May I Enter the Roundabout?
A Time-To-Contact Computation Based on Stereo-Vision

Maximilian Muffert*, Timo Milbich**, David Pfeiffer* and Uwe Franke*
* Daimler Research, Sindelfingen, Germany, Email: [firstname.secondname]@daimler.com
** HFT, Stuttgart, Germany, Email: timo_milbich@gmx.de

***Abstract*— This paper presents a stereo-vision based system for the recognition of dangerous situations at roundabouts. At first, we investigate the necessary field of view and viewing direction using videos taken by a panoramic camera. Using the insights of these tests we build up a stereo-vision system. This system is based on the well established disparity estimation scheme Semi-Global Matching and the recently introduced medium-level representation called Dynamic Stixel-World. A time-to-contact measure is defined that makes explicit use of the roundabouts structural characteristics. Using this measure enables us to create a system for driver warning or possible automated intervention. Our empirical studies reveal that the warning decision correctly mimics human driver decisions.**

I. INTRODUCTION

Most urban accidents occur at intersections. As a result, roundabouts have become highly popular (at least in Europe), since the number of hazard points is significantly smaller than on classical intersections [6]. Assuming right lane driving, the major risk is that a driver entering the roundabout overlooks traffic from the left side, as shown in Fig. 1. Additionally, roundabouts can lead to a higher traffic throughput.

However, in contrast to crossroads, roundabouts exhibit many different complex (non-straight) driving routes that other traffic participants can take.

Unfortunately, this causes problems for today's collision avoidance systems that usually assume straight motion of both the ego-vehicle and the opponent [2]. A short glance on Fig. 2 reveals that it is challenging to predict if an incoming car will leave the roundabout. In the given example, the motion vector of critical vehicle B intersects with the motion vector of our ego-vehicle E only in the very last moment.

Accordingly, it is our goal to develop a driver assistance system that helps to reduce the risk of accidents in such scenarios. To the best of our knowledge, this traffic situation has not been investigated so far. Since today high-quality cameras come at a very low price, we aim at a vision-based solution. Furthermore, we assume that the car knows from a map (e.g. navigation) that the driver approaches a roundabout.

Creating such a system requires the detection of other traffic participants and the estimation of their pose and motion state in order to compute the potential risk level.

For many similar problems, stereo-vision has already proven a powerful solution. For instance, Barth et al. [2] have shown that pose, size, and the full motion state (including acceleration and yaw rate) can be robustly measured by tracking oncoming vehicles.

Fig. 1: A typical traffic situation at an urban two lane roundabout.

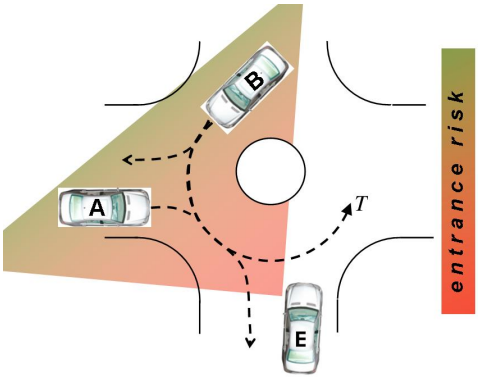


Fig. 2: The entrance risk at roundabouts grows with descending distance. The ego-vehicle E has to consider that both cars (A and B) will pass. Altogether, this is a very challenging situation for the environment perception.

Our recognition system uses preprocessing steps already developed for forward looking cameras. The depth analysis is based on Semi-Global Matching (SGM) using a real-time FPGA implementation [7], [8]. The detection of moving obstacles utilizes the so called Dynamic Stixel-World, a compact three-dimensional scene representation recently introduced by Pfeiffer et al. in [9], [10].

Based on this data, an object clustering is carried out. Subsequently, this information is used to compute a time-to-contact measure that allows to decide whether a safe entrance into the roundabout is possible.

The remaining paper is organized as follows: Section 2 gives a detailed overview about the addressed challenges at roundabout scenarios. The algorithms used to obtain the input data are sketched in Section 3. This also includes a brief

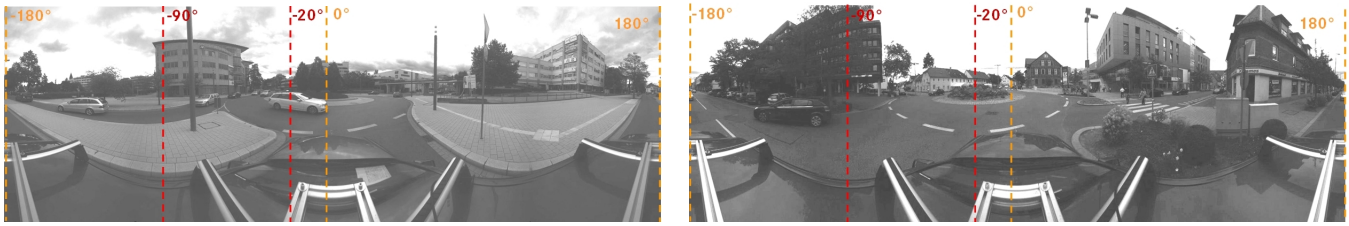


Fig. 3: Images of a spherical 360° camera showing typical urban roundabouts that are used for our field of view analysis. The painted horizontal lines represent the viewing direction of the ego-vehicle. For the core application of tracking other vehicles in roundabouts, we focus on the field of view from -90° to -20° relative to the ego-vehicle (red lines).

overview of the Stixel scene representation. Then, Section 4 describes the clustering of Stixels to objects and the time-to-contact computation is presented in Section 5. Finally, results are given in Section 6. We show different tracking results in roundabouts and compare the decision of our situation analysis with the decision of test persons. As it turns out, our system mimics a careful and defensive driver.

II. Problem Statement

A typical roundabout scenario in an urban environment is shown in Fig. 2. While the ego-vehicle E is waiting for a riskless entrance, vehicle B is driving inside the roundabout. At the same time, vehicle A is about to enter the circle as well. Altogether, this is a challenging situation, because it allows to evolve in many very different ways.

In a preliminary investigation we recorded traffic situations at about 50 different roundabouts. In the lab, the sequences were shown to different test persons who were asked to decide if they would enter the inner circle.

The time that another vehicle needs until it is in front of the ego-vehicle is called time-to-contact (TTC). According to the behavior of the test persons, a realistic TTC is in the range of 2-2.5 s. Usually, they considered other vehicles as relevant obstacles if they were in the lower left quarter of the circle.

A short calculation reveals the reason for this insight: on roundabouts with diameters of about 20 m people drive with 6-7 $\frac{m}{s}$. Accelerating up to this speed from a complete stop at the entrance takes about 2-3 s. Therefore, it is typically safe to enter the roundabout as long as no car is driving with that speed in the mentioned area.

However, there is one exception we have to take into account: vehicle B might have higher speed when it enters the roundabout. This forces us to extend the area we have to check for oncoming traffic for about the length of a car. Depending on the used sensor, one has to add a few more meters for reliably estimating the motion state of cars in the roundabout by tracking.

That means, if we want to track oncoming traffic participants until they pass in front of us or make a right turn, we require a field of view of at least 70°. This is in accordance with the evaluation of the image data that we took with a spherical 360° camera system (see Fig. 3).

As can be seen in Fig. 4, about 95 % of all potentially relevant objects are located between -20° and -90° with respect to the horizontal viewing direction of the ego-vehicle. Based on this data, we decided to work with 80° lenses looking at -50° to the left. The base line of our stereo camera system is 35 cm. It is worth mentioning that this set-up also allows surveillance at standard intersections.

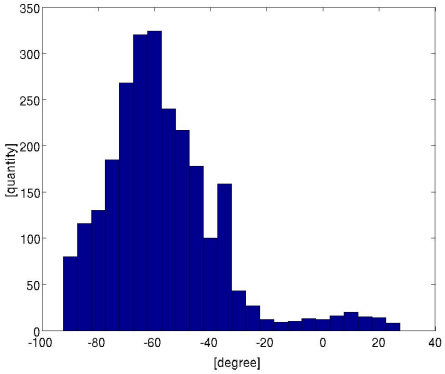


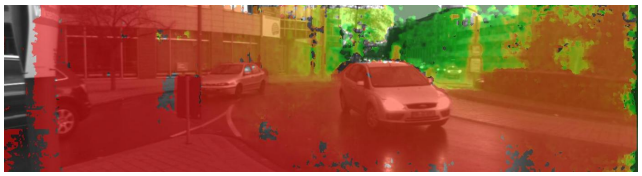
Fig. 4: Distribution of the viewing directions at approaching vehicles with respect to the heading direction of the ego-vehicle. It is shown that about 95 % of all observed vehicles move in the field of view from -20° to -90°.

III. The Dynamic Stixel World

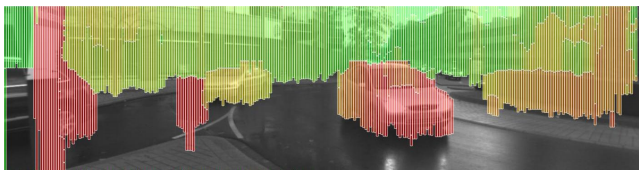
Detecting vehicles passing through roundabouts is achieved by relying on the Stixel representation proposed by Pfeiffer et al. [9] , [10].

A single Stixel is defined as a vertically oriented rectangle with a fixed width in the image (e.g. 5 px) and a variable height. Every object within the image is approximated by a set of adjacent Stixels. This way, Stixels allow for an enormous reduction of the raw input data, e.g. approximately 400,000 disparity measurements from a 1024×440 px stereo image pair are reduced to a few hundred Stixels only. At the same time, Stixels give easy access to the most task-relevant information such as free space and obstacles and thus effectively bridge the gap between low-level (pixel-based) and high-level (object-based) vision.

Stixels are extracted from a stereo image pair in two steps: the stereo computation, e.g. using Semi-Global Matching stereo (SGM) [7], [8], and the actual Stixel computation. According to [10], the three-dimensional scene is segmented into two different class types, namely ground and object. Both are expected as planar surfaces. The difference lies

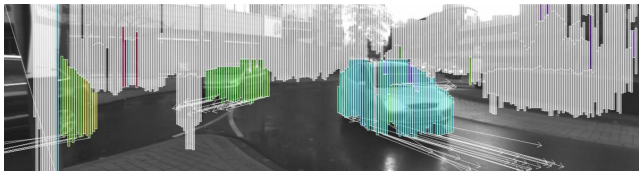


(a) SGM stereo input



(b) Static Stixel representation

(c) KLT-feature tracks



(d) Stixels with motion state

Fig. 5: An exemplary roundabout is shown. The given images denote the different steps when computing Stixels for the current scenario. Fig. (a) shows the dense stereo input obtained from using SGM, Fig. (b) shows the corresponding static Stixel representation. Fig. (c) shows the optical flow input used for tracking Stixels over time and Fig. (d) shows the final Stixel result with motion estimates.

in their orientation: ground is expected as horizontal while object is assumed as vertical with a constant depth. The segmentation is regularized by a set of physically motivated world model priors, such as gravity and ordering constraints. This way, the segmentation task leads to a typical maximum a posteriori (MAP) estimation problem. Solving for the most likely and thus optimum segmentation is achieved through the use of dynamic programming. The SGM result and the Stixel result are depicted in Fig. 5a and 5b.

Up to this point, the Stixel representation only describes the current three-dimensional world geometry (in both the image and in 3D). However, deciding whether a roundabout is occupied by other moving vehicles or not also requires additional velocity information.

For this purpose, the Stixel based tracking scheme proposed in [9] is chosen. Besides using stereo data, this scheme additionally requires optical flow information (see Fig. 5c) as well as the own vehicle's odometry. To this end, the first is computed by using the well-known feature-based KLT-tracker [11] and the latter is extracted by using visual odometry [1].

For estimating the motion properties of other objects, the obtained input data has to be combined properly. This is achieved by following the 6D-Vision principle suggested by Franke et al. [5]. This scheme uses Kalman filtering [13] to estimate both the position and velocity of three-dimensional point feature. The result is combined in a rich 6-dimensional state vector. However, since in our considered scenarios all relevant objects are expected to move earthbound, this state vector allows to be reduced to 4D, namely the longitudinal and lateral state components, such that $\overline{X} = (X, Z, \dot{X}, \dot{Z})^T$.

As a result, precise motion information is available for every Stixels independently. Stixels enriched with motion information are defined as dynamic Stixels. The Stixel-based tracking result for the exemplary scenario is depicted in Fig. 5d.

IV. Clustering Process

The goal of the following steps is to reliably estimate the position and velocity of relevant vehicles in roundabouts. For this purpose, independent Stixels $s_i \in \{1,..,\mathrm{I}\}$ are grouped to only a small number of clusters c_k with $k \in \{1,..,\mathrm{K}\}$ and $\mathrm{K} << \mathrm{I}$.

A successful clustering is based on well-considered geometrical and physical conditions. To this end, the following assumptions are made:

- **minimum number of Stixels**: due to their horizontal expansion, objects are represented by a minimum number of Stixels *minStix*.
- **geometrical characteristics**: the euclidean distance between two Stixels s_i and s_j is a relevant criterion for the spatial separation.
- **physical characteristics**: Stixels representing the same object have a uniform velocity and driving direction.

We applied a real-time clustering procedure which is based on the DBSCAN algorithm [4]. The key idea is that an arbitrary Stixel s_i of a cluster c_k has at least a minimum number of Stixel neighbors *minNeigh* within a given neighborhood threshold ε. In our approach we assume that the Stixel density within a cluster is considerably higher than outside of a cluster. Thus, clusters with $minStix < minNeigh$ are flagged as noise (Fig. 6, left).

The euclidean distance d is frequently used as a neighborhood criterion, but the DBSCAN [4] allows any kind of cost functions.

For our application it is not sufficient to use the euclidean distance $\mathrm{d} = \mathrm{dis}(s_i, s_j)$ as the only neighborhood constraint. This is because back-to-back driving cars with different driving directions tend to be merge to one object.

To this end, a second neighborhood criterion is defined which is the angle ϕ between the two motion vectors $u_i =$

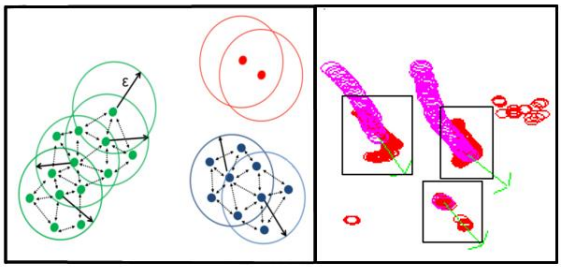


Fig. 6: Illustration of the DBSCAN. Left: the green and blue points represent two different clusters. The red points are considered as outliers. Right: the clustering results of the scene in Fig. 5d is shown. The black boxes represent the cluster positions and the green arrows describe their driving directions.

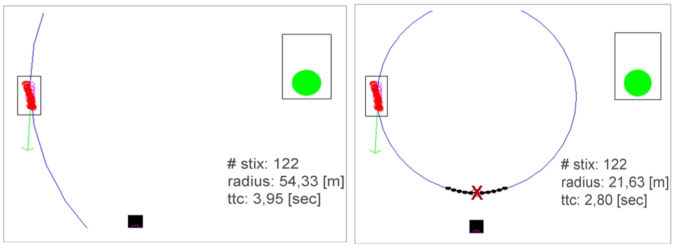


Fig. 7: The circle fit with and without the radius constraint after 10 frames of tracking. At the unconstrained solution the collision time takes too much time which is not consistent with the real situation. The black rectangle represent the ego position. The black dots are the predicted positions of the incoming vehicles. The red cross is the calculated collision point.

$(\dot{X},\dot{Z})_i^T$ and $u_j = (\dot{X},\dot{Z})_j^T$ of the Stixels s_i and s_j:

$$\phi = arccos\left(\frac{< u_i, u_j >}{\|u_i\| \ \|u_j\|}\right) \tag{1}$$

A drawback of the DBSCAN algorithm is its quadratic complexity $O(n^2)$, where in our case n equals the number of Stixels I. To reduce this burden, we use the modified l-DBSCAN [12] which is a hybrid clustering method with a runtime complexity of $O(n)$. Its key idea is to start with a coarser clustering of the complete data set. Each cluster is represented by its leader point l. Then, a fine clustering is carried out for which only those leader points are considered. Finally, after the grouping process, each cluster is represented by its mean position X_k, Z_k and its mean velocity $\dot{X}$ and $\dot{Z}$.

For a better understanding, the clustering result of the scene depicted in Fig. 5d is given in Fig. 6.

V. The Time-to-Contact-Computation

In the following, the estimation of the TTC_k is extracted which is performed at each time-step and for each detected cluster k. The goal is to predict whether a safe entrance into the roundabout is possible or not.

With the help of the mean points of each cluster and a nearest neighbor criterion the driven trajectories are estimated. The motion trajectory of the incoming vehicle is approximately represented by a circular shape. For this purpose, a circle is fitted to the driven mean positions $[X_k, Z_k]_t$ as soon as a cluster c_{kt} is steadily observed over time $t \in \{1,...,\mathrm{T}\}$.

The method of [3] is used for the circle estimation which is based on a least square fit. Hereby, the sum of the squares

$$F = \sum_{t=1}^{T} {v_t}^2 \tag{2}$$

is minimized where v_t is the error distance function defined as:

$$v_t = \sqrt{(X_{kt} - a_k)^2 \ + \ (Z_{kt} - b_k)^2} - R_k \ , \tag{3}$$

with the circle radius R_k and the circle center $[a,b]_k$. For a robust estimation the circle radius R_k is constrained. Therefore it is assumed that digital maps (e.g. navigation systems) will provide this geometrical information. An example for the circle estimation is shown in Fig. 7.

Furthermore the length of the circular arc ca_k which a vehicle will drive to a possible collision point is defined by the circle angle α. This angle α is calculated from the current vehicle position, the circle center and the position of the ego-vehicle. The estimation of ca_k is straightforward:

$$\mathrm{ca}_k = \pi \ R_k \ \frac{\alpha}{180^\circ} \ . \tag{4}$$

Finally, the TTC_k is determined by the estimated velocity $v_k = \sqrt{(\dot{X}_k^2 + \dot{Z}_k^2)}$ and the ca_k:

$$\mathrm{TTC}_k = \frac{\mathrm{ca}_k}{v_k} \ . \tag{5}$$

If TTC_k is below a given TTC_k threshold the system advises not to enter the roundabout. The TTC_k is updated at each time step which is exemplary shown in Fig. 9.

VI. Results

For our experiments we evaluate video material of typical roundabout scenarios recorded at rush-hour traffic. The used stereo camera system has 1400×1024 px image sensors with 80° FOV lenses and a focal length of 740 px. The images are cropped to 1400×400 px to focus on the relevant scene content. The dynamic Stixel algorithm, the clustering process and the TTC computation run on the CPU in real-time.

A. Results of the vehicle tracking and the TTC computation

Fig. 10 shows different tracking samples of a 80 s sequence of the two lane roundabout from Fig. 1. The incoming cars are recognized at an average distance of approximately 25 m. It is visible from the processed data that vehicles which passed us nearly drove a circular arc. Vehicles that turned off typically had an approximately linear driving path.

Due to side-by-side driving, in some cases, tracks of covered vehicles were lost and a new cluster re-initialization had to occur. Scattered outliers are observed at distances of approximately 20-25 m which, however, have shown no negative influence on the TTC computation.

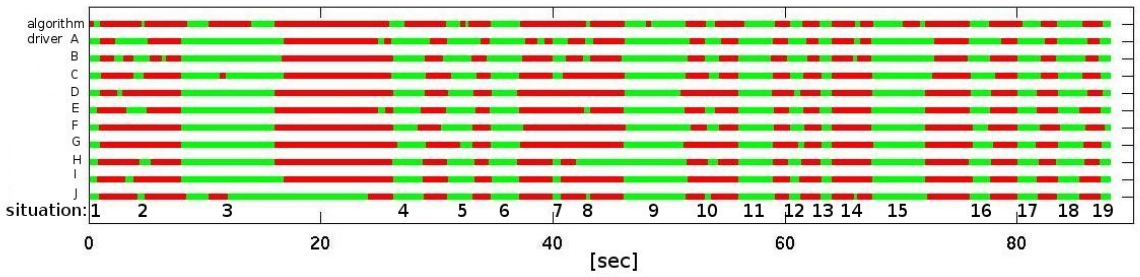


Fig. 8: The estimated stop and go phases (red and green) of our algorithm compared to the behavior of test persons for a 90 s sequence of a typical urban roundabout. In 13 of 19 independent traffic situations (1, 4-6, 9, 11-13, and 15-19) the algorithm decision closely corresponded to the human behavior. In two cases (situation 8 and 14), the algorithm only matched to five or less participants. Again, in four situations (3, 5, 9, and 15) the algorithm switched to red for a few frames while some of the participants decided otherwise. In these cases, the test persons recognized that the vehicles turned off.

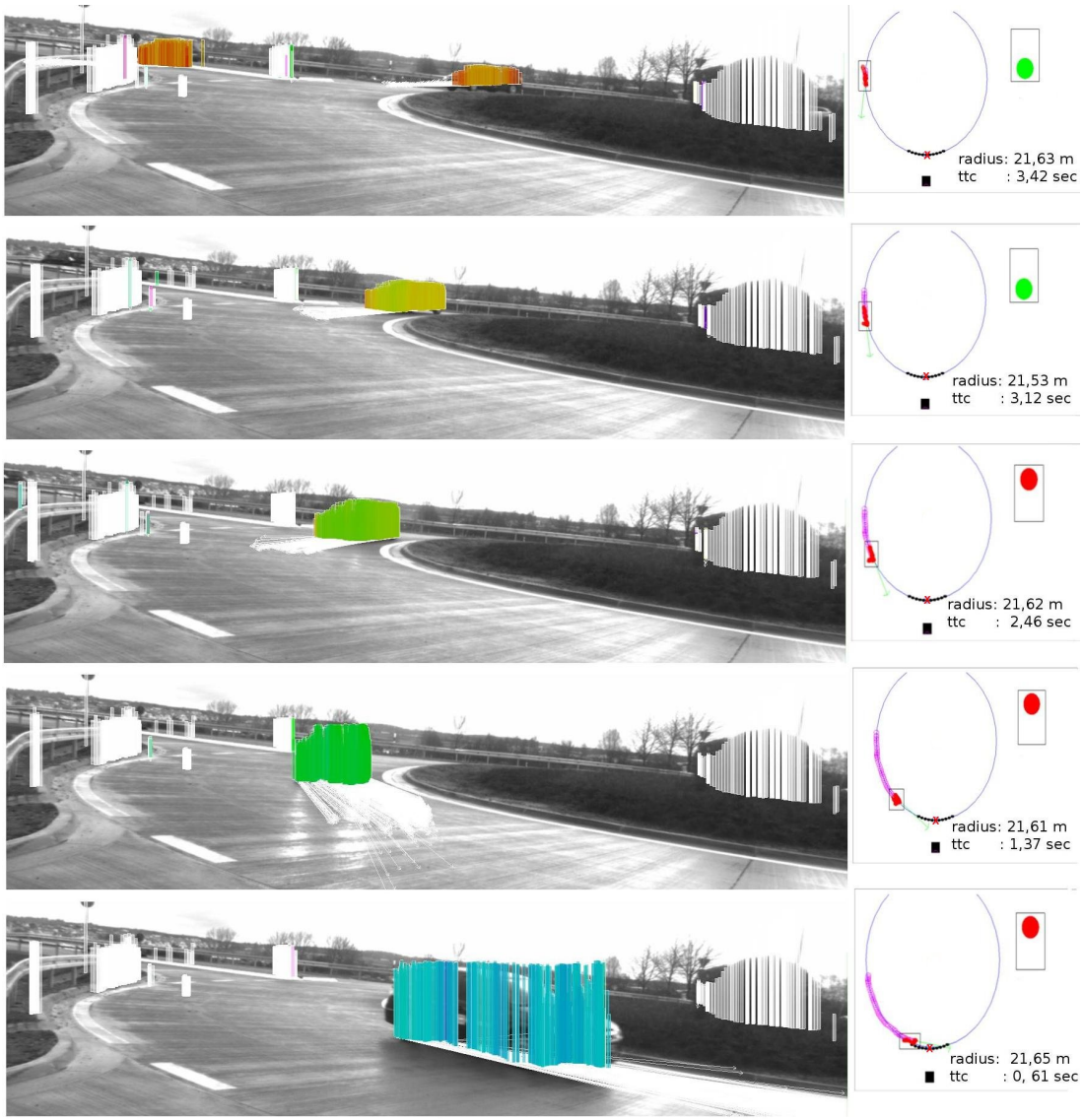


Fig. 9: A sequence set of a typical scene with an incoming vehicle at a roundabout. The images are illustrated together with the corresponding dynamic Stixel representation. On the right side the results of the clustering process and the TTC computation is shown for each scene.

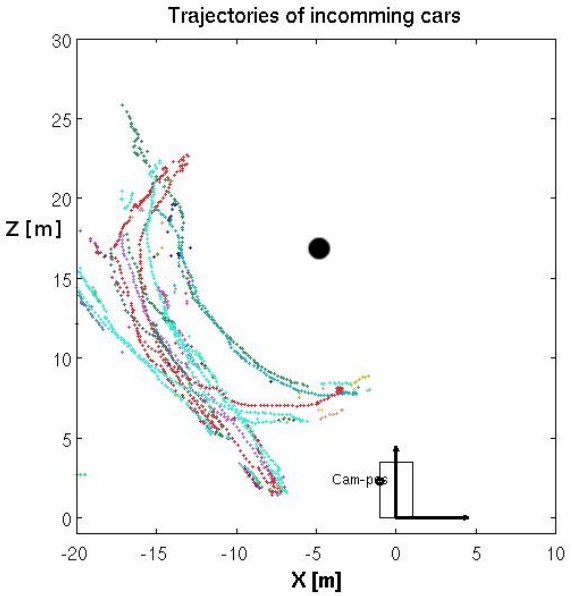


Fig. 10: The estimated trajectories of our tracking algorithm for a sample 80 s sequence. Roughly, 50 percent of all incoming vehicles drove a circular arc and passed the ego-vehicle (black rectangle). All other vehicles were leaving the roundabout.

Fig. 9 shows the dynamic Stixel representation with the TTC computation of a typical roundabout scene where an incoming car passed the ego-vehicle. The color value of the Stixels and the drawn arrows represent the direction the Stixel moves with respect to our ego-vehicle. The color saturation encodes the speed as well. In the third scene the estimated TTC is below the given threshold of 2.5 s. For this reason, the advise is not to enter the roundabout, as indicated by the red light.

B. Evaluation with test persons

For an evaluation of our TTC algorithm estimated stop and go phases have been compared with the driving behavior of test persons. They were shown recorded scenes and had to mark those time windows where they would enter the roundabout.

Fig. 8 compares the results of each test person with our algorithm for a 90 s sequence with 19 independent traffic situations. Note that this is just an extract of our evaluation with about 50 independent roundabout scenarios.

Generally, the red-green-phases of the algorithm corresponds to the phases of the test persons. Thereby, the geometrical assumptions of the driving behavior and the TTC threshold of 2.5 s are confirmed. In contrast to the test persons the TTC computation can not "recognize" turning off vehicles, such as shown in scene five (about 0:35 s) and scene seven (about 0:40 s). Apparently, our decision strategy shows the most defensive but also safest driving behavior.

VII. CONCLUSION

For the first time, a stereo-based time-to-contact computation for right of way situations at roundabouts was presented. For this purpose, urban roundabouts were observed to configure an optimal stereo camera setup.

Dense disparity images are used to compute the dynamic Stixel World which is a compact three-dimensional environment representation for urban traffic situations. This work proves the power of the dynamic Stixels which support our processing steps perfectly.

A well known clustering method was used to group independent dynamic Stixels representing the same object. This procedure allows reliable tracking of incoming vehicles at urban roundabouts. In order to handle such complex situations properly, we assume that all tracked vehicles drive on a circular arc. This has proven a defensive but safe assumption.

For a reliable time-to-contact computation a robust circle fit method was used which is supported by additional geometric constraints.

The system's estimated stop and go phases have been compared to the driving behavior of 10 different test persons. According to these tests, it has performed no misjudgment of the current right-of-way situations.

REFERENCES

[1] Hernán Badino. A robust approach for ego-motion estimation using a mobile stereo platform. In 1st *International Workshop on Complex Motion, IWCM*, Günzburg, Germany, October 2004. Springer.

[2] Alexander Barth and Uwe Franke. Tracking oncoming and turning vehicles at intersections. In *IEEE Conference on Intelligent Transportation Systems (ITSC)*, pages 861–868, Madeira Island, Portugal, September 2010.

[3] Nikolai Chernov. *Circular and linear regression: Fitting circles and lines by least squares*. CRC Press, Inc., Boca Raton, FL, USA, 2010.

[4] Martin Ester, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *KDD*, pages 226–231, 1996.

[5] Uwe Franke, Clemens Rabe, Hernán Badino, and Stefan Gehrig. 6d-vision: Fusion of stereo and motion for robust environment perception. In *German Association for Pattern Recognition (DAGM)*, Vienna, Austria, September 2005.

[6] Behörde für Stadtentwicklung und Umwelt. Planungshinweise für Stadtstrassen: Knotenpunkte & Kreisverkehre, Freie und Hansestadt Hamburg, Germany, 2009.

[7] Stefan Gehrig, Felix Eberli, and Thomas Meyer. A real-time low-power stereo vision engine using semi-global matching. In *International Conference on Computer Vision Systems (ICVS)*, 2009.

[8] Heiko Hirschmüller. Accurate and efficient stereo processing by semi-global matching and mutual information. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 807–814, San Diego, CA, USA, June 2005.

[9] David Pfeiffer and Uwe Franke. Efficient representation of traffic scenes by means of dynamic Stixels. In *IEEE Intelligent Vehicles Symposium (IV)*, pages 217–224, San Diego, CA, USA, June 2010.

[10] David Pfeiffer and Uwe Franke. Towards a global optimal multi-layer Stixel representation of dense 3D data. In *British Machine Vision Conference (BMVC)*, Dundee, Scotland, August 2011. BMVA Press.

[11] Carlo Tomasi and Takeo Kanade. Detection and tracking of point features. Technical report, School of Computer Science, Carnegie Mellon University, April 1991.

[12] P. Viswanath and Rajwala Pinkesh. l-dbscan : A fast hybrid density based clustering method. In *ICPR*, pages 912–915, 2006.

[13] Greg Welch and Gary Bishop. An introduction to the kalman filter. Technical report, Department of Computer Science, University of North Carolina at Chapel Hill, 1995.