Ground Truth and Performance Evaluation of Lane Border Detection

4 authors, including:

Some of the authors of this publication are also working on these related projects:

Multimedia Imaging View project

Automated glaucoma diagnosis using deep and transfer learning: Proposal of a system for clinical testing View project

Ground Truth and Performance Evaluation
of Lane Border Detection

Ali Al-Sarraf[1], Bok-Suk Shin[1], Zezhong Xu[2], and Reinhard Klette[1]

[1] Department of Computer Science, The University of Auckland Auckland,
New Zealand{aals005,b.shin,r.klette}@auckland.ac.nz
[2] College of Computer Information Engineering, Changzhou Institute of Technology
Changzhou, Jiangsu, China xuzz.1999@263.net

Abstract. Lane-border detection is one of the best-developed modules
in vision-based driver assistance systems today. However, there is still a
need for further improvement for challenging road and traffic situations,
and a need to design tools for quantitative performance evaluation.
This paper discusses and refines a previously published method to gener-
ate ground truth for lane markings from recorded video, applies two lane-
detection methods to such video data, and then illustrates the proposed
performance evaluation by comparing calculated ground truth with de-
tected lane positions. This paper also proposes appropriate performance
measures that are required to evaluate the proposed method.

1 Introduction

Vision-based driver assistance systems are already standard modules in modern
cars, supported by the availability of high-computing power and low-voltage
purpose-designed FPGA solutions, small and accurate cameras, that can fit in
any vehicle, and progress in the methodology of computer-vision solutions. Lane
border detection, a component of vision-based driver assistance solutions, has
been studied for more than twenty years, and there are robust solutions available
for road environments where lane markings are clearly visible, such as highways
or multi-lane main roads.

Vision-based lane detection supports, for example, lane departure warning,
lane keeping, lane centring, and so forth [1]. Despite the many available algo-
rithms and approaches, an ongoing concern [1, 9] is the lack of proper ground-
truth estimation to evaluate the efficiencies and accuracies.

A common way, how publications validate the accuracy of their approach, is
by using their naked eye for the validation. Building a ground-truth data base
can become really difficult as roads are not generic even within a single country,
let alone internationally. Roads can be well built or not, have proper markings or
no marks at all, they can be urban or countryside roads, have solid lane markers
or painted dashed lane borders,or differ in many other ways. The environment is
only one factor, another factor is the equipment, such as the type of camera used
(e.g. image resolution, grey-level or color, bits per pixel, or geometric accuracy).

The generation of ground truth for lane detection needs to reflect on many
parameters, to ensure the creation of a trust-worthy ground truth. Digitally
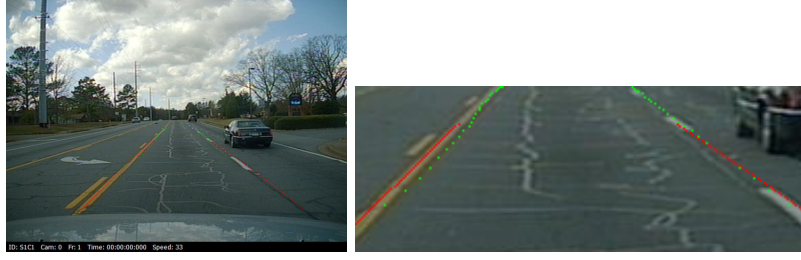
Fig. 1. *Left*: Ground truth and calculated lane markings. *Right*: Magnified window of the image shown on the left. The red lines are calculated ground truth, and the green dots are calculated lane-border positions. Note that algorithms for lane-border detection not necessarily provide curves; it might be just isolated dots.

simulated ground truth was created by Revilloud et al. [10] but they found out that, when adapting their lane detection algorithm to their synthetic ground truth, it did not work very well on real-world data as their approach mainly focuses on the ground texture for detecting lines. They stated a need for another solution for ground-truth generation.

There is another obvious solution: to generate ground-truth manually, supported by some graphics routines for drawing lines. However, considering frame rates of at least 25 Hz, and the need to generate ground truth for very long video sequences, this would be a very tedious task.

Borkar et al. [2] developed a technique using time slices and splines to generate ground truth from any type of road image sequence recorded in an *ego-vehicle* (i.e. the vehicle the vision system is operating in). The approach works reasonably well on clearly marked roads, but the involved interaction also comes with the risk of human error and limited usability. The method was easy to re-implement. It works well on long sequences, as long as there are some markings in the frames which identify the lanes. However when selecting points on lane markings and the points are not at the center of the drawn lines, then this may lead to errors.

In this paper, we provide an improvement that makes the ground-truth generation process easier to use, and which also helps to generate ground truth for a diversity of recorded video data. Our solutions use standard image-processing techniques for making the entire process easier, and to reduce the errors in point selections, thus going closer towards a fully automated solution. We also provide novel measures for comparing ground truth with calculated lane borders.

Figure 1 illustrates the subject of this paper. It shows ground truth together with results applying one lane-detection algorithm which generates isolated points rather than curves as output. A quantitative comparison between ground truth and estimated lane borders requires evaluation measures.

The paper is structured as follows. Section 2 provides a brief explanation of the drawing technique by Borkar et al. and of our improvements. Section 3 discusses how to quantify performance by introducing measures. Section 4 reports about some experiments for illustrating the approach. Section 5 concludes.

2   Ground Truth by Time Slices

Evaluations of computer vision techniques based on available ground truth became a widely accepted approach for improving methods, for identifying issues, and to help to overcome those issues. Current examples are the KITTI benchmark suite [4, 7] and some of the sequences on EISATS [3, 8]. These are websites offering long video sequences, more than 100 frames each, for testing vision algorithms on recorded road scenes. Such websites provide currently only manually generated ground truth for lane marking in a few frames (e.g. KITTI for about 200 frames). For really challenging video data, so far we can only apply subjective evaluations of algorithmic performances, such as demonstrated by the Robust Vision Challenge at ECCV 2012 [11]. This current situation illustrates the difficulty of providing usable ground truth for extensive lane-detection experiments.

We propose a solution for generating ground-truth data for lane sequences by extending the technique proposed in [2] with the aims of reducing errors in the point-selection step, and of increasing the efficiency of the use of this method.

Ground-truth data for lane sequences is defined by generated curves indicating where the actual lane border is located in an image. The process starts with creating an image called a *time slice* by selecting points from each frame of the given sequence. Next, spline interpolation is applied for those selected points to generate ground truth. See Figure 2.

For comparing with [2], we use the same two sequences as used in this paper (and made publicly available). The first sequence consists of 1372 frames, each having a resolution of 640×500 pixels, and the second of 400 frames, each having a resolution of 640×480 pixels. For generating $n > 0$ time slices, we extract n rows of pixels from each frame at fixed row locations. A distance between subsequent rows of around 20 to 30 pixels appears to be appropriate for a standard 640×480 VGA image format, and for n, a value between 3 and 5 is reasonable. Each of the n fixed rows, accumulated over time, defines one time slice: the row from Frame 1 goes into the bottom-most row, the row from Frame 2 into the next, and so forth.
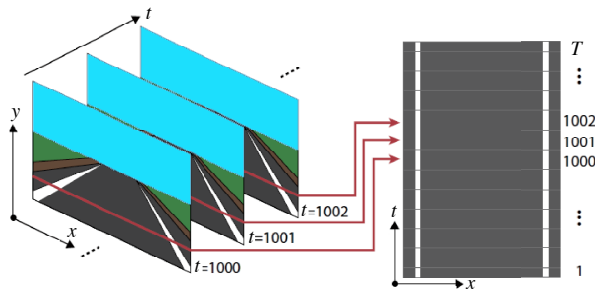

Fig. 2. A time slice is created by using a stack approach which combines single rows of pixels from each frame into one image (i.e. a sequence of rows). Figure follows [2] with modified notation.
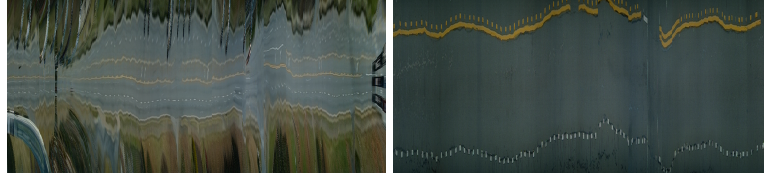
Fig. 3. Samples of two generated time slices for the first sequence. The images have 1372 rows because 1372 frames contributed to each time slice. *Left*: Time slice generated from row 270 (counted from the top, i.e. further away from the ego-vehicle). *Right*: Time slice generated from row 400 (i.e. closer to the ego-vehicle)

Figure 3 shows examples of two time slices calculated from the same image sequence but from two different fixed rows. After creating $n > 0$ different time slices, a number of points is manually selected on the left and right lane borders, and we apply cubic spline interpolation for curve fitting. This generates curves shown in Fig. 4 as white curves in one time slice.

After repeating the same curve generation on each of the $n > 0$ time slices, all the created points are propagated from those time slices into corresponding locations (on the fixed rows) in the original frames, thus having n points on each lane border in each frame. The authors of [2] re-apply then curve fitting (by interpolation) once again. This generates the proposed ground truth data for each data sequence. Figure 5 shows ground truth generated this way using our re-implementation of the original algorithm by [2]. The approach is easy and reasonably time-efficient in generating ground truth data on different types of video sequences.

However, the crucial problem we experienced is the uncertainty in selecting points in time slices, especially in time slices created by rows far away from the ego-vehicle. These points should ideally be at the center of a lane border. Figures 4 and 5 show visible deviations from an ideal center line.

The index of the fixed row for generating one time slice has impacts on accuracy: the further away from the ego-vehicle, the more likely that manually located points are not supporting ideal ground truth curves being at the center of lane markings.
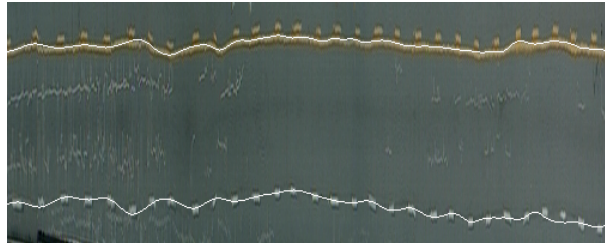

Fig. 4. The white curves are generated by applying curve fitting (interpolation) to manually selected points in the time slice shown on the right in Fig. 3. The interpolated curves follow the lane markings.

Fig. 5. Generated white lines after propagating points from time slices into the image sequence and interpolating those points, following the original algorithm.

We added a simple but useful modification, based on tests of various edge detectors on generated time slices. Finally we decided to use the Canny edge detector for detecting left and right edge points for lane markings in time slices and to constrain selected points to midpoints between such left and right edge points. Tracking of pairs of edge points and an automated selection of midpoints leads to improved ground truth generation in recorded frames. Besides an automated point selection, this extension also provides estimates for the width of lane markings (which is used for normalising our evaluation measure; see next section).

Figure 6 shows results of the implementation of the original method (red curves) and of the proposed method (blue curves). The blue curve, i.e. the generated ground truth, is at the center of both black border curves, defined for the left and right lane marking; see Fig. 6, bottom.

Fig. 6. Comparison between original and proposed method. *Top*: Generated ground-truth (red) using the original interactive approach. *Bottom*: Generated ground truth (blue) using our approach, also indicating lane-marking width

3 Performance Measures

In this section we describe two novel measures for comparing ground truth with calculated lane borders. In each frame, we only consider an interval of relevant

rows, with indices between $y_{\min}$ and $y_{\max}$. First, we describe a measure for comparing ground truth with detected lane borders also covering algorithms which provide isolated points for lane-border positions. For each row y, with $y_{\min} \leq y \leq y_{\max}$, we have the following cases for ground truth:

1. Ground truth provides points on a left and a right lane border (case GTB).
2. Ground truth only provides a point on the left border (case GTL).
3. Ground truth only provides a point on the right border (case GTR).
4. Ground truth provides no point in this row y (case GTN).

Analogously, we also have the cases BDB, BDL, BDR, and BDN for *border detection* (BD) by the applied detection algorithm. For example, we may not even see the right border of the lane in which the ego-vehicle is driving in at that moment. We consider cases such as GTL, GTR, and GTN as being defined by the circumstances.

In row y, with $y_{\min} \leq y \leq y_{\max}$, we use the notation $x_{y,L}^{GT}$ and $x_{y,R}^{GT}$ for detected ground truth points (if they exist at all), and $x_{y,L}^{BD}$ and $x_{y,R}^{BD}$ accordingly for a studied border-detection algorithm. The error $E_{IP}(y,t)$ is defined for row y, with $y_{\min} \leq y \leq y_{\max}$, and a selected Frame t of the input sequence, using the L_1-norm:

$$E_{IP}(y,t) = \begin{cases} ||(x_{y,L}^{GT}, x_{y,R}^{GT}) - (x_{y,L}^{BD}, x_{y,R}^{BD})||_1, & \text{if cases BDB and GTB} \\\\ ||x_{y,L}^{GT} - x_{y,L}^{BD}||_1, & \text{if cases BDL, and GTL or GTB} \\\\ ||x_{y,R}^{GT} - x_{y,R}^{BD}||_1, & \text{if cases BDR, and GTR or GTB} \\\\ 0 & \text{otherwise} \end{cases} \tag{1}$$

Let τ_y be the expected width of the lane marking in row y, provided by our ground-truth generation method. If the value $E_{IP}(y,t)$ is less than the value of τ_y, then we consider this as insignificant deviation from ground truth and replace value $E_{IP}(y,t)$ for this row by zero.

The reason for this normalisation is that lane-detection algorithms typically detect one side of the lane marking, or the center of the lane marking, depending on the methodology used. The use of a value τ was introduced in performance measures in [14] as an a-priori tolerance threshold value, identified with a constant representing the ideal lane-marking width as stated by the Federal Highway Administration (FHA) [15]. This ideal value can be accurate, but not all data sequences have that information available, and it also needs to be considered as a function of row index y. The difference is that we extract τ_y while creating the ground truth data.

The given error measure provides multiple options for measuring inconsistencies between the BD algorithm and the ground truth. We only formally state one option here, defined by the use of only those rows $y_{0,t} < y_{1,t} < \ldots < y_{m_t,t}$, with $y_{\min} \leq y_{i,t} \leq y_{\max}$, for $i = 0, \ldots, m_t$, where we have both GTB and BDB.

This gives us the error

$$E_{BD} = \frac{1}{T} \sum_{t=1}^{T} \left[\frac{1}{m_t + 1} \sum_{i=0}^{m_t} E_{IP}(y_{i,t}, t) \right] \qquad (2)$$

for a considered method BD for the whole sequence of T frames.

4   Experiments and Discussion

We illustrate a comparison between the proposed ground truth with results of two lane-border detectors using the proposed measure E_{BD}.

The first border detector is an adaptation of the technique described in [6]. This method uses a lane border model, originally proposed in [13], defined by isolated left and right lane border points in one image row, detected by applying a particle filter for each row. The second is a technique proposed in [12] which applies a less restrictive lane border model and a combination of particles defined by multiple image rows, thus also supporting temporary tracking of lane borders while lane markings are actually missing.

We illustrate results for two different data sets, the first sequence of data used by Borkar et al. [2] (see the specification of both sequences given above), and also one sequence from EISATS [3]. See Table 1.

Interestingly, we obtained individual measurements for those sequences (which come with different characteristics, defined by frame resolution, used cameras, and bits per pixel, but also by road conditions, traffic density, lighting, contrast, or weather conditions).

In the table, GT and BD size specify the total numbers of points generated from each sequence (kept constant for both sequences). We provide percentages of correct and misaligned border detections. As a result, the values of measure E_{BD} point out that the method of [12] is outperforming the method of [6] on those two sequences (more clearly on the more challenging EISATS sequence).

Table 1. Performance results. B1 for Seq #1 of [2], and E1 for EISATS Seq #1

Data	#Frames	Algorithm	GT size	BD size	Coverage	Correct	Misaligned	E_{BD}
B1	1250	[6]	131	30	36.64	63.75	36.25	7.98
B1	1250	[12]	131	30	36.64	79.83	20.16	6.11
E1	325	[6]	131	30	21.37	59.49	40.50	11.44
E1	325	[12]	131	30	21.37	70.23	29.76	6.02

5   Conclusions and Future work

We proposed an improvement to a currently published ground-truth generation method proposed in [2] for lane borders in recorded on-road video data. The addition works well on different types of video sequences, and also for identifying the width of lane border markings.

We defined an appropriate performance measure and applied it to selected lane-border detection algorithms. The evaluation is consistent with statements, for example, in paper [8], that diversities of *situations* requires adaptive selections of techniques for optimizing analysis results. The proposed framework can be used to identify better algorithms which correspond in their performance to a given situation or scenario.

For future work, we would like to test our ground truth method for more BD algorithms, and will also increase the range of considered situations.

References

1. Bar Hillel, A., Lerner, R., Levi, D., Raz, G.: Recent progress in road and lane detection: A survey. *Machine Vision Applications*, pages 1–19 (2012)
2. Borkar, A., Hayes, M., Smith, M.T.: An efficient method to generate ground truth for evaluating lane detection systems. In Proc. IEEE Int. Conf. Acoustics Speech Signal Processing, pages 1090–1093 (2010)
3. EISATS benchmark data base. The University of Auckland, `www.mi.auckland.ac.nz/EISATS` (2013)
4. Geiger, A., Lenz, P., Urtasun, R.: Are we ready for autonomous driving? The KITTI Vision Benchmark Suite. In Proc. IEEE Int. Conf. Computer Vision Pattern Recognition, pages 3354–3361 (2012)
5. Jiang, R., Klette, R., Wang, S., Vaudrey, T.: New lane model and distance transform for lane detection and tracking, In Proc. CAIP, LNCS 5702, pages 1044 – 1052 (2009)
6. Jiang, R., Klette, R., Vaudrey, T, Wang, S.: Lane detection and tracking using a new lane model and distance transform, *Machine Vision Applications*, **22**:721–737 (2011)
7. KITTI vision benchmark suite. Karlsruhe Institute of Technology, `www.cvlibs.net/datasets/kitti/` (2013)
8. Klette, R., Krüger, N., Vaudrey, T., Pauwels, K., Hulle, M., Morales, S., Kandil, F., Haeusler, R., Pugeault, N., Rabe, C., Lappe, M.: Performance of correspondence algorithms in vision-based driver assistance using an online image sequence database. *IEEE Trans. Vehicular Technology*, **60**:2012–2026 (2011)
9. McCall, J.C., Trivedi, M.M.: Video-based lane estimation and tracking for driver assistance: Survey, system, and evaluation. *IEEE Trans. Intelligent Transportation Systems*, **7**:20–37 (2006)
10. Revilloud, M., Gruyer. D., Pollard, E.: Generator of road marking textures and associated ground truth applied to the evaluation of road marking detection. In Proc. IEEE Int. Conf. Intelligent Transportation Systems, pages 933 – 938 (2012)
11. Robust Vision Challenge at ECCV 2012. See `hci.iwr.uni-heidelberg.de/Static/challenge2012/` (2012)
12. Shin, B.-S., Tao, J., Klette, R.: A superparticle filter for lane detection. Submitted (2014)
13. Zhou, Y., Xu, R., Hu, X., Ye, Q.: A robust lane detection and tracking method based on computer vision. *Measurement Science Technology*, **17**:736–745 (2006)
14. Borkar, A.: Multi-viewpoint lane detection with applications in driver safety systems. *PhD thesis*, (2011)
15. Federal Highway Administration: Manual Uniform Traffic Control Devices. See `http://mutcd.fhwa.dot.gov/` (Nov. 2009)