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

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

AALBORG UNIVERSITY DENMARK

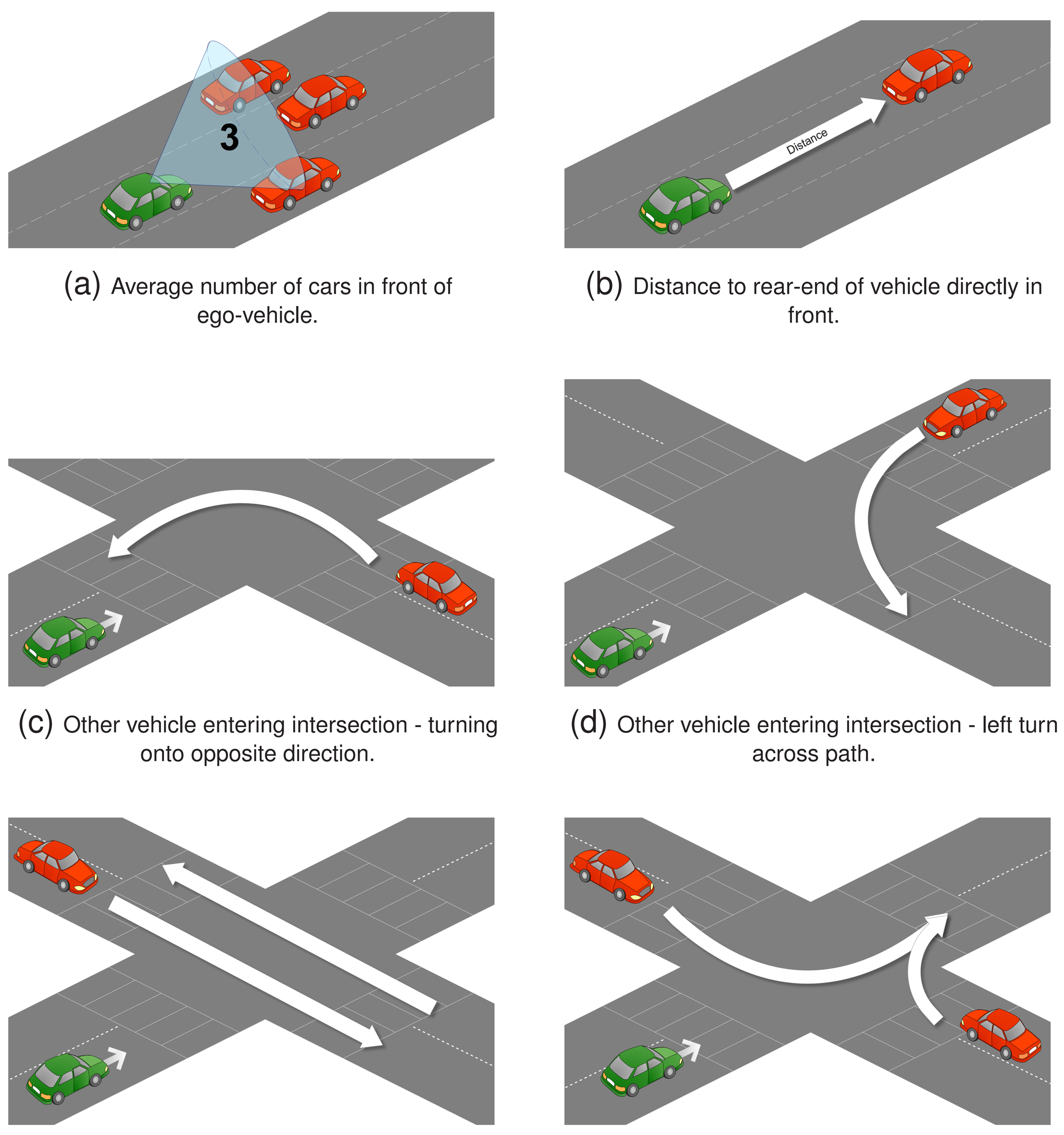
UC San Diego Jacobs School of Engineering

CVRR Computer Vision and Robotics Reseach 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).

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 *IEEE, Intelligent Vehicle Symposium*, 2015.

[2] F. Guo and Y. Fang, "Individual driver risk assessment using naturalistic driving data," *Accident Analysis & Prevention*, vol. 61, pp. 3 – 9, 2013, emerging Research Methods and Their Application to Road Safety Emerging Issues in Safe and Sustainable Mobility for Older Persons The Candrive/Ozcandrive Prospective Older Driver Study: Methodology and Early Study Findings.

[3] B. G. Simons-Morton, M. C. Ouimet, Z. Zhang, S. E. Klauer, S. E. Lee, J. Wang, R. Chen, P. Albert, and T. A. Dingus, "The effect of passengers and risk-taking friends on risky driving and crashes/near crashes among novice teenagers," *Journal of Adolescent Health*, vol. 49, pp. 587 – 593, 2011.

[4] S. P. McEvoy, M. R. Stevenson, and M. Woodward, "The impact of driver distraction on road safety: results from a representative survey in two australian states," *Injury prevention*, vol. 12, pp. 242–247, 2006.

[5] R. Satzoda and M. Trivedi, "Drive analysis using vehicle dynamics and vision-based lane semantics," *Intelligent Transportation Systems, IEEE Transactions on*, pp. 1–10, 2014.

[6] ——, "Overtaking amp; receding vehicle detection for driver assistance and naturalistic driving studies," in *Intelligent Transportation Systems (ITSC), IEEE*, 2014, pp. 697–702.

[7] M. M. Trivedi, T. Gandhi, and J. McCall, "Looking-in and looking-out of a vehicle: Computer-vision-based enhanced vehicle safety," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 8, pp. 108–120, 2007.

[8] E. Ohn-Bar, S. Martin, and M. M. Trivedi, "Driver hand activity analysis in naturalistic driving studies: challenges, algorithms, and experimental studies," *Journal of Electronic Imaging*, vol. 22, pp. 041 119–041 119, 2013.

[9] S. Sivaraman and M. M. Trivedi, "A review of recent developments in vision-based vehicle detection," *Intelligent Vehicles Symposium (IV), IEEE*, pp. 310–315, 2013.

[10] Z. Sun, G. Bebis, and R. Miller, "On-road vehicle detection: a review," *Pattern Analysis and Machine Intelligence, IEEE*, pp. 694–711, 2006.

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

[12] A. Broggi, A. Cappalunga, C. Caraffi, S. Cattani, S. Ghidoni, P. Grisleri, P. Porta, M. Posterli, and P. Zani, "Terramax vision at the urban challenge 2007," *Intelligent Transportation Systems, IEEE*, pp. 194–205, 2010.

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

[14] C. Rabe, U. Franke, and S. Gehrig, "Fast detection of moving objects in complex scenarios," in *Intelligent Vehicles Symposium, IEEE*, 2007, pp.

Related Work

We look at some resent research published with regards to automatic NDS data reduction as well as some notable work in improving disparity maps and object detection in traffic scenes.

In [2], statistics from near-crashes and crashes are used to identify factors related to driver associated risks, such as age, experience, gender, and demographic. Similar studies are conducted in [3], where results show that near-crash and crash percentage among teenagers were 75 % lower in the presence of adult passengers and 96 % higher among teenagers with risky friends. In [4] NDS are used to quantify distracting activities from e.g. a mobile phones, which result in loss of concentration. Results show that drivers in average are engaged in distracting activity every 6 minutes, something which could result in a near-crash or crash.

Manual data reduction for NDS is quite comprehensive and time consuming, it is therefore desirable to automate it by e.g. applying computer vision to understand the traffic scene. An example of one such study is seen in [5], where 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 [6], a system is developed for automatic labelling 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 [7]. LiLo fits well with using multiple inputs to understand the driver's behavior. An example of Looking-In is [8] where driver behavior with respect to hand activity is evaluated.

In [9], a review of the research conducted since 2005 in both monocular and stereo vision with respect to vision-based vehicle detection is presented. Before 2005, [10] conducted a review of on-road vehicle detection. Most of the research published with regard to vehicle detection is evaluating the methods on data, representing a limited part of the challenges. In [11] the accurate and efficient Semi-Global Matching(SGM) is introduced. 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 with 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 line 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 is directly related to the distance to the corresponding object. In the same paper, the LR-RL-consistency check is proposed for reducing noise in the calculated disparity image.

In [12, 13], 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.

In [14], 6D-vision is introduced, where features are found in the left monocular image and then located in 3D by using the stereo images. For each feature point, a Kalman filter is used to estimate a 6D vector consisting of the 3D position and 3D motion vector. [14] is able to do ego-motion compensation by identifying the static 6D points. The predicted static world points are then compared to the remaining points to isolate the ego-motion. In [15], this work is continued and by using 6D-vision, tracked feature points are represented as stixels, which is a vertical areas of equal disparity. The tracked objects are being classified using prior knowledge of vehicle shapes. Alternatively, objects can be classified using clustering in the disparity map, as seen in [12]. In [16, 17], temporal and scene priors from good conditions are used with the purpose of improving the disparity map in adverse weather conditions, such as night, rain, and snow. Using these priors, the object detection rate improves on a database of 3000 frames including bad weather while reducing the false positive rate.

Proposed System

A Bumblebee XB3 stereo camera is used to acquire **stereo images** with an average rate of 15 FPS in an resolution of 1280x960. For generating the **disparity map**, the OpenCV's SGBM[11] implementation is used. **Noise** is removed using LR-RL consistency check, temporal information, and a monocular color check. The **road surface** is found by searching for the most significant line in the V-disparity[18] using RANSAC. Additionally, the line parameters are filtered using a Kalman filter to smooth out faulty road surface detections. Objects are projected to 3D using the camera's properties, which result in 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 outliers are removed. **Clusters** are found by creating a k-d tree, which organize points according to distance to neighbors.

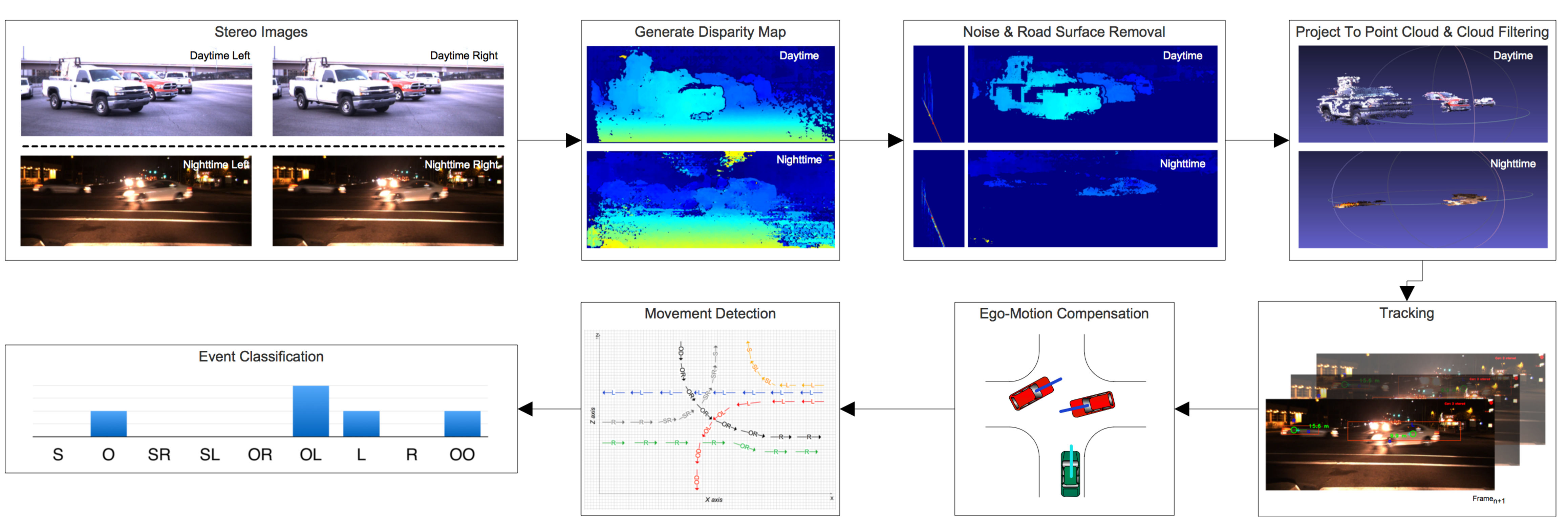


Figure 2: Critical event detection process flow.

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 approaching an intersection, we utilize the *LIBVISO2: C++ Library for Visual Odometry 2* [19]. 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 **event classification**.

An overview of the system is shown in Figure 4. The final 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

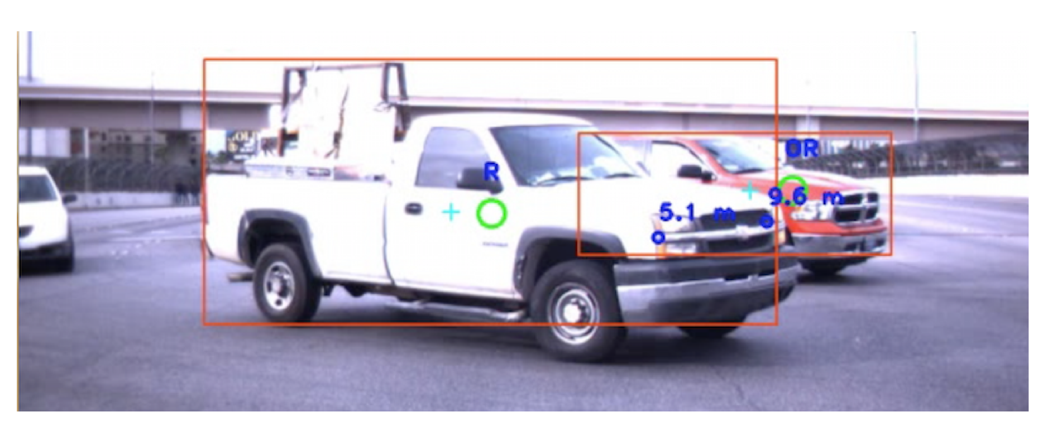


Figure 3: Partially occluded vehicle detected in a left turn.

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.

- 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.

Acknowledgment

Thanks to our colleagues at the LISA-CVRR Laboratory.