Lane Detection and Estimation from Surround View Camera Sensing Systems

1st Ting Yuan
Shanghai Jiao Tong University
tyuan@sjtu.edu.cn

2nd Wenqi Cao
Shanghai Jiao Tong University
wenqicao@sjtu.edu.cn

3rd Shuqi Zhang
Shanghai Jiao Tong University
steven_zhang@sjtu.edu.cn

4th Kaipei Yang
University of Connecticut
kpyang@uconn.edu

5th Markus Schoen
Daimler AG
markus.schoen@daimler.com

6th Bharanidhar Duraisamy
Daimler AG
bharanidhar.duraisamy@daimler.com

***Abstract*—Autonomous driving poses unique challenges for vehicle environment perception systems. It is highly desirable that we utilize existing vehicle-equipped driver-assistant sensors, without hardware change, to achieve driverless performance. Current product level vehicle surround view camera module (denoted concisely as SVS) is served as a panoramic view visual aid tool for low-automation applications. With proper statistical analysis, the multiple mono-camera information can be very useful for higher vehicle intelligence without significant hardware change. In this study, we focus on lane detection and estimation from a SVS only system. The major difficulty lies in the fact that mono-cameras of the SVS are non-cooperative and essentially of protractor nature; this would lead to large uncertainty on object depth information and incomplete lane observations. We process the highly distorted data in a multi-stage manner. We first utilize a neural network classifier to yield labeled lane-relevant objects. The lane marks/edges point clouds are processed by a truncated Gaussian random field model for the spatial filtering and a fading memory model for the temporal filtering. Then we present polynomial fitting scheme and a statistical analysis of the fitting errors reveals good lane and ego-vehicle orientation cues. In a parking lot real world study, we show promising lane detection and estimation performance of significant practical implications for lane keeping capability in high-automation applications.**

***Index Terms*—lane detection and estimation, surrounding vision, goodness-of-fit**

I. Introduction

Autonomous driving (AD) poses unique challenges for vehicle perception. The transition from human operator in advanced driver assistance systems (ADAS) to intelligent driverless automation addresses the necessity of acquiring deeper information from vehicle sensors. This requires us to not only explore new perception modalities of automotive sensors but also fully carve for more comprehensive capability in existing ADAS sensor systems [12][23].

The surround view camera systems (denoted concisely as SVS) is a widely used ADAS module [1][2]. It, through geometric alignment and composite synthesis, provides stitched panoramic view of vehicle surrounding environment and utilizes a pre-defined reference line for driver-assistant purpose. As shown in Fig. 1, the camera image if projected to ground coordinates would unavoidably cause significant distortion due to its protractor nature. The big depth uncertainty could lead to extreme difficulty in lane detection and estimation [9]. To overcome such a drawback, cooperative mono-cameras can be used to form a stereo vision system [7][13] or a LiCam (Lidar+Camera) incorporating 3-D point information into mono-camera image which creates super-pixel representation [8]. However, these solutions require extra hardware/sensor configurations.

In this study, we aim at achieving satisfying AD performance with minimum hardware change using existing product level SVS sensing systems. We carry out an image-based ground-coordinate lane detection and estimation in a multi-stage manner. We first process raw images based on a neural network classifier, yielding segmented pixel-wise image semantics. Then the semantic data labelled as lane marks/edges are selected and projected in 2-D ground coordinates, which is quantified by a grid representation [5]. We further apply a spatial filter and a temporal filter for outlying and smoothing purpose. Through an adaptive polynomial fitting of the filtered data and a proper statistical analysis on the fitting errors, we reveal that the processed semantic data contains rich information about lane shape and ego-vehicle orientation. In a parking lot scenario, we show promising lane detection and estimation performance of significant practical feasibility for intelligent lane keeping capability.

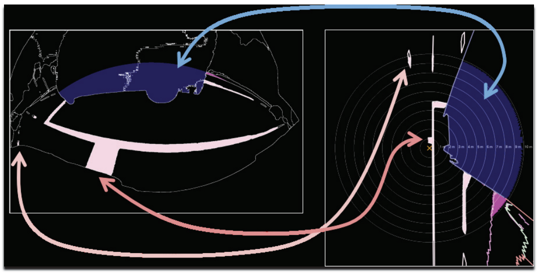


Fig. 1: Image space and ground plane correspondence.

Our major contributions in the work lie in following aspects: i) the data processing is carried out in ground plane (rather than in image space [9]) and can be naturally integrated into autonomous driving applications; ii) no map is required as prior information [18][19]; iii) we focus on statistical inference whilst meticulous on real-world statistical properties of data in each processing stage.

The paper is organized as follows. Section II introduces a work pipeline for the surround view camera system and its geometric properties in ground plane. Section III presents the processing details for the highly distorted semantic data. Section IV presents the data statistical analysis for lane estimation in a parking lot scenario. Summaries and conclusions are given in Section V.

II. Surround View Camera System

The automotive SVS assists the driver in parking by allowing top-down view of the 360 degree vehicle surroundings. A composite view of the vehicle surroundings is synthesized and reconstructed in real time as a visual aid tool [1][2].

We will extract more intelligent information for the sensing system. Fig. 2 shows a work pipeline from SVS raw images to ground plane

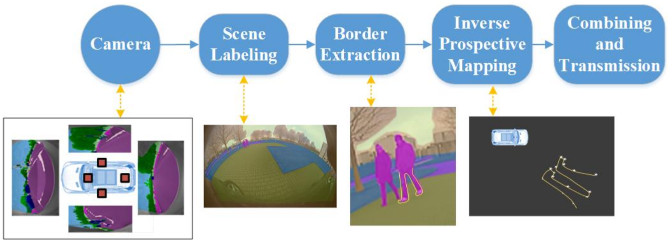


Fig. 2: SVS raw image labeling and prospective mapping.

semantic points. In the scene labeling stage, semantic segmentation is carried out based on a pre-trained GoogLeNet extended by a fully-connected layer trained on a proprietary data set of common objects on roads, e.g., vehicles, lane markings and curbs. This yields pixel wise label for the objects [4]. In the border extraction, the object edge pixels is selected and associated data is projected into ground plane (according to an empirical transform matrix). We can form, via Douglas-Peucker algorithm [15], polygon lines consisting of a list of labeled connected points.

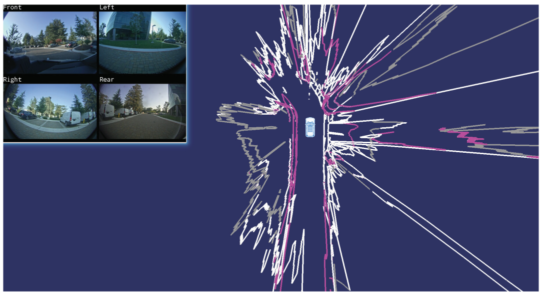


Fig. 3: SVS semantic contours (white: lane marking; magenta: sidewalk; gray: obstacle).

Fig. 3 shows the so-called SVS semantic contours, which are highly distorted in coordinate transform from image space to ground plane: the edges-associated data are extremely uncertain and any labeling errors can lead to disastrous results. This barely provides *direct* cues for good scene understanding.

III. Lane Recognition and Detection

We must properly quantify the SVS contour uncertainty to facilitate statistical analysis. To achieve so, we use a grid-based fuzzy logic scheme to simplify the data geometry representation. Then a spatial filter and a temporal filter are used as for outlying and smoothing purpose.

A. Grid Representation

We use only the SVS contour data labeled as lane marking. The corresponding SVS point clouds, as shown in Fig. 4(a), have two noticeable statistical properties: i) non-uniform uncertainty according to geometry and ii) the uncertainty is non-quantifiable with any known probability density function. We therefore carry out a scan-line polygon filling [24], as shown in Fig. 4(b), to erase the non-uniformity for possible uncertainty quantification.

The quantification is implemented based on classic grid mapping process using inverse measurement model [5][20]. The corresponding existing evidence are obtained via binary Bayesian filter (BBF) or Dempster-Shafer fuzzy logic [17][14], based on data accumulated over a small time window. Fig. 5 shows a result with gray-scale indicating existing probability.

(a) SVS point cloud (b) Filled polygon

Fig. 4: Scan-line polygon filling.

Fig. 5: Grid based filled polygons (over a small time window).

B. Spatial and Temporal Filtering

Now we can re-account for the non-uniform uncertainty. This is compensated for based on spatial geometry importance and data up-to-dateness. Fig. 6 illustrates the spatial filtering using a truncated Gaussian random field model and the temporal filtering using an exponential fading model.

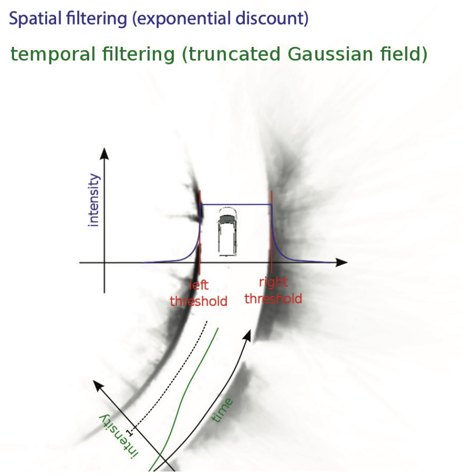
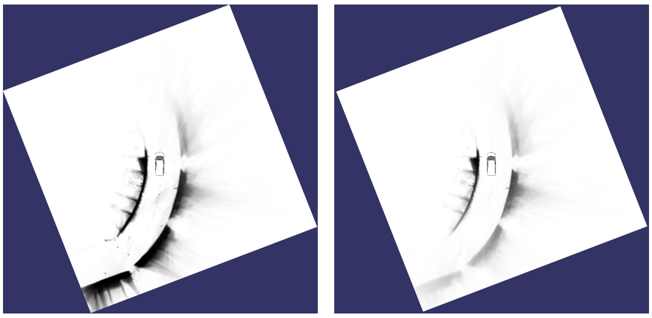


Fig. 6: The spatial-temporal filtering for SVS filled polygons.

The spatial filtering is carried out independently for different vehicle directions (left, right, front and rear). We define x^l as the distance from vehicle to the expected closest lane edge. By shifting x^l value, we can have different processed SVS filled polygons $\{C_i(k)\}$, $i = 1, 2, ..., n_k$. For each of the processed filled polygons, we carry out a boxing fitting [10]. The optimal x^l is the one with minimum area among the fitting boxes.

The temporal filtering is quite straightforward and an exponential coefficient as fading memory over data sequence is used. The temporal data processing largely alleviate errors from occasional neural network wrong labeling.



(a) Spatial filtering result (b) Temporal filtering result

Fig. 7: Results from the spatial and temporal filtering.

Fig. 7 shows the result after the spatial and temporal filtering. It can be seen that we could extract deeper information for decent lane estimation.

IV. Lane Estimation in a Parking Lot Study

After the grid representation and the spatial-temporal filtering, the lane estimation is achieved by an adaptive-order polynomial fitting with the filtered data.

Starting with fitting order $n = 1$, we test the following goodness-of-fit satisfying a specific (say, 95%) probability threshold [21]:

$$J_{N_c} \triangleq \mathbf{e}^{n\prime} \left(R_{N_c}\right)^{-1} \mathbf{e}^n \sim \chi^2_{N_c} \tag{1}$$

where $\mathbf{e}^n \triangleq \left[e_1^n, e_2^n, ..., e_{N_c}^n\right]'$, $R_{N_c} \triangleq \text{diag}\left\{\tilde{\sigma}^2, \tilde{\sigma}^2, ..., \tilde{\sigma}^2\right\}$, $\{e_i^n\}, i = 1, 2, ..., N_c$ is the fitting errors from the filtered data (of N_c points) and $\tilde{\sigma}$ is the fitting-error associated standard deviation (SD)[1].

Moreover, we can also have a *run-time* fitting-error SD as $\sigma_f = \sqrt{\frac{\sum(e_i^n - \bar{e}^n)^2}{N_c - 1}}$, where $\bar{e}^n$ is the mean of $\{e_i^n\}$.

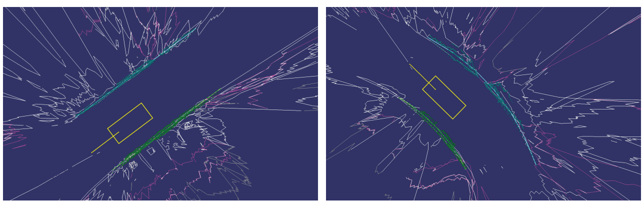


Fig. 8: Lane estimation via polynomial fitting.

we conclude that:

[1]Considering the fitting errors as one kind of measurement data, the associated SD can be *approximated* from a preliminary study.

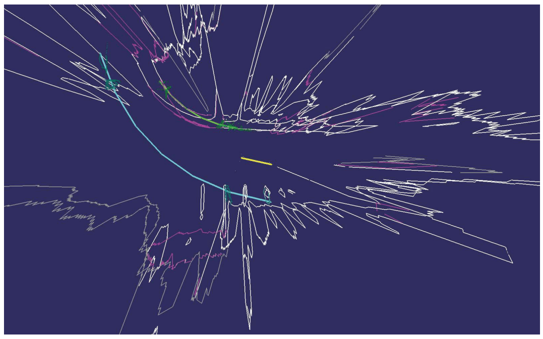


Fig. 9: Lane prediction in the case of SVS data missing or insufficient (e.g., turning in a T cross-road).

1). The fitting order indicates the curvature of road: The goodness-of-fit test is carried out via a chi-square distribution with approximated measurement SD $\tilde{R}$ [21]. Fig. 8 shows a first order fitting and a second order fitting results.

2). The run-time fitting-error SD could indicate lane width: the SD with $2\,\sigma_f$ (under Gaussian assumption) can be considered as a reasonable lane width estimation.

3). Difference of the run-time fitting error SDs from the left and right sides can indicate turning direction of the vehicle: when the vehicle makes turns, the image distortion on one side will be bigger than the other and therefore the run-time fitting error SDs. Similarly, heading direction if comparing the front and back.

4). Lane prediction can be achieved from optimal polynomial fitting: Fig. 9 shows a short-time prediction performance.

V. Summaries and Conclusions

In this study, we aim at using ADAS SVS sensing system to achieve AD performance with careful handling on data statistical properties. A multi-stage data processing approach is carried out for the highly-distorted SVS data. We first use a neural network for pixel-wise segmentation and labeling. Then we employ grid-based fuzzy logic to bring uniformity for the SVS contour non-stationary geometry uncertainties and simplify the corresponding quantification. We further design a spatial filter and a temporal filter to re-account for the non-uniformity, alleviating data labeling errors and depth uncertainty. After these processing, an adaptive polynomial fitting scheme is used on filtered data over a small time window. A statistical analysis on the properties of the fitting errors reveals that: i) we can achieve a good lane estimation on its location and shape; ii) we can roughly estimate the ego-vehicle orientation in real time; iii) a decent lane prediction even with incomplete data is also possible.

REFERENCES

[1] Altera, “White Paper: Generating Panoramic Views by Stitching Multiple Fisheye Images”, May 2009.

[2] V. Appia, H. Hariyani, S. Sivasankaran, et al., “Surround View Camera Systems for ADAS on TI’s TDAx SoCs”, Oct. 2015.

[3] K. Bengler, K. Dietmayer, et al., “Three Decades of Driver Assistance Systems: Review and Future Perspectives”, *IEEE Intelligent Transportation Systems Magazine*, vol.6(4):6-22, 2014.

[4] S. Bittel, et al., “Pixel-wise Segmentation of Street with Neural Networks”, *arXiv:1511.00513*, Nov. 2015

[5] K. Werber, M. Rapp, J. Klappstein, M. Hahn, et al., “Automotive Radar Gridmap Representations”, *IEEE MTT-S International Conference on Microwaves for Intelligent Mobility (ICMIM)*, 2015.

[6] A. Elfes, Using Occupancy Grids for Mobile Robot Perception and Navigation, *Computer*, vol. 22, no. 6, pp. 46–57, June 1989.

[7] U. Franke, C. Rabe, H. Badino and S. Gehrig, “6D-Vision – Fusion of Motion and Stereo for Robust Environment Perception, *DAGM Symposium 2005*, Vienna, pp.216223, 2005.

[8] U. Franke, “Towards a Comprehensive Super-Pixel Representation of Traffic Scene Images”, invited talk, *IEEE-ICCV 2017 Workshop*, Venice, Italy, Oct. 2017.

[9] A. B., Hillel, R. Lerner, D. Levi and G. Raz, “Recent Progress in Road and Lane Detection: a survey”, *Machine vision and applications*, vol(3)25: 727-745, 2014.

[10] J. Peukert, “Object Generation from a Dynamic Occupancy Gridmap Considering Autocorrelation”, Master Thesis, *Institute of Systems Optimization, Karlsruhe Institute of Technology*, Karlsruhe, Germany, Feb. 2016.

[11] J. Javier Yebes, P. F. Alcantarilla, L. M. Bergasa, et al., “Surrounding View for Enhancing Safety on Vehicles”, *Proceedings of the 2012 IEEE Intelligent Vehicles Symposium Workshops*, 2012.

[12] J. Levinson, J. Askeland, J. Becker, et al., “Towards Fully Autonomous Driving: systems and algorithms”, *2011 IEEE Intelligent Vehicles Symposium (IV)*, pp:163-168, 2011.

[13] J. C. McCall and M. M. Trivedi, “Video-based Lane Estimation and Tracking for Driver Assistance: Survey, System, and Evaluation”, *IEEE transactions on intelligent transportation systems* vol.7(1): 20-37, 2006.

[14] D. Nuss, T. Yuan, G. Krehl, et al., “Fusion of Laser and Radar Sensor Data with a Sequential Monte Carlo Bayesian Occupancy Filter”, *2015 IEEE IV Symposium*, Korea, June 2015.

[15] D. K. Prasad, K. Dilip, et al., “A Novel Framework for Making Dominant Point Detection Methods Mon-parametric”, *Image and Vision Computing*, 30 (11): 843859, 2012.

[16] S. H. Qian, J. K. Tan, et al., “Road Region Estimation and Obstacles Extraction using a Monocular Camera”, 2013.

[17] S. Thrun, et al., Probabilistic Robotics, *MIT Press Campridge*, 2005.

[18] A. Wedel, U. Franke, et al., “Realtime Depth Estimation and Obstacle Detection from Monocular Video.” *Lecture notes in computer science*, 2006.

[19] K. Yamaguchi, A. Watanabe, T. Naito, and Y. Ninomiya,“Road Region Estimation using a Sequence of Monocular Images”, *19th International Conference on Pattern Recognition (ICPR 2008)*, pp:1-4, 2008.

[20] T. Yuan, D. Nuss, et al., “Fundamental Properties of Dynamic Occupancy Grid Systems for Vehicle Environment Perception”, *2015 IEEE International Workshop on CAMSAP*, Cancun, Mexico, Dec. 2015

[21] T. Yuan, Y. Bar-Shalom, P. K. Willett and D. Hardiman, “Impact Point Prediction for Thrusting Projectiles in the Presence of Wind”, *IEEE Trans. Aerosp. Electronic Systems*, 2013.

[22] B. Zhang, V. Appia, I. Pekkucuksen, et al.,“A Surround View Camera Solution for Embedded Systems”, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp.662-667, 2014.

[23] A. Ziebinski, R. Cupek, H. Erdogan and S. Waechter, “A Survey of ADAS Technologies for the Future Perspective of Sensor Fusion”, *International Conference on Computational Collective Intelligence*, Springer International Publishing, pp.135-146, 2016.

[24] https://www.cs.uic.edu/~jbell/CourseNotes/ComputerGraphics/PolygonFilling.html