Day and Night-Time Drive Analysis using Stereo-Vision for Naturalistic Driving Studies

AALBORG UNIVERSITY DENMARK

Mark P. Philipsen, Morten B. Jensen, Ravi K. Satzoda, Andreas Møgelmose, Mohan M. Trivedi & Thomas B. Moeslund*

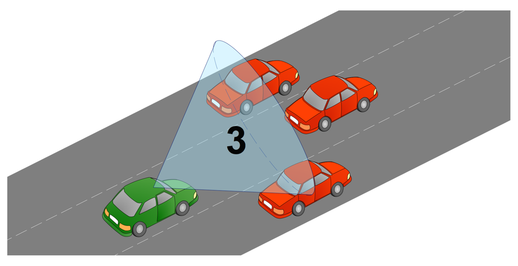
UC San Diego Jacobs School of Engineering

CVRR Computer Vision and Robotics Reserach Laboratory

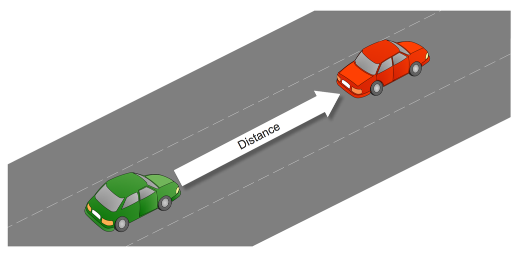
LISA: Laboratory for Intelligent & Safe Automobiles

Introduction

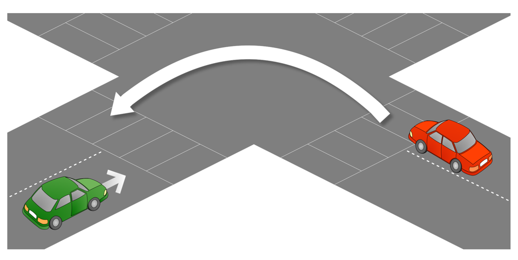
We introduce a stereo-vision based system that enables automatic event detection based on detection and tracking of vehicles in scenarios that usually would be highly problematic for monocular detectors. The purpose is to provides insight into patterns and behaviors of drivers during near-crashes and crashes. The proposed system will be limited to only handling a handful of events, which especially benefit from the extra dimension gained with stereo-vision. The considered events are illustrated in Figure 1.



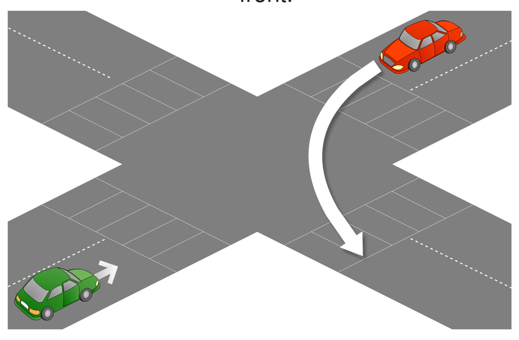
(a) Average number of cars in front of ego-vehicle.



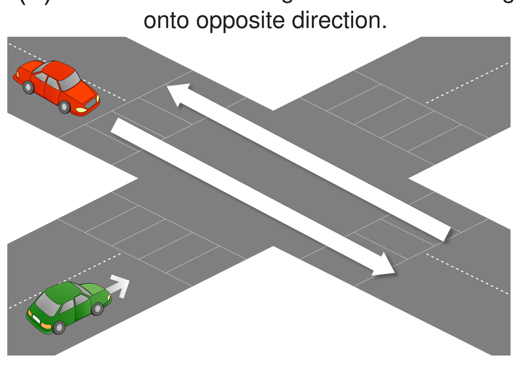
(b) Distance to rear-end of vehicle directly in front.



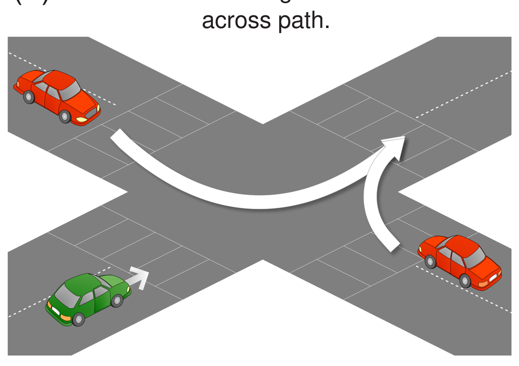
(c) Other vehicle entering intersection - turning onto opposite direction.



(d) Other vehicle entering intersection - left turn across path.



(e) Other vehicle entering intersection - straight across path.



(f) Other vehicle entering intersection - turning onto same direction.

Figure 1: Automatically detectable critical events. Green car is ego vehicle moving towards intersection. Red cars are other vehicles [1].

Monocular systems usually have problems dealing with occlusions since they detect based on appearance, something which require a large amount of training data and many not still be able to deal with all of the many possible viewpoints and light conditions.

Our contributions are:

- Using stereo-vision for automatic event detection in both day and nighttime, with focus on intersections (Figure 1a, 1c, 1d, 1e, 1f).
- Introducing a new event: Average distance to vehicles directly in front of the ego vehicle. (Figure 1b).

Related Work

In [2], monocular computer vision and information from the CAN bus is used to automatically detect 23 drive events, including lane position, vehicle localization within lanes, vehicle speed, traffic density, and road curvature. In [3], a system is developed for automatic labeling of driver behavior events relying on overtaking and receding vehicle detection. This type of system is categorized as Looking-In and Looking-Out (LiLo), which is discussed in depth in [4]. LiLo fits well with using multiple inputs to understand the driver's behavior.

[5] introduced the accurate and efficient Semi-Global Matching (SGM) algorithm from finding stereo correspondence. SGM make use of epipolar geometry, and in most cases a set of rectified stereo images. Horizontal lines in the images are used as a scan lines, matches are then found for all possible disparities in a 1D disparity search. A match for a pixel in the left image is found in the right image by searching through the corresponding horizontal row and locating the most similar block to a reference block around the original pixel in the left image. The offset between these pixels is known as the disparity, which, if correctly matched, is directly related to the distance to the corresponding object. A final match it based on the outcome of a smoothed path cost which is calculated in a number of directions for each disparity. In the same paper, the LR-RL-consistency check is proposed for reducing noise in the calculated disparity image.

In [6], the so-called V-disparity is generated and used for separating objects from the ground/road surface. The V-disparity examines the vertical coordinates in a (u,v) image coordinate system and is constructed using a disparity map from, e.g. the SGM algorithm. What is especially histograms are calculated for each row in the disparity. Significant surfaces in the disparity map will then show up as lines in the V-disparity.

Acknowledgment

Thanks to our colleagues at the LISA-CVRR Laboratory.

Proposed System

A Bumblebee XB3 stereo camera is used to acquire **stereo images** with frame rate of 16 FPS in a resolution of 1280x960 each. For generating the **disparity map**, OpenCV's CPU implementation of SGBM is used, it is based on the stereo correspondence algorithm proposed in [5]. **Noise** is removed using a modified LR-RL consistency check where consistency conflicts are solved by keeping the lowest disparity value, instead of rejecting the pixel entirely. An additional temporal consistency check and a monocular color check is also introduced. The **road surface** is found by searching for the most significant line in the V-disparity [7] using RANSAC. Additionally, the line parameters are filtered using a Kalman filter to smooth out faulty road surface detections. Following this the disparity pixels belonging to the road surface or noise below is removed and what remains is provisionally regarded as objects of interest. These objects are projected to 3D using the camera's properties, which result in a 3D **point cloud** representations of the segmented objects. The acquired point clouds are post processed using a band-bass filter to remove near and distant points, downsampled using a voxel grid, and outlier removal. **Clusters** are found by creating a k-d tree, which organize points according to distance to neighbors.

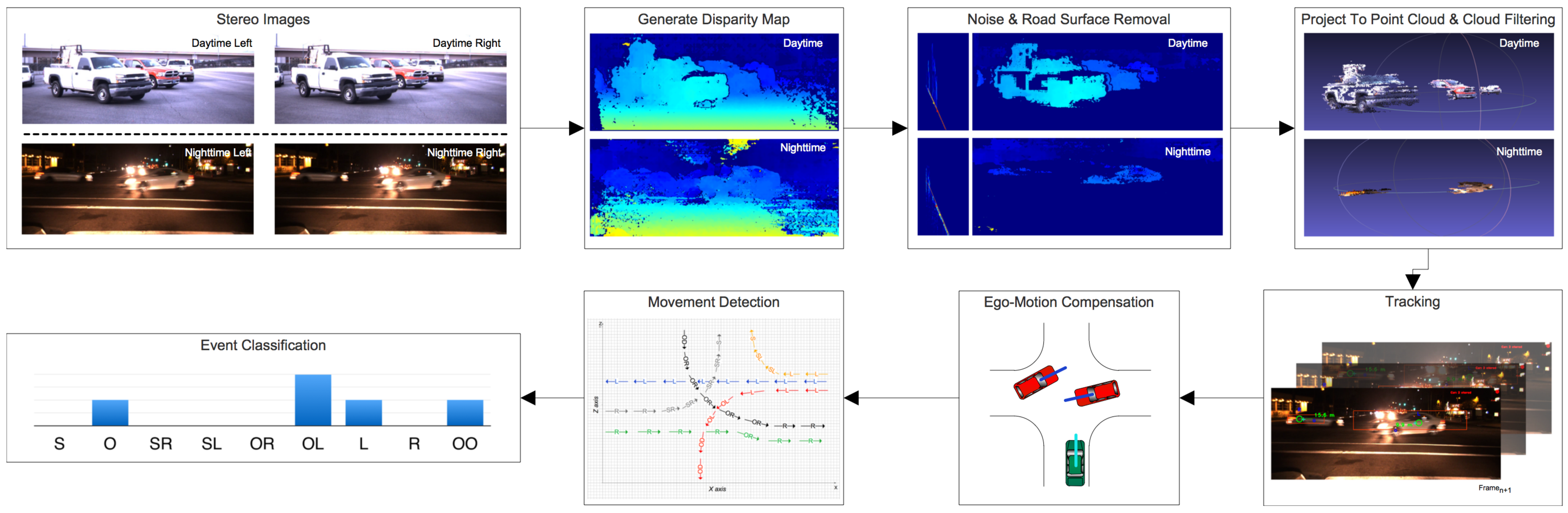


Figure 2: Critical event detection process flow.

The clusters' center points are used for nearest neighbor **tracking** between frames and for determining the distance from ego-vehicle to detected vehicles. For **ego-motion compensating** while the ego-vehicle is approaching an intersection, we utilize *LIBVISO2: C++ Library for Visual Odometry 2* [8]. Events are detected by looking at the **movement** history of other vehicles. For all detected vehicles, individual frame to frame movements are categorized to form a histogram of movements for the sequence, which is used for **event classification**. An overview of the system is shown in Figure 2. The final system output is an event report.

Results

The proposed system is evaluated on 4,992 day and 3,933 night time frames. In Table 1 the results are seen, where GT is the *ground truth* of events manually labeled and SO is the *system output*. The proposed system have an overall precision of 0.78 and recall of 0.72.

Table 1: Summary of event report analysis. The syntax for the results are [SO/GT]. P and R are abbreviations for precision and recall, respectively.

Drive Behavior Event	Daytime	Nighttime	P	R
From Left To Right - straight across path	35/32	5/19	0.95	0.63
From Right to Left - straight across path	45/34	11/33	0.87	0.67
Left turn across path	5/5	20/1	0.75	1
Turn onto opposite dir.	32/37	41/15	0.68	0.93
Short turn onto same dir.	7/5	9/5	0.63	1
Long turn onto same dir.	1/16	1/8	1	0.09
Avg. number of cars	1.67/1.74	1.6/1.3	NA	NA
Avg. distance to car	8.73 m	10.98 m	NA	NA

In Figure 4 an example of the output from the system is seen. The new NDS event introduced with this work is the distance to vehicle in front of the ego-vehicle, which is measured using stereo-vision. In Figure 4 an example of this is seen where the distance of 12.6 meters is written in blue just above the center. This distance measure is also visible in Figure 3.

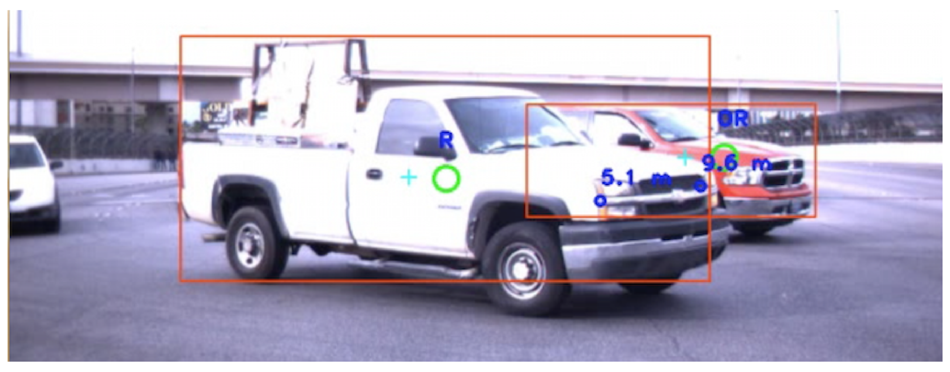


Figure 3: Partially occluded vehicle detected in a left turn.

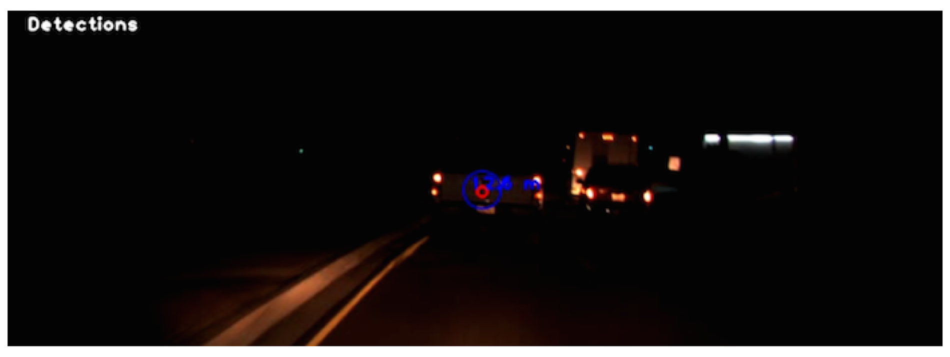


Figure 4: Distance to vehicle in front of ego-vehicle.

Concluding Remarks

The use of stereo-vision is considered beneficial, especially in scenarios with partly occluded cars, in such cases most monocular systems will fail. Distance information from the ego vehicle to the surroundings provides useful information with regards to drive patterns before and during a crashes.

Our work can be summarized as the following bullets:

- Introduction of a novel stereo based critical event analysis approach.
- Experimental analysis shows very promising detection, trajectory and event classification rates.
- Ongoing research involves extensive experimental validation and a day and night time critical event detection module for public use.

References

[1] M. P. Philipsen, M. B. Jensen, R. K. Satzoda, M. M. Trivedi, A. Møgelmose, and T. B. Moeslund, "Night-time drive analysis using stereo-vision for data reduction in naturalistic driving studies," in *Intelligent Vehicle Symposium (IV)*, 2015.

[2] R. Satzoda and M. Trivedi, "Drive analysis using vehicle dynamics and vision-based lane semantics," *Intelligent Transportation Systems(ITSC)*, pp. 1–10, 2014.

[3] ——, "Overtaking and receding vehicle detection for driver assistance and naturalistic driving studies,," in *ITSC*, 2014, pp. 697–702.

[4] M. M. Trivedi, T. Gandhi, and J. McCall, "Looking-in and looking-out of a vehicle: Computer-vision-based enhanced vehicle safety," *ITSC*, vol. 8, pp. 108–120, 2007.

[5] H. Hirschmuller, "Accurate and efficient stereo processing by semi-global matching and mutual information," in *Computer Vision and Pattern Recognition (CVPR)*, 2005, pp. 807–814.

[6] R. Labayrade, D. Aubert, and S. S. Ieng, "Onboard road obstacles detection in night condition using binocular ccd cameras," in *in ESV*, 2003.

[7] R. Labayrade, D. Aubert, and J.-P. Tarel, "Real time obstacle detection in stereovision on non flat road geometry through" v-disparity" representation," in *IV*. IEEE, 2002, pp. 646–651.

[8] A. Geiger, J. Ziegler, and C. Stiller, "Stereoscan: Dense 3d reconstruction in real-time," in *IV*, 2011.