

Related Work

We look at some recent 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.

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-pass 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.

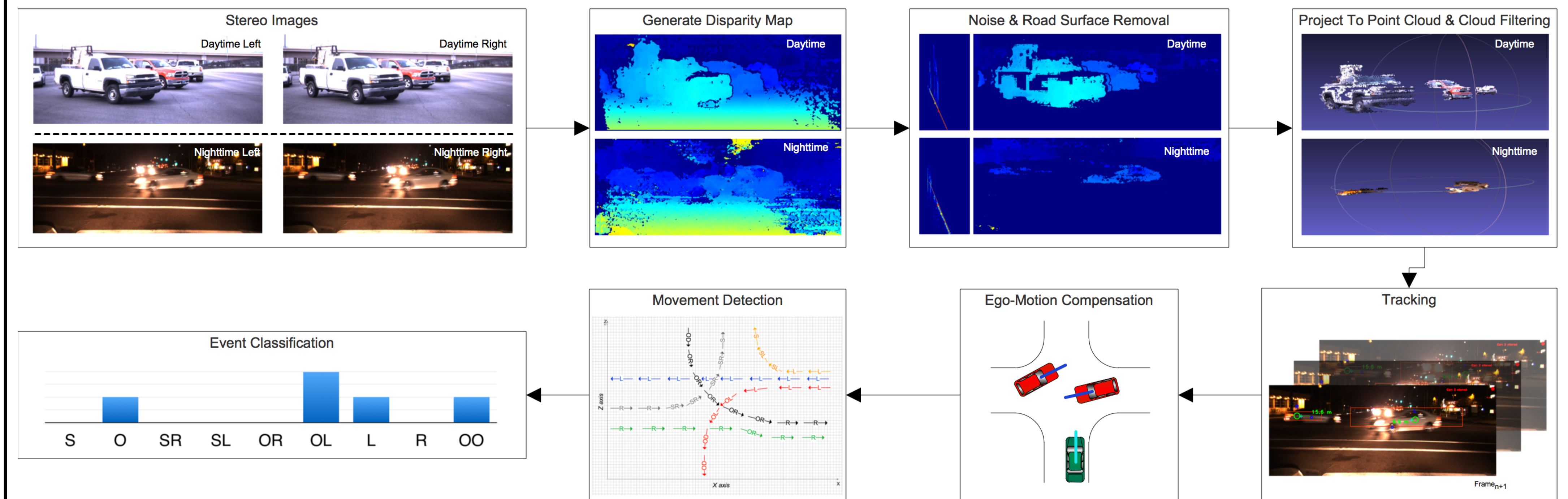


Figure 2: Critical event detection process flow.

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 2. 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
Right - straight across path	35/32	5/19	0.95	0.63
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

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 Ravi K. Satzoda and Andreas Møgelmoose for guidance.