

The configuration of our stereo camera rig is depicted in Fig. 2. Two consecutive stereo images are considered. The world reference frame is denoted with $\mathbf{O}_W$ and coincides with the left camera frame. Consequently the extrinsic calibration of the setup $\{\mathbf{R}_C, \mathbf{t}_C\}$ is known and assumed to be constant. Assuming a proper, calibration we will use rectified input images for our algorithm, so that we only have to deal with a horizontal search line for correspondences in the stereo image pairs.

A. Scene Flow Computation

Interest points $\mathbf{x} = [u, v]^\top$ are detected in two consecutive stereo image pairs at time $k, k-1$ using the algorithm of [14]. Only reliable feature correspondences which match in a loop are kept ($\mathbf{x}_{L,k-1} \leftrightarrow \mathbf{x}_{R,k-1} \leftrightarrow \mathbf{x}_{R,k} \leftrightarrow \mathbf{x}_{L,k} \leftrightarrow \mathbf{x}_{L,k-1}$). Since the signal-to-noise ratio is rather low considering only the previous frame, detected interest points are associated and stored as tracklets for up to 5 time steps.

Since rectified images are used and the disparities are estimated at sub-pixel accuracy, the extracted interest points $\mathbf{x}_{k-1}$, $\mathbf{x}_k$ are mapped to 3D points in the world coordinate system $\mathbf{X}_{k-1}, \mathbf{X}_k$ with $\mathbf{X} = [X, Y, Z]^\top \in \mathbb{R}^3$. The reconstruction is given by

$$X = \frac{(u_L - c_{u,L}) \cdot b}{d} \tag{1}$$

$$Y = \frac{(v_L - c_{v,L}) \cdot b}{d} \tag{2}$$

$$Z = \frac{b \cdot f}{d} \tag{3}$$

where b denotes the baseline of the stereo system, c_u and c_v the principal point of the camera and f the focal length for the rectified images. The center of the left-handed world coordinate system is $\mathbf{O}_W$ with the X axis pointing to the right as depicted in Fig. 2.

The velocity $\mathbf{V}$ of every world point is assumed to be constant within the tracked time t of 5 frames ($t = 0.5\,\mathrm{s}$). Thus, the velocity is computed as the first order derivative of the world points $\mathbf{X}_{k-\Delta t_i}$:

$$\mathbf{V} = \frac{\Delta \mathbf{X}_{k-\Delta t_i}}{\Delta t} \tag{4}$$

The velocity is measured at discrete time steps Δt_i, $i = 0 \ldots 5$ at a constant sampling rate of $1/\Delta t = 10\,\mathrm{Hz}$. This assumption of an equispaced grid leads to the calculation of the coefficients for the first order derivative f' depending on the tracked time steps according to [12]:

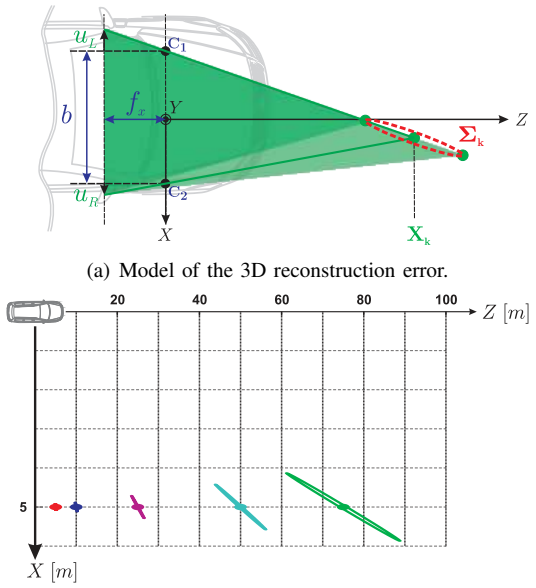
$$f'(\mathbf{x}) \approx a_0 f(\mathbf{x}) + a_1 f(\mathbf{x} - 1) \ldots + a_5 f(\mathbf{x} - 5) \tag{5}$$

Since the proposed algorithm is designed for automotive applications, the delay between occurrence and detection of an object should be minimal. Therefore, backward differences are used to determine the scene flow. The coefficients are listed in Table I. An example of the resulting scene flow is shown in the system overview in Fig. 2.

Δt_i	a_0	a_1	a_2	a_3	a_4	a_5
1	1	−1				
2	3/2	−2	1/2			
3	11/6	−3	3/2	−1/3		
4	25/12	−4	3	−4/3	1/4	
5	137/60	−5	5	−10/3	5/4	−1/5

TABLE I

Coefficients for finite backward differences.



(a) Model of the 3D reconstruction error.

(b) Quadratically growing reconstruction error.

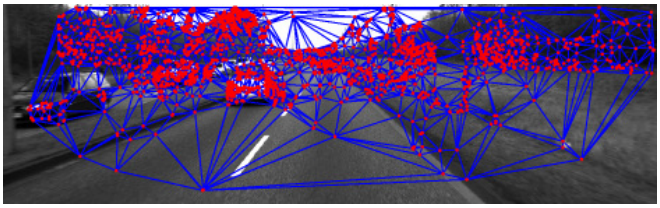
Fig. 3. **Resulting noise of 3D reconstruction.** One pixel noise of the disparity $u_L - u_R$ leads to the 3D position $\mathbf{X}_k$ and the associated covariance ellipsoid Σ_k (a). Error propagation leads to a quadratically growing error of the 3D position with the largest uncertainty perpendicular to the line of sight (b).

B. Scene Flow Clustering

The computed scene flow is now clustered into groups describing a similar motion. For this purpose, a graph-like structure is built up where interest points are considered as nodes and the adjacent nodes are connected by edges. To obtain this structure, a Delaunay triangulation [2] as shown in Fig. 4(a) is applied to determine neighboring interest points for further processing steps.

Considering the absolute scene flow difference $\mathbf{V}_i - \mathbf{V}_j$ of adjacent nodes i, j and removing the corresponding edges does not lead to a satisfying solution as shown in Fig. 4(b). The reconstructed 3D position is error-prone due to measurement noise. The resulting error of the 3D position is computed by linear error propagation. The error grows quadratically with the distance as depicted in Fig. 3, thus a fixed threshold on the scene flow difference is not suitable.

Therefore, the Jacobian $\mathbf{J}$ of the scene flow for a 3D world point $\mathbf{X}$ is calculated according to equation 6 with respect to all image coordinates u_L, u_R, v for all possible time steps $k - \Delta t_i$.

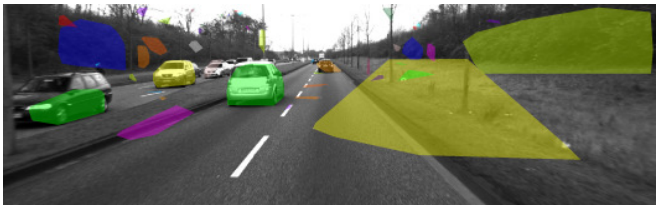


(a) Interest point triangulation

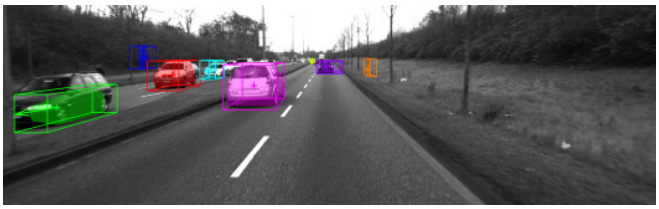
(b) Fixed threshold

(c) Thresholding of Mahalanobis distance

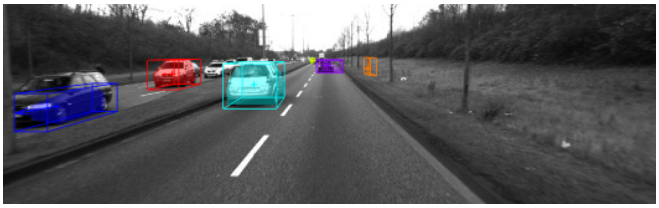
Fig. 4. **Graph building and clustering.** Detected interest points are connected using a Delaunay Triangulation (a). A fixed threshold to remove the edges considering the scene flow difference does not lead to a satisfying solution (b). Considering error propagation and applying a threshold on the resulting Mahalanobis distance groups similar objects in the scene (c).



(a) Remaining connected components



(b) Geometrically checked objects



(c) Tracked objects

Fig. 5. **Object detection.** All remaining connected components of the computed graph are found by depth-first search (a). Geometric features are taken into account to remove static parts of the scene. Objects are covered by a bounding box, center and the velocity of the object are marked in their respective color (b). Object tracking and expecting a detection in at least two time steps leads to the final detection (c).

$$\mathbf{J} = \frac{d\mathbf{V}}{d\mathbf{x}} = \begin{bmatrix} \frac{\partial V_X}{\partial u_{L,k-\Delta t_i}} & \frac{\partial V_X}{\partial u_{R,k-\Delta t_i}} & \frac{\partial V_X}{\partial v_{k-\Delta t_i}} \\ \frac{\partial V_Y}{\partial u_{L,k-\Delta t_i}} & \frac{\partial V_Y}{\partial u_{R,k-\Delta t_i}} & \frac{\partial V_Y}{\partial v_{k-\Delta t_i}} \\ \frac{\partial V_Z}{\partial u_{L,k-\Delta t_i}} & \frac{\partial V_Z}{\partial u_{R,k-\Delta t_i}} & \frac{\partial V_Z}{\partial v_{k-\Delta t_i}} \end{bmatrix}, \; i = 0 \dots 5 \tag{6}$$

The covariance $\mathbf{\Sigma}$ of the scene flow is given by

$$\Sigma = \mathbf{JSJ}^\top \tag{7}$$

where $\mathbf{S}$ is the diagonal measurement noise matrix assuming a measurement noise of $0.5\,\text{pixel}$. Assuming the noise to be Gaussian, for two adjacent nodes i and j, the Mahalanobis distance Δ is given by

$$\Delta(\mathbf{V}_i, \mathbf{V}_j) = \sqrt{(\mathbf{V}_i - \mathbf{V}_j)^\top \Sigma_{i,j}^{-1} (\mathbf{V}_i - \mathbf{V}_j)} \tag{8}$$

according to [7]. Each edge is weighted 0 (or removed) if Δ exceeds a certain threshold. Consequently, the remaining subgraphs contain nodes with small scene flow differences and similar motion describing a rigid object as shown in Fig. 4(c). The remaining connected components in Fig. 5(a) of the graph are detected by depth-first search [4].

Since the egomotion of the observing vehicle is not known, the algorithm responds to static parts of the scene as well. However, static parts describe large 3D volumes or are typically part of the ground plane. Therefore, objects exceeding reasonable dimensions are neglected as well as objects that are part of or not standing on the ground plane. Both latter criteria are achieved by estimating the parameters $\pi_1 \dots \pi_4$ of the ground plane

$$E : \pi_1 X + \pi_2 Y + \pi_3 Z + \pi_4 = 0 \tag{9}$$

using a random sample consensus (RANSAC) algorithm.

C. Object Association

The last step is to form tracks produced by the same object within the considered sequence. The Observation-to-Track Association is handled by a GNN approach [3]. The position $\mathbf{X}_{k-1}^{(i)}$ of a detected object io is predicted for the current time step $\mathbf{X}_{p,k}^{(i)}$ assuming a constant motion within Δt. For every predicted object, a constant track gate is placed around $\mathbf{X}_{p,k}^{(i)}$ considering acceptable measurement and prediction error as shown in Fig. 6. If two observations are associated with one predicted track, the unassigned observation will initiate a new track. Objects are assumed to appear at least in two consecutive time steps for confirmation and only one miss is accepted before deletion.

The object association results in the detected objects in Fig. 5(c). The blue erroneous object in Fig. 5(b) is not detected since it only appears in this very time step. However, objects are detected with one frame delay such as the police car, which will initiate a track from the next frame on.

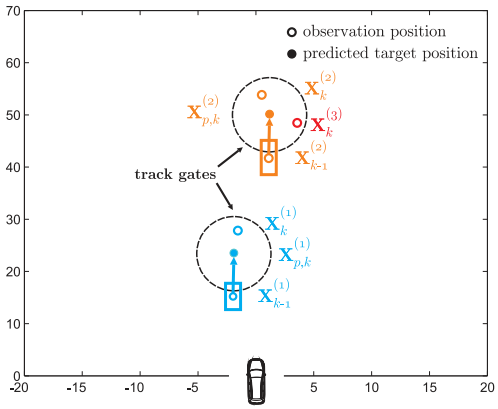


Fig. 6. **Global nearest neighbor algorithm.** Track gates are set around the predicted positions $\mathbf{X}_{p,k}^{(i)}$. For unambiguous gates one observation is assigned to exactly one prediction ($\mathbf{X}^{(1)}$). For a conflicting prediction such as $\mathbf{X}_{p,k}^{(2)}$, the unassigned observation $\mathbf{X}_k^{(3)}$ initiates a new track.

V. Experiments

For all experiments, we used grayscale images with a resolution of 1392 $\times$ 512 pixels. The image sequences cover rural, suburban and inner city traffic scenarios. For all images, at least 2000 interest points were detected. Our algorithm is implemented in Matlab processing at least one frame per second on one core of an Intel Core2Duo with 2.4 GHz and 4 GB RAM.

To evaluate the capability of the proposed algorithm to detect objects continuously, the rural traffic scenario in Fig. 7 is considered. Two vehicles in front are visible within 115 frames in a distance of approximately 10 m and 35 m respectively. Using the GNN approach described in chapter IV-C, both cars are detected continuously within the whole sequence. Approaching cars in this sequence are detected within at most three time steps after they are fully visible and at a distance up to 25 m. For a greater range of up to 60 m, objects are detected within five time steps. Since the maximum speed for this scenario is usually limited to 60 km/h, we consider this time period as sufficient. After detection of an object at a distance of 50 m, there are 2 s remaining for an appropriate reaction. Since our approach is class-independent, uncommon but moving objects are detected as well, e.g. the wheelchair user on the left sidewalk in the second frame. We tested the appearance-based detector of [11] using this image. Since this class was not trained, the detection of a person or a bike in this image fails, although for the class car there were detected objects at close and far range.

Figure 8 shows that the detection at close and far range up to 60 m works well in a rural environment. For inner city scenarios as depicted in Fig. 9, cars and pedestrians are detected. The algorithm fails here for objects in a far range for more than 50 m since stereo reconstruction exhibits a heavy noise level in depth direction and the background is highly textured and therefore only few interest points are detected on objects of interest. The tracks of turning cars are very short due the assumption of continuous motion.

Fig. 7. **Results on a rural road.** The cars in front of the observer vehicle are tracked continuously within 15 s. The class-independent method detects the wheelchair user in the upper left area of the second frame. Static objects such as the advertisement or trees are detected as objects since egomotion is not compensated.

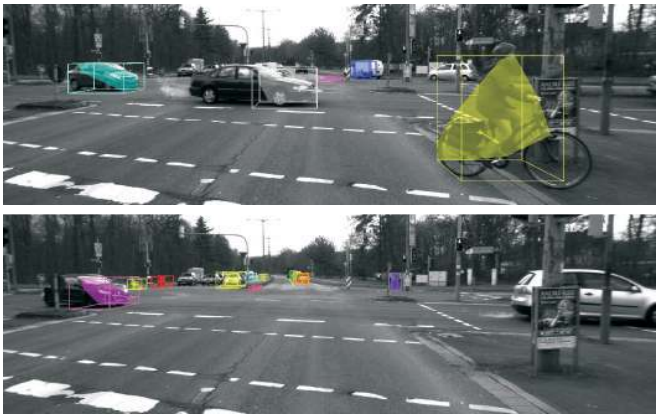


Fig. 8. **Results on a rural intersection.** Close range objects as the cyclist or the crossing traffic in the first frame are detected as well as objects in the far range in the second frame.

Fig. 9. **Results on inner city intersections.** (a) Slowly moving and small objects such as the pedestrian in the first two frames are detected in a range up to 30 m. However, the track of such a small object is interrupted and it is not continuously tracked. Similarly moving groups of pedestrians are detected as one object since the scene flow difference is not unique and the number of detected interest points is to low. (b) Moving objects in this sequence are detected, but especially for the far range static objects are detected as well. (c) Turning cars and partly occluded objects, which where fully visible and observed in a previous frame, are detected. Since continuous velocity is assumed, the Observation-to-Track association fails for sharply turning objects.

VI. Conclusion and Future Work

We presented a novel approach for class-independent object detection for inner-city traffic scenarios. The proposed algorithm uses computationally efficient sparse interest points in consecutive stereo images to compute the scene flow. Clustered points describing a similar scene flow result in a robust detection of independently moving rigid objects in the scene. Our results compared to [22] indicate, that this approach outperforms a similar method utilizing only optical flow for object detection. Due to the class-independent approach, objects are detected where an appearance-based object detectors fails which works well for known object classes. For typical inner-city traffic scenarios unknown objects or objects in an uncommon view may occur and in this case these approaches may be not sufficient for DAS applications.

Our next steps will include a more sophisticated multiple target tracking to improve the distinction between traffic participants and erroneously detected objects, handle partly occluded objects as well as erroneous track gates. Additionally, tracking of objects coming to a stop will be improved. Further, we want to include egomotion compensation and a motion model for detected interest points and objects. Thereby rotation of detected objects will be explicitly considered. Furthermore, a global solution to remove edges in the graph structure will be investigated. Although we want to follow our class-independent approach for object detection in the first step, semantic information should be provided after the initial object detection for higher-level reasoning, e.g. for navigation decisions of a DAS. Since most of the existing large object databases, e.g. [10], are not providing the required data for our approach, a ground truth data base for our application will be built up for evaluation.

VII. Acknowledgment

The authors would like to thank the “Karlsruhe School of Optics and Photonics” and the “Deutsche Forschungsgemeinschaft” for supporting this work.

REFERENCES

[1] A. Bak, S. Bouchafa, and D. Aubert. Detection of independently moving objects through stereo vision and ego-motion extraction. In *IEEE Intelligent Vehicles Symposium*, pages 863 –870, 2010.

[2] C. B. Barber, D. P. Dobkin, and H. Huhdanpaa. The quickhull algorithm for convex hulls. *ACM Transactions on Mathematical Software*, 22(4):469–483, 1996.

[3] S. S. Blackman. Multiple hypothesis tracking for multiple target tracking. *IEEE Aerospace and Electronic Systems Magazine*, 19(1):5 –18, 2004.

[4] T. H. Cormen. *Introduction to algorithms*. MIT Press, Cambridge, 3. ed. edition, 2009.

[5] N. Cornelis, B. Leibe, K. Cornelis, and L. Gool. 3d urban scene modeling integrating recognition and reconstruction. *International Journal of Computer Vision*, 78(2-3):121–141, 2007.

[6] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 886–893, 2005.

[7] L. Devroye, L. Györfi, and G. Lugosi. *A probabilistic theory of pattern recognition*. Applications of mathematics. Springer, New York, 1996.

[8] A. Ess, B. Leibe, K. Schindler, and L. van Gool. Robust multiperson tracking from a mobile platform. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(10):1831–1846, 2009.

[9] A. Ess, K. Schindler, B. Leibe, and L. V. Gool. Object detection and tracking for autonomous navigation in dynamic environments. In *International Journal of Robotics Research*, volume 29, pages 1707–1725, 2010.

[10] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The PASCAL Visual Object Classes Challenge 2010 (VOC2010) Results. http://www.pascal-network.org/challenges/VOC/voc2010/workshop/index.html.

[11] P. Felzenszwalb, D. McAllester, and D. Ramanan. A discriminatively trained, multiscale, deformable part model. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1 –8, june 2008.

[12] B. Fornberg. Classroom note: Calculation of weights in finite difference formulas. *SIAM Review*, 40:685–691, September 1998.

[13] U. Franke, C. Rabe, H. Badino, and S. K. Gehrig. 6d-vision: Fusion of stereo and motion for robust environment perception. In *Deutsche Arbeitsgemeinschaft für Mustererkennung e. V. Symposium*, pages 216–223, 2005.

[14] A. Geiger, J. Ziegler, and C. Stiller. Stereoscan: Dense 3d reconstruction in real-time. In *IEEE Intelligent Vehicles Symposium*, Baden-Baden, Germany, June 2011.

[15] A. Hartley, Richard ; Zisserman. *Multiple view geometry in computer vision*. Cambridge Univ. Press, Cambridge, 2. ed., 5. print. edition, 2008.

[16] D. Hoiem, A. A. Efros, and M. Hebert. Putting objects in perspective. volume 80, pages 3–15, Hingham, MA, USA, October 2008. Kluwer Academic Publishers.

[17] S. Kammel, J. Ziegler, B. Pitzer, M. Werling, T. Gindele, D. Jagzent, J. Schröder, M. Thuy, M. Goebl, F. v. Hundelshausen, O. Pink, C. Frese, and C. Stiller. Team annieway's autonomous system for the 2007 darpa urban challenge. *Journal of Field Robotics*, 25:615–639, September 2008.

[18] H. Lategahn, W. Derendarz, T. Graf, B. Kitt, and J. Effertz. Occupancy grid computation from dense stereo and sparse structure and motion points for automotive applications. In *IEEE Intelligent Vehicles Symposium*, pages 819 –824, 2010.

[19] B. Leibe, N. Cornelis, K. Cornelis, and L. J. V. Gool. Dynamic 3d scene analysis from a moving vehicle. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2007.

[20] B. Leibe, E. Seemann, and B. Schiele. Pedestrian detection in crowded scenes. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages 878–885, Washington, DC, USA, 2005. IEEE Computer Society.

[21] F. Moosmann and T. Fraichard. Motion estimation from range images in dynamic outdoor scenes. In *IEEE Conference on Robotics and Automation*, pages 142 –147, May 2010.

[22] D. Muller, M. Meuter, and S.-B. Park. Motion segmentation using interest points. In *IEEE Intelligent Vehicles Symposium*, pages 19 –24, 2008.

[23] S. Nedevschi, S. Bota, and C. Tomiuc. Stereo-based pedestrian detection for collision-avoidance applications. *IEEE Transactions on Intelligent Transportation Systems*, 10(3):380–391, 2009.

[24] S. Nedevschi, R. Danescu, T. Marita, F. Oniga, C. Pocol, S. Sobol, C. Tomiuc, C. Vancea, M. Meinecke, T. Graf, T. B. To, and M. Obojski. A sensor for urban driving assistance systems based on dense stereovision. *IEEE Intelligent Vehicles Symposium*, pages 276 –283, june 2007.

[25] D. Pfeiffer and U. Franke. Efficient representation of traffic scenes by means of dynamic stixels. In *IEEE Intelligent Vehicles Symposium*, pages 217 –224, 2010.

[26] A. Torralba, K. Murphy, and W. Freeman. Sharing features: efficient boosting procedures for multiclass object detection. *IEEE Conference on Computer Vision and Pattern Recognition*, 2:II–762 – II–769 Vol.2, june-2 july 2004.

[27] S. Vedula, S. Baker, P. Rander, R. Collins, and T. Kanade. Three-dimensional scene flow. In *IEEE Conference on Computer Vision*, volume 2, pages 722 – 729, 1999.

[28] T.-D. Vu, J. Burlet, and O. Aycard. Grid-based localization and online mapping with moving objects detection and tracking: new results. In *IEEE Intelligent Vehicles Symposium*, pages 684 –689, 2008.

[29] A. Wedel, A. Meißner, C. Rabe, U. Franke, and D. Cremers. Detection and segmentation of independently moving objects from dense scene flow. In *International Conference on Energy Minimization Methods in Computer Vision and Pattern Recognition*, pages 14–27, Berlin, Heidelberg, 2009. Springer-Verlag.