Journal of Science and Technology ***49****, Number 5 (2011) 1-12*

STEREO VISION-BASED LANE DETECTION AND TRACKING FOR INTELLIGENT VEHICLES

Chunzhao Guo, Takayuki Yamabe, Seiichi Mita

Department of Electronics and Information, Toyota Technological Institute

Received September 30, 2011

ABSTRACT

Lane detection is one of the key issues for intelligent vehicles. In this paper, we present a lane detection approach designed to navigate an autonomous vehicle through challenging traffic scenes based on stereo vision. In the method, both intensity and geometry cues of the road scenes are utilized and integrated for detecting and tracking the targets based on Hidden Markov Models to deal with challenging conditions and situations. It can capture both painted and physical lane boundaries. Furthermore, the geometry relationships between the stereo camera in the moving vehicle and the road are dynamically estimated and calibrated. Therefore, more accuracy and robustness can be expected in the proposed system. Experimental results in various real challenging traffic scenes show the effectiveness of the proposed system.

Keywords. Lane detection; stereo vision; intelligent vehicles; navigation

1. INTRODUCTION

Roadways are typically marked with painted and physical boundaries to assist the safe and efficient transportations. Real-time lane detection and localization from a moving vehicle using on-board sensor data is one of the key issues for many intelligent transportation systems. Within the last few years, research on this topic has expanded into a wide variety of applications from navigating a fully autonomous driving system to providing road information to a driver assistance system. The problem of finding lanes can be divided into three sub-problems: lane feature detection, lane boundary estimation, and lane tracking [1]. Lane feature detection refers to the use of on-board sensors to extract the road markings, including painted markings and other environmental markings such as color or texture discontinuities that define the safe and legal regions of driving. Lane boundary estimation is that of using the detected lane features to estimate the shape and positions of the lane boundaries based on a certain representation of the road. Lane tracking is to infer the shape and positions of the lanes using the observations. Previous related work for lane detection can be categorized into two main types of methods: LIDAR-based methods [2, 3] and vision-based methods [4, 5, 7 - 10]. LIDAR sensors are useful

in rural areas for helping to detect physical road boundaries, such as roadside fences and curbs. Some systems in DARPA Urban Challenge [6] also use LIDAR to detect painted lane boundaries since LIDAR is insensitive to changing environments so that the reliable detection of the lane boundaries can be expected in practice. However, only changes in the intensity of the returned laser data are indicative of painted road markings while the number of LIDAR beams is very limited. Therefore, such systems usually rely on GPS with other a priori information such as RNDF (Route Network Definition File) or GIS (Geographic Information System). Moreover, the GPS has limitations on the spatial and temporal resolution and the map data may be outdated and inaccurate. While such systems can perform extremely well in certain situations, vision can be utilized to perform well in a wide variety of situations since it can deliver a great amount of information. Vision-based lane detection problems have been extensively studied. McCall and Trivedi provide an excellent survey [7].

In this paper, we present a lane detection approach designed to navigate an autonomous vehicle through challenging traffic scenes based on stereo vision. The proposed approach is developed in the scope of the stereovision-based navigation system integrated in our experimental autonomous vehicles shown in Figure 1. The left one is a super-compact electric vehicle equipped with two computers, a stereo camera, four single beam laser scanners, a GPS and other sensors. The right one is a hybrid vehicle equipped with six computers, a stereo camera, two four-beam laser scanners, a GPS and other sensors.

Figure 1. Our experimental autonomous vehicles

Figure 2 shows the flow diagram of the proposed system. The geometry relationships between the stereo camera and the road are dynamically estimated and calibrated for each and every input image pair to utilize the underlying geometry cues more precisely in the following steps, so as to enhance the accuracy and robustness. Subsequently, an Artificial Neural Network (ANN) classifier is applied to the left and right inverse perspective mapped images (IPM images) in parallel to detect painted lane markings. Next, the Normalized Cross Correlation (NCC) is employed to reveal the geometry cue between the preserved left and right lane markings, which is then integrated with the intensity cue by constructing a weighted graph, reflecting the belief of each pixel as a lane feature. Finally, the left and right lane boundaries are estimated in a Hidden Markov Model (HMM).

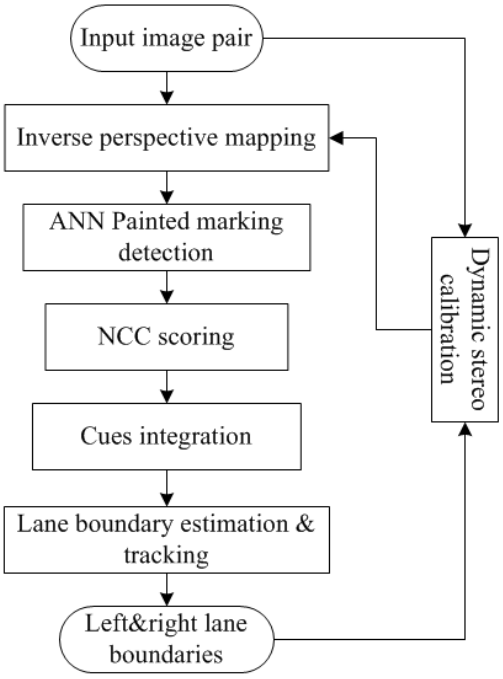

Figure 2. Flow diagram of the proposed approach

2. USE OF GEOMETRY CUES

Most of the vision-based methods use monocular camera to extract the lane boundary by employing features with specific intensity, color and texture as visual cues on the road surface. These methods are mainly supposed to be used in well-arranged environments such as urban streets and highways, since the visual cues can be affected by a number of factors that are not easily measured and change over time, such as the materials, lighting and weather conditions, etc. Therefore, the invariant cues are necessary to overcome these difficulties, especially in challenging traffic scenes. Geometry information is a very important invariant cue, particularly for land vehicles. Stable and accurate reconstruction of the 3D structure by computing the disparity map is difficult due to the requirement of solving the correspondence problem for every pixel. However, for the problem imposed on autonomous vehicle navigation when assuming the road is nearly flat, the complete 3D structure reconstruction is not necessary; all that we need to consider are the planar road plane, since the road region is safe to drive as long as it is flat and the painted lane boundary also lies on the road plane. In the proposed system, in order to utilize geometry cues while maintaining a manageable computational load, we employ inverse perspective mapping (IPM) for lane boundary detection.

2.1. Inverse Perspective Mapping

In the proposed system, we employ IPM for lane boundary detection due to the following three reasons. First, IPM remaps the road points in the left and right images into points in the same world coordinate (Z = 0) and the result image represents a top view of the road region in

front of the vehicle, revealing the geometry relationships of the road between the stereo pair and the road plane. Second, IPM normalizes the size of the lane markings and reduces the range of lane boundaries, which is good for the use of the ANN classifier. Third, the IPM image is accordant with the world coordinate so that the detection results in the IPM image are easy to be used to update the navigation map.

IPM can be performed with the knowledge of the camera, such as viewpoint position, viewing direction, aperture, resolution, etc. [4]. An alternative simpler way for IPM is a linear transformation on homography represented by a 3×3 matrix H_I, as shown in Figure 3, which can correspond each road point u on the image plane to a point x on the road plane π by the following equation,

$$x = H_I u \tag{1}$$

H_I can be derived from a simple external camera calibration with four reference points [11].

2.2. Dynamic Stereo Calibration

Many stereovision-based road and lane detection methods assume that the cameras are calibrated beforehand and the geometry relationships between cameras and the road plane are known and fixed. However, these assumptions are not practical in real applications since the vehicle may tilt and the cameras may vibrate. Therefore, the geometry relationships, e.g. the homography matrices of IPM in the proposed system, must be estimated and calibrated dynamically for each and every frame, since the precise geometry cues are crucial to achieve high accuracy as well as robustness of road detection in challenging traffic scenes.

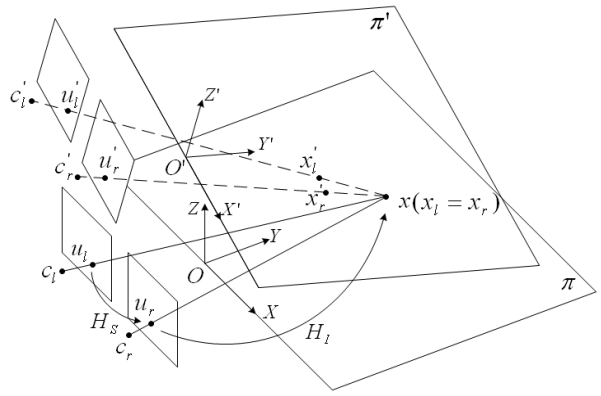

Figure 3. Geometry relationships between cameras and the road plane

As shown in Figure 3, c_l, c_r, π are the normal positions of the left and right cameras and the road plane, respectively. The homography matrices of IPM H_{Il}, H_{Ir} obtained from the external camera calibration can map the image points u_l, u_r to x_l, x_r in the world coordinate and $x_l = x_r$ since u_l, u_r are from the same road point x in the same coordinate. Once the position and pose of the cameras change and move to $c_l^{'}, c_r^{'}$ due to vehicle tilt or camera

vibration, the world coordinate that the pre-calculated H_{Il}, H_{Ir} correspond to also changes along with the cameras from O to O'. In this case, H_{Il}, H_{Ir} will map $u_l^{'}, u_r^{'}$ to $x_l^{'}, x_r^{'}$ and $x_l^{'} \neq x_r^{'}$.

Assume the transformation from O' to O needs to translate by δ and rotate by θ about the X-axis, ϕ about the Y-axis and ψ about the Z-axis, the relationship between O' to O can be described by

$$x = Rx' + T \tag{2}$$

where R, T are the rotation and translation matrices, respectively. We combine (1) and (2) in homogeneous coordinates and then the mapping from a road point u' on the shifted image plane to a point x on the road plane can be represented by the following equation,

$$x = \mathrm{A}\mathrm{H}_I u' \tag{3}$$

where H_I is the pre-calculated homography matrix in homogeneous coordinates and

$$\mathrm{A} = \begin{pmatrix} \cos\phi\cos\psi & \sin\theta\sin\phi\cos\psi - \cos\theta\sin\psi & \cos\theta\sin\phi\cos\psi + \sin\theta\sin\psi & \delta_x \\ \cos\phi\sin\psi & \sin\theta\sin\phi\sin\psi + \cos\theta\cos\psi & \cos\theta\sin\phi\sin\psi - \sin\theta\cos\psi & \delta_y \\ -\sin\phi & \sin\theta\cos\phi & \cos\theta\cos\phi & \delta_z \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{4}$$

In the present case, in order to estimate the current homography matrices of IPM $H_{Il}^{'}, H_{Ir}^{'}$ that can map $u_l^{'}, u_r^{'}$ to x where $x_l^{'} = x_r^{'}$, we estimate the shift vector $S = (\theta, \phi, \psi, \delta_x, \delta_y, \delta_z)$ by minimizing the total intensity error function $E(S)$ in equation (5) between the stereo image pair, using Levenberg-Marquardt Algorithm (LMA) [12].

$$E(S) = \sum_{u_l^{'}, u_r^{'} \in Rc} \sum_k \omega_k (\Phi_l^k(\mathrm{A}\mathrm{H}_{Il} u_l^{'}) - \Phi_r^k(\mathrm{A}\mathrm{H}_{Ir} u_r^{'}))^2 \tag{5}$$

where $\Phi_l(\cdot)$ and $\Phi_r(\cdot)$ are the feature vectors we constructed for the image pair, respectively, and ω_k is the weight of the k-th element. For color images, the vector is nine dimensional that consists of three color values plus a six dimensional color gradient vector in x and y direction. For grayscale images, it is three dimensional that consists of the intensity value plus a two dimensional gradient vector. R_c is the computational region which is actually the planar road region of the image.

3. LANE BOUNDARY DETECTION AND TRACKING

Roadways are typically marked with painted boundaries to assist the safe and efficient transportations, especially in urban streets. In such traffic scenes, vehicles must drive not only safely but also legally. The problem of finding lane boundaries can be divided into three sub-problems: lane feature detection, lane boundary estimation, and lane tracking. In the proposed

system, we apply ANN classifier to detect painted lane markings and the left and right lane boundaries are estimated in an HMM.

3.1. ANN Painted Lane Marking Detection

We apply machine learning to the left and right IPM images in parallel for lane marking detection, since the painted lane boundaries are constructed not randomly but with limited patterns. A comparative study of both classification performance and computation time on various painted markings classification methods is presented in [5]. Based on that, we choose to use an ANN classifier with two layers and seven hidden nodes since it is fastest, whereas the performances are still good. For training, we have gathered image patches of 100 painted lane markings and 100 non-painted markings. For detection, the ANN classifier is applied on a small image patch of 9×3 windows around each and every pixel of the IPM images. Pixels whose ANN scores exceed a loose threshold will be preserved as the painted lane markings. Non-maximum suppression is then applied to the preserved pixels to ensure well-localized positions.

3.2. NCC Scoring Between IPM Images

We compute the NCC value between the left and right processed IPM images to measure the similarity of corresponding pixel locations. The aim of this step is to utilize the underlying geometry cue since the entire lane markings lie on the road plane and all the road points are mapped into the same world coordinate. For each detected lane marking pixel in the left IPM image, the NCC in (6) is computed with the pixel at the same location in the right IMP image.

$$NCC = \frac{\sum_{(i,j)\in W}[f_1(i,j)-\overline{f_1}][f_2(i,j)-\overline{f_2}]}{\sqrt{\sum_{(i,j)\in W}[f_1(i,j)-\overline{f_1}]^2\sum_{(i,j)\in W}[f_2(i,j)-\overline{f_2}]^2}} \tag{6}$$

where, W is the computational window, $f_1(i,j)$ and $f_2(i,j)$ are the image blocks in the left and right IPM images, respectively. $\overline{f_1}$, $\overline{f_2}$ are the average values of the blocks. Pixels whose NCC values exceed a loose threshold will be further preserved as the lane features.

3.3. Cues integration

Sensing the real world is an inherently uncertain process. Many previous approaches model uncertainty for lane estimation based on noisy observations of binary classified lane features, in which false positives are treated equally as the true positives. In the proposed system, we intuitively model the uncertainty in the lane feature detection since the uncertainty happens at the very beginning. In particular, we construct a weighted graph by integrating the intensity and geometry cues, reflecting the belief of each pixel as a lane feature, which assures that each pixel has a weight so as to play different role when estimating lane boundaries using particle filter. In this way, the uncertainty of lane feature detection and the uncertainty of lane boundary estimation can be integrated for the probabilistic reasoning to deal with challenges due to varying appearances of lane boundaries.

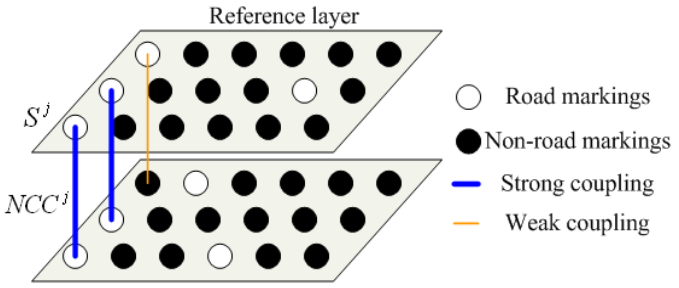

Figure 4. Weighted graph with integrated cues

The weighted graph $G = (V, E, B)$ is shown in Figure 4, where each node $v^j \in V$ represents a pixel, associated with the ANN score S^j as the node weight. Each edge $e^j \in E$ connects pixels at the same location between the left and right IPM images, associated with the NCC value NCC^j. The integrated weight $b^j \in B$ of a node in the reference layer, which is constructed from the left IPM image, is computed as follows,

$$b^j = \alpha_1 \exp(k(1 - NCC^j)) + \alpha_2 S^j / S^{\max} \tag{7}$$

where k is a convergence parameter, α_1, α_2 are retuning parameters for the cues integration. $S^{\max}$ is the maximum value of S^j.

3.4. Boundary estimation based on HMM

Hidden Markov models have been successfully used for the applications in temporal pattern recognition such as speech, handwriting, gesture recognition, etc. In the proposed approach, we formulize the drivable road boundary detection in a HMM, as shown in Figure 5. Unlike the regular HMM, where data is usually over time, the proposed HMM is over the rows of the image from bottom to up, and the states are the horizontal positions of image pixels in a row. The observation is derived from the weighted graph. The most likely sequence of hidden states, i.e. the lane boundaries, lies in the pixels with high confidence.

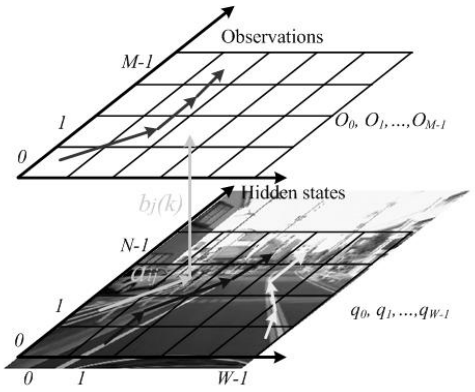

Figure 5. HMM for lane boundary detection. The grids coorespond to the image pixels

We employ the Viterbi algorithm to solve the detection problem in the HMM, which can choose the best state sequence that maximizes the likelihood of the state sequence for the given observation sequence in the following recursive equations,

$$\begin{aligned} X_{0,j} &= b_j(O_0) \cdot \pi_j \\ X_{i,j} &= b_j(O_i) \max_k (X_{i-1,k} \cdot a_{k,j}) \end{aligned} \tag{8}$$

where $X_{i,j}$ is the maximal probability of state sequences of the i columns that end in state j. Taking the log of equation (8), we see the above equations are equivalent to the following,

$$\begin{aligned} X_{0,j} &= V_{0,j} + P_{0,j} \\ X_{i,j} &= V_{i,j} + \max_k (X_{i-1,k} + t(k,j)) \end{aligned} \tag{9}$$

where, $V_{i,j}$ is the observation measure of the pixel (i, j) as a boundary point. $t(k, j)$ is the cost of the path from pixel $(i-1, k)$ to pixel (i, j). $X_{i,j}$ is the best score up to the i-th column.

Here, we define $V_{i,j}$ as

$$\begin{aligned} V_{i,j} &= \lambda_1 b^{(i,j)} + \lambda_2 X^{t-1}_{\mathcal{T}(i,j)} \\ &+ \lambda_3 \left(\frac{\sum_{k=j-\Delta w}^{j} | I_{Il}(i,k) - I_{Ir}(i,k) | - \sum_{k=j}^{j+\Delta w} | I_{Il}(i,k) - I_{Ir}(i,k) |}{2\Delta w} \right) \end{aligned} \tag{10}$$

where, $b^{(i,j)}$ is the integrated weight at pixel (i, j), i.e. the score of the pixel as a lane feature. $X^{t-1}_{\mathcal{T}(i,j)}$ is the probability of the current pixel as a lane boundary point from the previous detection result. $\mathcal{T}$ indicates the transformation according to the motion of the host vehicle from $t-1$ to t. The first term calculates the measure from the image evidence. The second term calculates the measure from the temporal support, i.e. tracking. The third term calculates the intensity gradient value at pixel (i, j) based on the difference between the left and right IPM images, where the points on the road plane will coincide very well while other points do not, according to (1). Therefore, the use of the third term allows the proposed approach to be able to detect the physical boundaries in the scenarios without painted lane markings. wl, wr give the calculation range in the scan line, Δw is a constant number for the gradient computation. $\lambda_1, \lambda_2, \lambda_3$ are weighting parameters. Since the lane boundary should not vary dramatically between adjacent rows, we define $t(k, j) = | j - k |$. The optimum path obtained by Viterbi algorithm corresponds to the drivable road boundary of the nearly flat road.

4. EXPERIMENTAL RESULTS

In the experiments, the proposed approach has been implemented to test a wide variety of typical but challenging scenarios with changing environments without code optimization. The experimental specification is shown in Table 1. The computational time is under 100 ms/frame, which can satisfy the real-time applications. The computational time will be further reduced by

optimizing the codes and utilizing special image processing hardware.

Table 1. Experimental Specifications

CPU	Xeon(R) X5550 @2.67 GHz
Memory (RAM)	12.0GB
Operating system	Window Xp Professional Sp2
Programming language	C++
Experimental data	Stereo images
Resolution	320 × 240
Computational time	< 100 ms/packet

Since the road detection in normal environments is well demonstrated, here, we mainly focus on a wide variety of challenging environments. Totally, we obtained example results for more than 3000 frames, and the proposed system detected the accurate drivable road boundary for more than 97% of the total frames. Typical experimental results are shown in Figures 6-8.

Figure 6. Example results in various lighting conditions and complex on-road patterns

Figure 6 illustrates our approach's ability to overcome noise due to various lighting conditions and localize the lane boundary in spite of strong backlight, heavy shadows, complex on-road patterns, etc. Figure 7 gives the example results on curve/bending/sloped roads, which substantiates the robustness of the proposed approach; despite the flat-road assumption is used. Figure 8 shows the example results on physical/painted lane boundaries, which illustrate our approach can also work well in such environments.

Figure 7. Example results on curve/bending/sloped roads

Figure 8. Example results with physical/painted lane boundaries

5. CONCLUSION

In this paper we presented a stereovision-based system to the problem of accurate and robust detection of the lane boundaries in challenging traffic scenes using integrated cues. Our first contribution is the automatic extrinsic parameters optimization. Our second contribution is the weighted graph constructed with integrated intensity and geometry cues based on a "soft" detection of the lane features, reflecting the belief of each pixel as a lane feature. Our third contribution is the sophisticated measure of the probability of the state sequence to find the most likely lane boundaries. All of the contributions improve the accuracy as well as robustness of lane boundary detection without any a priori knowledge. Experimental results prove the effectiveness of the proposed system on a wide variety of challenging road environments.

REFERENCES

1. A. Huang, S. Teller - Lane Boundary and Curb Estimation with Lateral Uncertainties, Proc. of the 2009 International Conference on Intelligent Robots and Systems, St. Loius, USA, Oct. 2009, pp. 1729-1734.
2. B. Ma, S. Lakshmanan, and A. O. Hero - Simultaneous detection of lane and pavement boundaries using model-based multisensor fusion, IEEE Trans. Intell. Transp. Syst. **1** (5) (2000) 135-147.
3. A. V. Reyher, A. Joos, H. Winner - A lidar-based approach for near range lane detection, Proc. Conf. on Intelligent Vehicles Symposium, 2005, pp. 147-152.
4. M. Bertozzi and A. Broggi - GOLD: A parallel real-time stereo vision system for generic obstacle and lane detection, IEEE Trans. Image Process. **7** (1) (1998) 62–81.
5. Z. Kim - Robust lane detection and tracking in challenging scenarios, IEEE Trans. Intelligent Transportation Systems **9** (1) (2008) 16–26.
6. http://www.darpa.mil/grandchallenge/index.asp
7. J. C. McCall and M. M. Trivedi - Video-based lane estimation and tracking for driver assistance: Survey, system, and evaluation, IEEE Transactions on Intelligent Transport Systems **7** (1) (2006) 20-37.
8. S. Sehestedt, S. Kodagoda, A. Alempijevic, and G. Dissanayake - Robust lane detection in urban environments, Proc. IEEE Int. Workshop on Intelligent Robots and Systems, San Diego, CA, USA, Oct. 2007.
9. S. Sehestedt, S. Kodagoda, A. Alempijevic, and G. Dissanayake - Robust lane detection in urban environments, Proc. IEEE Int. Workshop on Intelligent Robots and Systems, San Diego, CA, USA, Oct. 2007.
10. S. Ieng, J. Tarel, and R. Labayrade - On the design of a single lane-markings detectors regardless the on-board camera's position, Proc. IEEE Intell. Veh. Symp., 2003, pp. 564–569.

11. R. Hartley and A. Zisserman - Multiple View Geometry in Computer Vision. Cambridge, U.K.: Cambridge Univ. Press, 2000.
12. P. Gill and W. Murray - Algorithms for the Solution of the Nonlinear Least-squares Problem, SIAM J. on Num. Ana. (1978) 977-992.

Corresponding author:

Chunzhao Guo,

Department of Electronics and Information, Toyota Technological Institute,

2-12-1, Hisakata, Tempaku-ku, Nagoya 468-8511, Japan

Email: *guo@ toyota-ti.ac.jp*