AN EFFICIENT METHOD TO GENERATE GROUND TRUTH FOR EVALUATING LANE DETECTION SYSTEMS

Amol Borkar, Monson Hayes

Center for Signal and Image Processing (CSIP)
Georgia Institute of Technology
Atlanta, GA
{amol,mhh3}@gatech.edu

Mark T. Smith

Institut för Tillämpad Informationsteknik
Kungliga Tekniska Högskolan
Stockholm, Sweden
msmith@kth.se

ABSTRACT

In this document, a new and efficient method to specify the ground truth locations of lane markers is presented. The method comprises of a novel process called *Time-Slicing* which provides the user with a unique visualization of the video. Coupled with automation via spline interpolation, the quick generation of necessary ground truth information is achieved. Videos recorded from a vehicle while driving on local city roads and highways are marked with ground truth information for use in testing. The performances of a variety of lane detection systems are compared to the ground truth and the error is computed for each system. Finally, quantitative analysis shows that the reference lane detection system presented in [1] produces the most accurate lane detections which is depicted by the smallest error.

Index Terms**—** Lane Detection, Lane Tracking, Matched Filtering, Inverse Perspective Mapping, Ground Truth.

1. INTRODUCTION

Improving public safety on roads has always been a topic of interest to automotive research. As vehicles can commute at high speeds on freeways, the consequences of an unattended moment can be catastrophic. High fatality rates from traffic accidents are evident world wide. In 2002, the Secretary for the Transport Ministry of Malaysia cited 4.9% of traffic related accidents as fatal [2]. The Ministry of Public Safety of China reported 667,507 traffic related accidents in 2003 as fatal [3]. In the United States, the Fatal Analysis Reporting (FAR) system of National Highway Traffic Safety Adminstration (NHTSA) reported 41,059 casualties and over 2.5 million injuries in 2007 [4]. The high death tolls have suggested embedding smart systems in a vehicle to aid a driver during a commute giving rise to Driver Assistance systems.

Driver Assistance (DA) systems are systems that provide aid to the driver once inside the vehicle. DA systems are gaining significant interest as powerful electronics capable of handling complex tasks are capable of being integrated into the automobile. Lane departure warning, lane change assistance, blind spot monitoring and adaptive cruise control are a few examples of popular DA systems.

This paper introduces a novel technique to generate ground truth information which is used to evaluate the quality of a few lane detection implementations. Following the introduction, common lane detection methods and ground truth generation techniques are evaluated and their weaknesses are highlighted. Then, the new ground truthing procedure is explained in detail and the reference lane detection system is briefly described. The performances of a variety of systems are assessed on real world data marked with ground truth. Finally, the conclusion and planned improvements are discussed.

2. PRIOR RESEARCH

Generating ground truth information for testing the quality of any system is generally difficult. In lane detection applications, the most common approach is to mark the lane locations in every frame while iterating through the video set. This is the simplest procedure to produce the ground truth; however, it is an extremely slow process and can takes several days to produce the data for one video set. As generating this type of data is very time consuming and extremely tedious, it is commonly avoided.

Besides generating ground truth, techniques to improve current day lane detection systems are also constantly under investigation. Some of the most popular techniques involve using the classical Hough transform [5] or *Middle-to-Side* approach to find lane boundaries in an edge map [1, 6, 7, 8]. Both techniques generally perform well when the road surface is devoid of artifacts. Detecting lane markers via color segmentation and color space transformations has also been proven quite effective [1, 8, 9]. Unfortunately, color based approaches rely strongly on the presence of white-colored illumination to reduce difficulty in thresholding.

Lane detection is a crucial component of many DA systems because it serves as the foundation for applications such as Lane Departure Warning (LDW), Lane Change Assistance (LCA), and, autonomous driving. From the examples above, it is clear that the feature extraction stages in existing research are constrained to operation under certain illumination and road conditions. In addition, without ground truth information, the lane detection results are qualitative and based purely on visual inspection by either one or multiple users. With the visual approach, problems of subjectivity arise where one user might grade the performance of a particular system better than another user. This can lead to inconsistency in results when a system is tested with an assortment of data. Hence, it is necessary to develop a method to efficiently produce ground truth information to allow a systematic comparison between various implementations as well as maintain consistency in the results.

3. GROUND TRUTH

In a camera captured image, marking the individual lane marker locations is a common approach to generate ground truth for a particular frame. However, due to the absence of any sort of assistance or automation, it is a very slow process. Additionally, when dealing hours of video footage, the ground truth generation task could take several days to complete. Therefore, a technique to efficiently generate ground truth is essential. One such method is with the creation of *Time-Sliced* images.

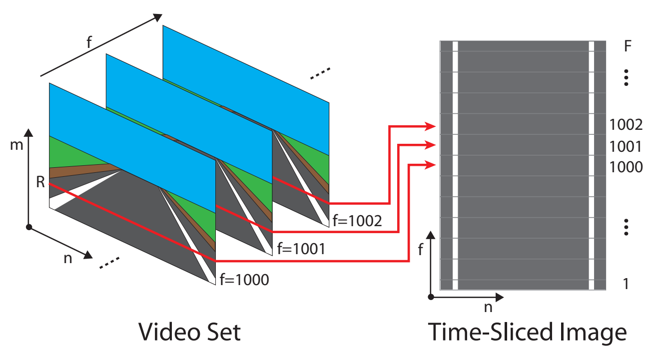
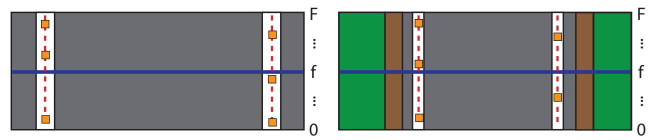


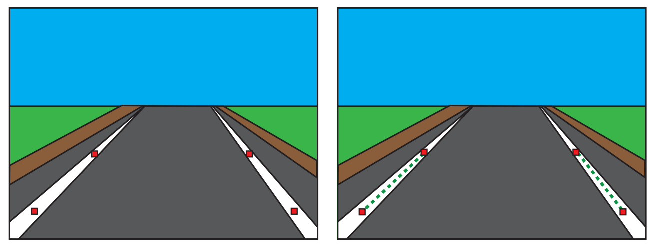
Fig. 1: Creating a *Time-Sliced* Image. Rows of pixels from an video set are copied using a stack approach.

A *Time-Sliced* (TS) image is created by stacking an empty image with a row of pixels from frames in the video set. To elaborate, in a video set that contains F frames, each frame $f \in$ [1, F] can be described as an image having dimensions $M \times N$. In contrast, a *TS* image has dimensions $F \times N$ for each row R, where $R \in$ [0, M). Due to the stacking of pixels, each value of R produces a unique *TS* image. The *TS* image has a width N which is the same as in each frame f since an entire row of pixels is stacked. It's height is usually F since F rows are stacked. An inherent property of the *TS* image is that traversing along the rows is equivalent to traversing through the different frames or time stamps in the video set due to the stacking of rows. This idea is illustrated in Fig. 1.



(a) *Time-Sliced* image for R=350. Frame f in blue. User marked points in orange and interpolated locations represented by dashed red line.

(b) *Time-Sliced* image for R=220. Frame f in blue. User marked points in orange and interpolated locations represented by dashed red line.



(c) Left and right lane markers locations for R=220 and 350 in frame f.

(d) Interpolated values between R=220 and 350 shown by dashed green line.

Fig. 2: Generating ground truth for each frame from *Time-Sliced* images.

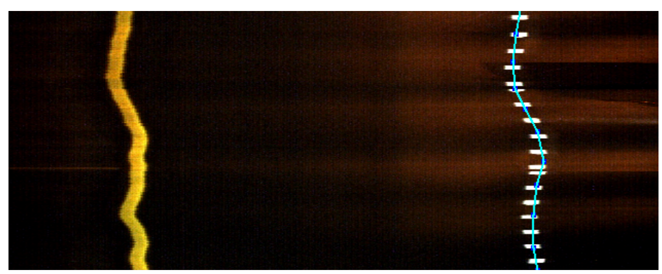


Fig. 3: A small section of a *Time-Sliced* image.

A custom designed tool in Visual C# [10] allows a user to mark various points in the *TS* image. These points tend to correspond to lane marker centers. Curve fitting with interpolation is then used to string the points together. Cubic spline is the chosen interpolation scheme as it produces a smooth curve. The interpolated values recovered from spline interpolation are used to determine the lane marker locations for the intermediate frames. In this particular implementation, 3 different values of R are chosen to produce 3 unique *TS* images for a video set which leads to recovering lane locations on the 3 rows. Another application of spline interpolation along the 3 rows allows to determine the lane locations for the intermediate rows in each frame f. With the lane marker locations recovered for the desired rows in image, ground truth is successfully generated for that particular frame. Fig. 2 illustrates the ground truth generation process for two different values of R. The automation provided by interpolation greatly helps in improving efficiency and at the same time, reduces the requirement of detailed input from the user. A small section of a *TS* image is shown in Fig. 3 with the automated ground truth locations highlighted in light blue.

4. LANE DETECTION SYSTEM [1]

The camera captured image undergoes preprocessing in the form of temporal blurring and grayscale conversion. Then, Inverse Perspective Mapping (IPM) is applied to remove perspective and transform the image into a bird's-eye view. An adaptive threshold converts the grayscale image into binary and then a low-resolution Hough transform is computed to find a set of candidate lane markers. The candidates markers are further scrutinized in an matched filtering stage to extract the lane marker centers. Random Sample Consensus (RANSAC) is used to estimate parameters for fitting a mathematical model through the recovered lane markers. Finally, the Kalman filter predicts the parameters of each lane marker line from one frame to the next.

5. EXPERIMENTAL ANALYSIS

5.1. Error Calculation

The Federal Highway Administration states the official width of the lane markers as 6 inches [11]. In the IPM image, this width is translated to a distance in pixel units and used as a buffer around the ground truth locations. Consequently, lane marker estimates that fall within this buffer are categorized as having no error. As a result, the error in each frame, $\mathbf{Er}(f)$ is computed as the average distance between the ground truth locations and the estimated lane markers for all defined rows in the IPM image.

$$\lambda_{(i,f)} = \mathbf{max}(|Gt_{(i,f)} - X_{(i,f)}| - \frac{W}{2}, 0) \qquad (1)$$

$$\mathbf{Er}(f) = \sum_{i=1}^{N} \frac{\lambda_{(i,f)}}{N} \text{ feet} \qquad (2)$$

In Eq. (1), $Gt_{(i,f)}$ is the ground truth location of the lane marker and $X_{(i,f)}$ is the estimated lane location on row i of frame f. W is the width of the buffer around the ground truth location and λ is the measured distance. Consequently, the error in each frame is calculated by averaging all applicable λ values in frame f as shown in Eq. (2). The λ values are measured in the IPM image instead of in the camera captured images as inter-pixel distances remain uniform in the perspective-free image [1].

5.2. Automated Ground Truth Evaluation

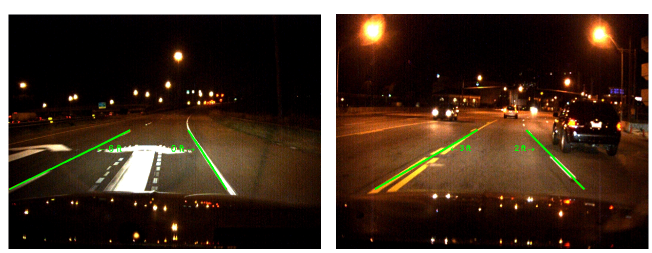
Since the generation of the ground truth data is largely automated with the help of *Time-Slicing*, it is essential to compare it's accuracy to reference ground truth data that is created manually in it's entirety. A collection of video clips consisting of driving on straight roads, curving roads and, lane changes were used in the evaluation. Tests were performed by marking the ground truth locations of lane markers in every frame of the video clips and computing $\mathbf{Er}(f)$. In these tests, X in Eq. (1) is the lane marker estimate calculated using the automated ground truth process and W is set to 0. An $\mathbf{Er}(f) < 0.5$ implies that the estimates are within a few inches of the ground truth portraying good accuracy. Table 1 shows the errors produced by the automated process depicting it's high accuracy. In addition, the *Time-Slicing* approach also saves enormous amounts of time in generating ground truth.

Table 1: Evaluation of the automated ground truth.

Type of Test	Avg. $\mathbf{Er}(f)$ ft.
Straight Road	0.167
Curving Road	0.182
Lane Change	0.189

5.3. Results

A color camera installed below the rear-view mirror is used to capture video. A variety scenarios were undertaken to generate a realistic set of driving conditions that one would encounter on local roadways. Over a total of five hours of video was captured in the process and each video was marked with the ground truth locations of lane markers using the automated process described earlier. The results are quantified in terms of accuracy per minute and categorized as follows: i) a correct detection occurs when more than half of the lane marker estimates are inside the ground truth buffer, ii) an incorrect detection occurs when more than half of the lane marker estimates are outside the ground truth buffer, and iii) a missed detection occurs when no lane is detected when one is visible. The error is calculated only when a correct detection occurs. Table 2 shows a comparison of performance between common lane detection implementations.



(a) (b)

Fig. 4: Reference system [1] performing lane detection accurately.

As expected, the reference system [1] is able to outperform the other implementations shown by the quantitative results in Table 2. Despite the presence of many surface artifacts, the reference system is still able to locate lane markers

Table 2: Performance comparison between different lane detection implementations.

Method	Road Type	Traffic	Avg. Rate Per Minute			
			Correct	$\mathbf{Er}(f)$ ft.	Incorrect	Misses
Reference [1]	Isolated Highway	Light	98.2%	0.032	1.49%	0.31%
		Moderate	97.94%	0.061	2.05%	0%
	Metro Highway	Light	98.27%	0.14	0.89%	0.41%
		Moderate	97.50%	0.21	1.95%	0.55%
	City	Variable	87.21%	1.49	8.18%	6.61%
Middle-to-Side	Isolated Highway	Light	90.5%	3.06	9.5%	0%
		Moderate	78.85%	6.85	21.41%	0%
	Metro Highway	Light	82.15%	3.05	17.85%	0%
		Moderate	76.81%	10.33	23.02%	0%
	City	Variable	69.92%	10.94	31.09%	0%
Hough	Isolated Highway	Light	80.37%	2.48	19.63%	0%
		Moderate	79.70%	5.4	20.29%	0%
	Metro Highway	Light	74.65%	14.2	25.35%	0%
		Moderate	73.22%	10.23	26.77%	0%
	City	Variable	65.16%	14.87	34.83%	0%

on the road as shown in Fig. 4. Incorrect and missed detections occurred occasionally and were mainly contributed to by wear and tear of the road surface. In addition, the error values of $\mathbf{Er}(f) < 0.5$ implies that the lane marker estimates are in fact very close to the ground truth locations. In contrast, the Hough Transform and *Middle-to-Side* approaches for lane detection did not fare as well. The relatively larger error values of $\mathbf{Er}(f) > 0.5$ suggest that the lane marker estimates frequently fluctuate more than a few inches away from the ground truth locations. In addition, the lack of a threshold to classify pixels as lane markers or not causes the two approaches to constantly yield 0% miss-rate.

6. CONCLUSION

A new and efficient process to generate ground truth is presented in this paper. With the introduction of *Time-Slicing*, the process benefits from automation via spline interpolation that greatly reduces the time spent to produce ground truth. A reference lane detection system [1] is briefly described and a detailed comparison to other common lane detection implementations is also presented. The results are deduced quantitatively using set categories and error calculations rather than visual inspection. The data sets used for testing were created by recording videos while driving on Interstate highways and city streets in and around Atlanta, GA. Even in the presence of a many artifacts on the road surface, the reference system was able to yield overall performance that was superior to other implementations as shown in Table 2.

7. FUTURE WORK

Image processing techniques will be used to accurately determine the exact widths of the lane markers on each row in every frame of the video set. Additionally, test sets and ground truth files will be available for free download for use by other researchers.

8. REFERENCES

[1] A. Borkar, M. Hayes, and M. Smith, "Robust lane detection and tracking with ransac and kalman filter," in *2009 IEEE International Conference on Image Processing*, 2009.

[2] A. M. Muad, A. Hussain, S. A. Samad, M. M. Mustaffa, and B. Y. Majlis, "Implementation of inverse perspective mapping algorithm for the development of an automatic lane tracking system," in *TENCON 2004. 2004 IEEE Region 10 Conference*, vol. 1, 2004, pp. 207–210.

[3] N. N. Zheng, S. Tang, H. Cheng, Q. Li, G. Lai, and F. Y. Wang, "Toward intelligent Driver-Assistance and safety warning systems," *IEEE Intelligent Systems*, vol. 19, no. 2, pp. 8–11, 2004.

[4] National Highway Traffic Safety Administration, "Fatality Analysis Reporting System (FARs) Encyclopedia," 2008. [Online]. Available: http://www-fars.nhtsa.dot.gov/Main/index.aspx

[5] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 3rd ed. Addison-Wesley Pub (Sd), Mar. 1992.

[6] C. C. Wang, S. S. Huang, and L. C. Fu, "Driver assistance system for lane detection and vehicle recognition with night vision," in *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2005)*, 2005, pp. 3530–3535.

[7] H. Wang and Q. Chen, "Real-time lane detection in various conditions and night cases," in *Proceedings of the IEEE Intelligent Transportation Systems Conference*, 2006, pp. 17–20.

[8] A. Borkar, M. Hayes, M., and S. Pankanti, "A layered approach to robust lane detection at night," in *2009 IEEE Workshop on Computational Intelligence in Vehicles and Vehicular Systems*, 2009.

[9] T. Y. Sun, S. J. Tsai, and V. Chan, "HSI color model based Lane-Marking detection," in *IEEE Intelligent Transportation Systems Conference*, 2006, pp. 1168–1172.

[10] Microsoft Visual Studio 2008. [Online]. Available: http://www.microsoft.com/visualstudio/en-us/default.mspx

[11] Federal Highway Administration, "Manual Uniform Traffic Control Devices. Part 3: Markings," Nov. 2004. [Online]. Available: http://mutcd.fhwa.dot.gov/pdfs/2003/Ch3.pdf