A Real-Time Approach to Stereopsis and Lane-Finding

Camillo J. Taylor	Jitendra Malik	Joseph Weber
EECS Dept. U.C. Berkeley	EECS Dept. U.C. Berkeley	EECS Dept. Caltech
camillo@cs.berkeley.edu	malik@cs.berkeley.edu	weber@vision.caltech.edu
(510) 642 5029	(510) 642 7597	(818) 395 4859

Abstract— **This paper will report new results we have obtained in applying stereo vision algorithms to the problem of autonomous vehicle navigation on highways. The project consists of two parts: lane extraction and obstacle detection.**

Our lane extraction system is based on a parameterized model for the appearance of the lanes in the images. This model captures the position, orientation and width of the lane as well as the height and inclination of the stereo rig with respect to the road. A robust lane recognition procedure is employed to recover the position of the vehicle within the lane. This scheme is able to recover and track the lane markers in real time (20Hz) even in the presence of a significant number of spurious lane features.

We have developed a new real time stereo system (20Hz) that has been optimized for use in a highway navigation system. The algorithm first extracts vertical line features from a subset of the scan lines in the left and right images. Correspondences are made between line features in the left and right images by a region-based correlation scheme. The results of this process can then be passed to the grouping and tracking process described by Weber, Koller, Luong and Malik in Intelligent Vehicles 95.

The paper will also describe the implementation of these algorithms on our network of TMS320C40 DSPs. This computational architecture allows us to handle the demanding computational and I/O requirements of these applications.

These algorithms have been tested on video data obtained from a test vehicle that was driven in typical highway traffic. Results from these experiments will be presented.

I. Introduction

This paper will report new results we have obtained in applying stereo vision algorithms to the problem of autonomous vehicle navigation on highways. The goal of our project was to design, implement and test a vision system that would take the video signals from a pair of CCD cameras and compute in real time (20 Hz) the position and orientation of the car with respect to the lane and the position of other obstacles on the roadway including other vehicles.

This paper consists of two parts: Section III will describe a real-time stereo system (20 Hz) which computes the position of salient obstacles in the roadway, including other vehicles, with respect to the controlled car while Section IV describes a real-time lane tracking system (20 Hz) which keeps track of the position and orientation of the vehicle with respect to the lane markers.

The estimate for the position of the vehicle within the lane would be used as input to a lateral control system while the positions of the detected obstacles would be passed to a longitudinal control system which would regulate the speed of the vehicle.

A number of researchers have demonstrated vehicle navigation systems that use vision as a primary control input [DM92], [THKS88], [MKLT95], [Pom95]. Our work [WKLM95]is most similar to that of Dickmanns et al in that we make use of models of the roadway and the obstacles in the field of view. However, our systems employ information from a pair of stereo cameras which allows us to directly measure the three-dimensional structure of the scene.

Our algorithms have been implemented on a network of TMS320C40 Digital Signal Processors in order to achieve real-time performance. The implementations are described in more detail in the sequel.

II. Description of imaging setup

Figure 1 shows the stereo configuration that was used to acquire images in our experiments. The vectors X_w, Y_w, and Z_w define the ground plane frame of reference while X_s, Y_s and Z_s represent the frame of reference attached to the stereo rig. The stereo rig was mounted on the car at a height d above the ground plane and at an inclination θ. These parameters depend upon the state of the cars suspension system and may vary over time. In our experiments, the stereo baseline, b was fixed at 10.5 centimeters.

Equations 1 through 4 define the relationship between the coordinates of a point with respect to

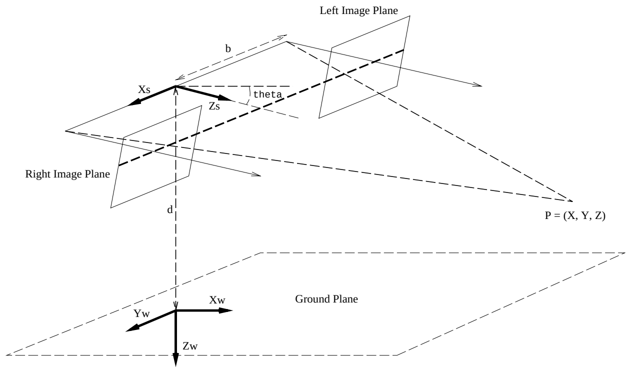


Fig. 1. Stereo configuration used in our experiments.

the stereo rigs frame of reference and the coordinates of its projection in the left and right images.

$$u^l = s_x((X_s + b)/Z_s) + c_x^l \quad (1)$$
$$v^l = s_y(Y_s/Z_s) + c_y^l \quad (2)$$
$$u^r = s_x((X_s - b)/Z_s) + c_x^r \quad (3)$$
$$v^r = s_y(Y_s/Z_s) + c_y^r \quad (4)$$
$$(5)$$

The parameters s_x and s_y refer to the scale factors in the x and y directions on the image plane and are assumed to be the same in both cameras. (c_x^l, c_y^l) and (c_x^r, c_y^r) refer to the images of the centers of projection of the left and right cameras respectively.

Equation 6 gives the relationship between the coordinates of a point with respect to the world frame and its coordinates in the stereo rig frame.

$$\begin{pmatrix} X_s \\ Y_s \\ Z_s \\ 1 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -\sin\theta & 0 & \cos\theta & (d\cos\theta) \\ \cos\theta & 0 & \sin\theta & (d\sin\theta) \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{pmatrix} \quad (6)$$

Consider the set of points that lie on the ground plane, that is, points with world coordinates of the form $(X_w, Y_w, 0, 1)$. We can derive the following equation which relates the stereo disparity of these points, $(u^l - u^r)$, to their row coordinates in the left image, v^l.

$$(u^l - u^r) = \frac{2bs_x\cos\theta}{ds_y}(v^l - h^l) + (c_x^l - c_x^r) \quad (7)$$

Where $h_l = (c_y^l - s_y \tan\theta)$ denotes the projection of the horizon of the ground plane in the left image. This expression is refered to as the *Helmholtz shear equation* [WKLM95] and it places a lower bound on the disparities we would expect between features in the left and right images on any given row of the image pair.

III. The Stereo Algorithm

The goal behind any stereo algorithm is to establish correspondences between points in the left and right images. Once this has been accomplished, it is a simple matter to compute the 3-D coordinates of the matched point via triangulation as shown in Figure 1.

For this application, we have chosen to break the stereo matching procedure into two stages. In the first stage we run a vertical edge extraction procedure on corresponding rows in the left and right images. This is accomplished by convolving the image rows with a Canny edge filter [Can86] and selecting the local maxima above a certain threshold in the resulting array. This is typically the most computationally intensive stage in our algorithm, it takes approximately 1 millisecond per row on our C40 network.

In the second stage of the matching process, we compare each edge in the left image row to potential correspondents among the edges in the right image row. The candidate matches are evaluated by computing the normalized correlation between a window of pixels centered around each edge. These windows are typically 20 to 30 pixels wide. The stereo algorithm selects the best correspondent for each edge in the left image on the basis of these correlation values; correlation values below a certain threshold are rejected as unreliable.

The main advantage of this scheme is that it dramatically lowers the computational complexity of the stereo process. Since we only compare edges in the left and right images, the stereo algorithm does not have to consider nearly as many possible matches as it would otherwise have to. Another reason for considering matches between regions with significant contrast changes is that correlation based matching schemes are most accurate when they are applied to these types of features. Similar edge based schemes have been successfully applied in indoor environments and at slower frame rates by a number of researchers [KTB89], [CSSP92].

The results of this stereo matching procedure could be passed to a grouping algorithm such as the one described by Weber, Koller, Luong and Malik [WKLM95] which would group the matched edges into coherent obstacles which could be tracked over time. Temporal integration schemes like the Kalman filter could be employed to improve our estimates for the position and velocity of these obstacles with respect to the test vehicle.

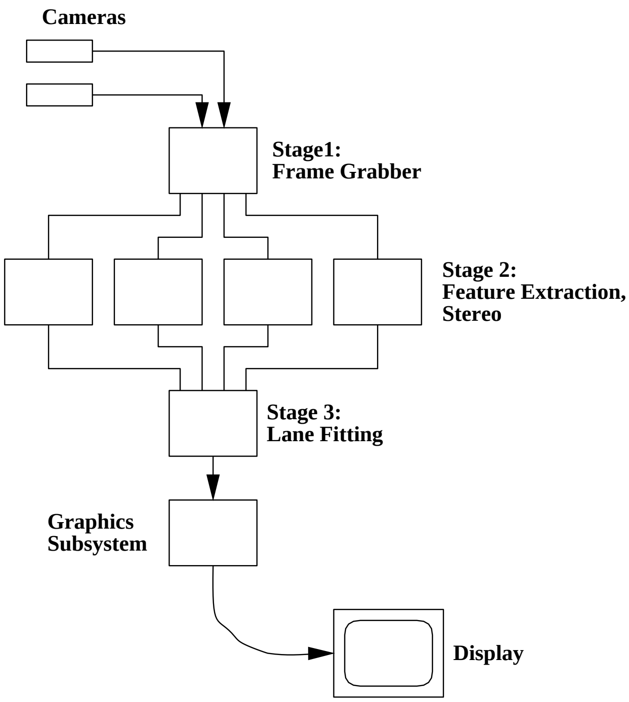
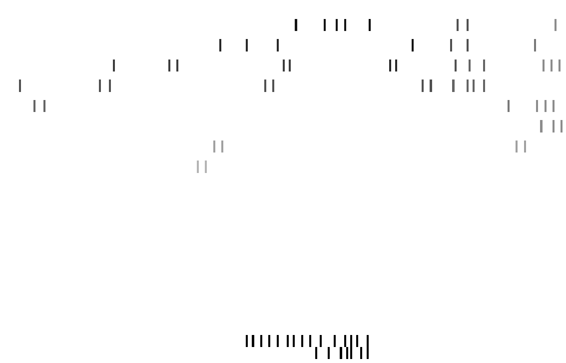


Fig. 2. Network of TMS320C40s used to implement the real-time lane tracking and stereo systems.

We have chosen to implement our stereo algorithm on a network of TMS320C40 Digital Signal Processors which are arranged in a network as shown in Figure 2. The processors are arranged in a pipeline to improve system throughput. The processor at the head of the pipeline is equipped with a frame grabber which it uses to capture pairs of images from the left and right cameras. It then sends the image data on to a set of four processors which constitute the second stage of the pipeline. Processors in this stage perform the feature extraction and stereo matching procedures. The stereo results are collected together by the processor at the third stage of the pipeline which passes the results on to the final stage where the results are displayed on a VGA monitor. This organization allows us to exploit the parallelism inherent in the stereo algorithm since each of the four processors in stage two computes stereo correspondences for a different region of the image. Additional parallelism is realized by allowing processors at different stages in the pipeline to work independently so that multiple stereo pairs can be processed simultaneously.

Our current system captures and processes stereo pairs from the cameras at a rate of 20 frames per second. Each of the images in the pair is digitized at a resolution of 340 columns by 240 rows. The system performs its stereo calculations on 20 of the rows in the image pair as shown in figure 3. There is a delay of 100 milliseconds between the time that an image is captured and the time that the results of the stereo computation are made available.

(a)



(b)

Fig. 3. (a) The left image of a stereo pair (b) The vertical edge features that the system extracted and found correspondences for in the right image, the gray value of the edge segment corresponds to its disparities, brighter features have larger disparities

IV. Lane Recognition

The lane tracking module is designed to provide estimates for the position and orientation of the car within the lane. Our approach to lane tracking is based upon a parametric model of the lane geometry, the tracking algorithm computes estimates for the parameters of this model from feature measurements in the left and right images. At each time instant, the lane tracker predicts where the lane markers should appear in the current image based on its previous estimates for the lane position. It then extracts possible lane markers from the left and right images. These feature measurements are passed to a robust estimation procedure which recovers the parameters of the lane along with the orientation and height of the stereo rig with respect to the ground plane.

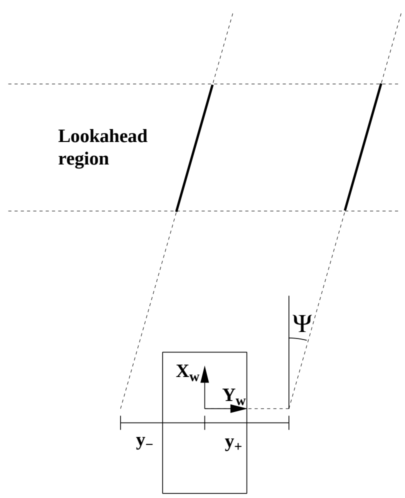


Fig. 4. Figure showing the relationship between the vehicle and the highway lane

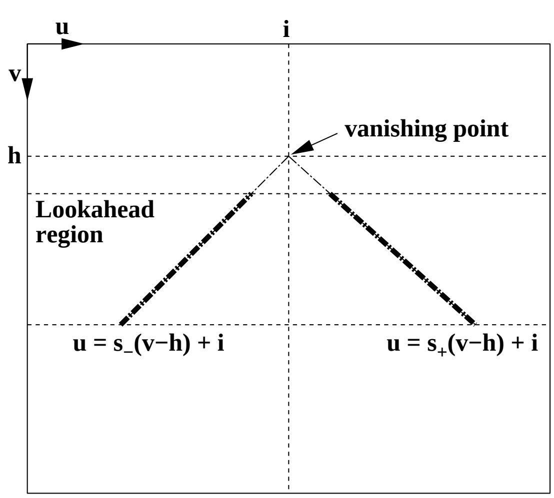


Fig. 5. The appearance of the roadway in the image plane

Figure 4 shows the relationship between the vehicle and a section of the highway lane some distance ahead while Figure 5 shows how that portion of the highway would appear in the left camera.

Note that the model for the appearance of the lane markers in the image is captured by four parameters: two parameters, h^l and i^l which represent the vanishing point of the lane in the image and another two, s^l_+ and s^l_- which denote the slopes of the left and right lane markers respectively. Equations (9) through (11) describe the well known relationship between these image parameters and the parameters of the road and camera models.

$$h^l = c^l_y - s_y \tan\theta \qquad (8)$$

$$i^l = c^l_x - s_x \frac{\tan\psi}{\cos\theta} \qquad (9)$$

$$s^l_+ = \frac{s_x}{s_y}\{\cos\theta(\frac{y_+ + b}{d}) - \sin\theta\tan\psi\} \qquad (10)$$

$$s^l_+ = \frac{s_x}{s_y}\{\cos\theta(\frac{y_- + b}{d}) - \sin\theta\tan\psi\} \qquad (11)$$

Note that the vanishing point in the image is simply the projection of the road direction vector onto the image plane. As such, there is a particularly simple relationship between the coordinates of the vanishing point, h^l and i^l, the pitch and yaw parameters, θ and ψ. The slopes of the left and right lane markers in the image, s^l_+ and s^l_- indicate the lateral position of the vehicle with respect to these lines.

Prior knowledge of the range of values that the model parameters can assume serves to limit the range of possible values for the image parameters. More specifically, knowledge about the range of reasonable values of roll and pitch angles limits the area in the image in which the vanishing point could lie while knowledge about the range of road widths and the approximate height of the stereo rig above the ground limits the difference in slope between the left and right lane markers.

The first stage of the lane tracking system is responsible for detecting and localizing possible lane markers in the left and right images. The lane markers are modeled as white bars of a particular width against a darker background. Regions in the image which satisfy this intensity profile can be identified through a template matching procedure. It is important to remember that the width of the lane markers in the image changes linearly as a function of the image row. This means that different templates are used for different rows in the left and right images. The Helmholtz shear equation can be used to verify that candidate lane markers actually lie on the ground plane.

Once a set of candidate lane markers has been recovered, the lane tracker applies a robust fitting procedure to find the set of model parameters which best match the observed data. A robust fitting strategy is absolutely essential in this application because on real highway traffic scenes the feature extraction procedure will almost always return a number of extraneous features that are not part of the lane structure. These extra features can come from a variety of sources, other vehicles on the highway, shadows or cracks in the roadway etc. These distractions can confuse naive estimation procedures based on least squares techniques.

A number of other researchers [KL95], [PJ92] have also proposed robust estimation techniques for road recognition. Most of these techniques have been too computationally demanding for real time implementation. Our research demonstrates that it is in fact possible to implement such schemes in real-time.

The lane fitting procedure is divided into two stages. In the first stage, linear Hough transforms are performed on the left and right lane markers independently. The Hough transform procedure computes the scores associated with a set of candidate lines through the observed lane markers. These scores indicate how well the lines conform to the observed data.

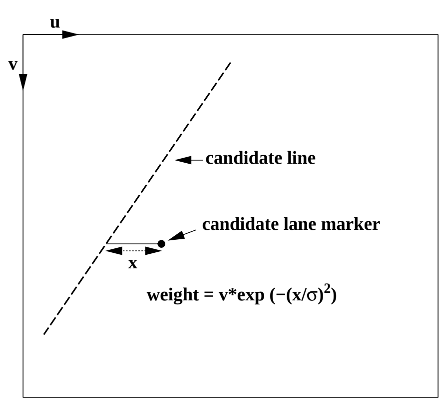


Fig. 6. Figure showing how a straight line is ranked by a particular image measurement

For each candidate line, the contribution of a given image measurement is based on the lateral distance between that measurement and the line and is weighted by the position of the feature in the image as shown in Figure 6. Feature points that are lower in the image (and hence closer to the vehicle) are given a greater weight than lane markers that are further away. This strategy biases the extraction procedure towards solutions that fit more closely in the near field. It also reflects the fact that features in the near field are considered more reliable because they are larger and can be localized more accurately.

In the second stage of the fitting procedure, the system selects a pair of candidate lines, one for the left lane marker and one for the right, that satisfy all of the applicable constraints and which have the best combined score. More specifically, the system ensures that the two selected lines will intersect within a particular area on the image determined by constraints on the pitch and yaw angles. It also guarantees that the difference between the slopes of the left and right lane markers will lie within a specified range which reflects the fact that there are constraints on the width of the highway lane and the height of the stereo rig. The combined score for the lane interpretation provides an indication of the systems confidence in its estimate.

If one of the lane markers is occluded then the system estimates the probable location of that lane from the position of the other lane marker and from a set of default values for the height of the stereo rig, the pitch angle, and the lane width.

A simple weighted averaging scheme is used to combine the last 20 measurements of the image parameters into a single estimate, the score associated with each estimate is used as a weighting factor. This temporal averaging scheme helps to eliminate high frequency noise in the measurements.

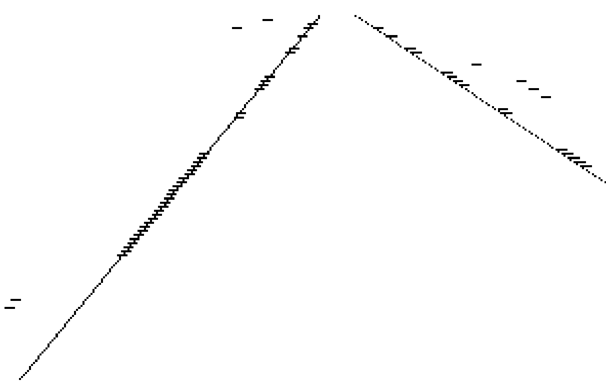
The lane tracking system has been implemented on the network of TMS320C40 Digital signal processors described in the previous section. Once again, the computation has been divided into a four stage pipeline. The processor at the head of the pipeline captures stereo images from the video source and distributes them to the four processors in the second stage. The processors in stage two extract candidate lane markers from different regions of the stereo pair. The processor in stage three collects the results from stage two and performs the robust fitting procedure described earlier. The stage four processor is responsible for displaying the extracted lane markers and the estimates for the lane position on a VGA monitor.

V. Conclusions

This work demonstrates the feasibility of using a real-time stereo system to recover the position of a vehicle within a highway lane and to detect obstacles in the roadway.

We have presented an approach to recovering sparse depth maps from stereo data in real time using vertical edge features. This approach works well on typical highway scenes where the vast majority of objects of interest contain some form of vertical structure. Our current implementation

(a)



(b)

Fig. 7. (a) The left image of a stereo pair (b) The lane markers horizontal markings represent candidate lane markers in the image while the straight line represents the tracking systems estimate for the position of the lanes

can capture and process stereo pairs at a rate of 20 Hz.

We have also developed a robust feature based lane recognition system which works in real time (20Hz) on our network of TMS320C40 Digital Signal processors. The system is able to recognize and track the roadway even in the presence of a large number of spurious markings by employing robust estimation techniques based on the Hough transform.

Our current research work is directed towards the development of control algorithms which would use the data provided by the vision system for lateral and longitudinal control tasks.

Acknowledgments: This work was supported by CALTRANS through the PATH program under grants MOU 131 and MOU 257 and by the AHS Consortium under AHS-B3.

References

[Can86] John Canny. A computational approach to edge detection. *IEEE Trans. Pattern Anal. Machine Intell.*, pages 679–98, November 1986.

[CSSP92] James L. Crowley, Patrick Stelmaszyk, Thomas Skordas, and Pierre Puget. Measurement and integration of 3-D structures by tracking edge lines. *International Journal of Computer Vision*, 8(1):29–52, July 1992.

[DM92] Ernst D. Dickmanns and Birger D. Mysliwetz. Recursive 3-d road and relative ego-state recognition. *IEEE Trans. Pattern Anal. Machine Intell.*, 14(2):199–213, February 1992.

[Ken89] S.K. Kenue. Lanelok: Detection of lane boundaries and vehicle tracking using image-processing techniques. In *SPIE Mobile Robots IV*, 1989.

[KL95] Karl Kluge and Sridhar Lakshmanan. A deformable-template approach to lane detection. In *Proceedings Intelligent Vehicles 1995*, pages 54–59, 1995.

[KT92] K. Kluge and C. Thorpe. Representation and recovery of road geometry in yarf. In *Proceedings Intelligent Vehicles 1992*, pages 114–119, 1992.

[KTB89] David J. Kriegman, Ernst Triendl, and Thomas O. Binford. Stereo vision and navigation in buildings for mobile robots. *IEEE Trans. on Robotics and Automation*, 5(6):792–803, December 1989.

[Mas92] Ichiro Masaki. *Vision-Based Vehicle Guidance*. Springer Verlag, 1992.

[MKLT95] Larry Matthies, Alonzo Kelly, Todd Litwin, and Greg Tharp. Obstacle detection for unmanned ground vehicles: A progress report. In *Proceedings Intelligent Vehicles 1995*, pages 66–71, 1995.

[PJ92] Amy Polk and Ramesh Jain. A parallel architecture for curvature-based road scene classification. In Ichiro Masaki, editor, *Vision-Based Vehicle Guidance*, chapter 14, pages 284–299. Springer Verlag, 1992.

[Pom95] Dean Pomerleau. Ralph: Rapidly adapting lateral position handler. In *Proceedings Intelligent Vehicles 1995*, pages 54–59, 1995.

[THKS88] Chuck E. Thorpe, Martial Hebert, Takeo Kanade, and Steve Shafer. Vision and navigation for the Carnegie-Mellon navlab. *IEEE Trans. Pattern Anal. Machine Intell.*, 10(3):362–373, May 1988.

[WKLM95] Joseph Weber, Dieter Koller, Q.T. Luong, and Jitendra Malik. New results in stereo-based automatic vehicle guidance. In *Proceedings Intelligent Vehicles 1995*, pages 530–535, 1995.