Title
A Combined Approach To Stereopsis And Lane-finding

Authors
Malik, Jitendra
Taylor, Camillo J.
Weber, Joseph
et al.

Publication Date
1997

CALIFORNIA PATH PROGRAM
INSTITUTE OF TRANSPORTATION STUDIES
UNIVERSITY OF CALIFORNIA, BERKELEY

A Combined Approach to Stereopsis and Lane-Finding

Jitendra Malik, Camillo J. Taylor, Joseph Weber, Dieter Koller, Quang-Tuan Luong
University of California, Berkeley

California PATH Research Report
UCB-ITS-PRR-97-27

This work was performed as part of the California PATH Program of the University of California, in cooperation with the State of California Business, Transportation, and Housing Agency, Department of Transportation; and the United States Department of Transportation, Federal Highway Administration.

July 1997

ISSN 1055-1425

A Combined Approach to Stereopsis and Lane-Finding
A final report on M.O.U. **131**

Principal Investigator: Prof. Jitendra Malik
EECS Dept. U.C. Berkeley

Postdoctoral Researcher: Camillo J. Taylor
EECS Dept. U.C. Berkeley

Postdoctoral Researcher: Joseph Weber
EECS Dept. U.C. Berkeley

Postdoctoral Researcher: Dieter Koller
EECS Dept. U.C. Berkeley

Postdoctoral Researcher: Quang-Tuan Luong
EECS Dept. U.C. Berkeley

March 31, 1997

Contents

List of Figures

PATH Goal Statement

The research reported herein is a part of the Program on Advanced Technology for the Highway, PATH, within the Institute of Transportation Studies, at the University of California, Berkeley. PATH aims to increase the capactity of the most used highways, to decrease traffic congestion, and to improve the safety and air quality. It is evolutionary and voluntary. It is a cooperative venture of automakers, electonic companies, local, state and federal governments and universities.

Disclaimer Statement

The contents of this report reflect the views of the authors who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the STATE OF CALIFORNIA or the FEDERAL HIGHWAY ADMINISTRATION. This report does not constitute a standard, specification, or regulation.

Chapter 1

Introduction

This report describes the results of the research efforts carried out under MOU131. The overall goal of this resarch was to develop a machine vision system for use in vehicle control applications. Towards this end we performed research to extend and improve the capabilities of machine vision algorithms. We also demonstrated real-time implementations of some of these algorithms which could actually be used as part of a closed-loop control system.

In this report we propose a new approach for vision based longitudinal and lateral vehicle control. The novel feature of this approach is the use of binocular vision. We integrate two modules consisting of a new, domain-specific, efficient binocular stereo algorithm, and a lane marker detection algorithm, and show that the integration results in a improved performance for each of the modules.

Longitudinal control is supported by detecting and measuring the distances to leading vehicles using binocular stereo. The knowledge of the camera geometry with respect to the locally planar road is used to map the images of the road plane in the two camera views into alignment. This allows us to separate image features into those lying in the road plane, e.g. lane markers, and those due to other objects which are dynamically integrated into an obstacle map. Therefore, in contrast with the previous work, we can cope with the difficulties arising from occlusion of lane markers by other vehicles. The detection and measurement of the lane markers provides us with the positional parameters and the road curvature which are needed for lateral vehicle control. Moreover, this information is also used to update the camera geometry with respect to the road, therefore allowing us to cope with the problem of vibrations and road inclination to obtain consistent results from binocular stereo.

Chapter 2 of this report describes the research that has been done on imrpoving and extending the capabilities of our machine vision system. Chapter **3** describes our real-time implementation of lane-finding and stereopsis capabilities. Chapter 3 discusses some of our conclusions and future work.

Chapter 2

An Integrated Stereo-Based Approach to Automatic Vehicle Guidance

2.1 Introduction

We propose an approach and develop a system for vision based longitudinal and lateral vehicle control which makes extensive use of binocular stereopsis. Former work on autonomous vehicle guidance has mostly concentrated on road following. Some of these approaches [DM92, Pom92] perform very well in uncrowded traffic scenes, where lane markers are clearly visible and not obstructed by other vehicles. However, in real traffic scenes, other vehicles are usually present, and this raises two problems. First, they are potential obstacles, which are to be detected. This problem has been addressed using optical flow interpretation [Enk90, CE90], stereopsis [ZJB+90, OOS+90, Mat92], or a combination of both [CMT91]. These approaches are often computationnaly expensive. Second, lane markers are often obstructed by other vehicles, which might defeat the algorithms which do not take explicitly occlusion into account.

Novel aspects of our system include (a) exploitation of domain constraints to simplify the search problem in finding binocular correspondences (b) temporal integration of the results of the stereo analysis to build a reliable depth map of obstacles (c) dealing with crowded traffic scenes where substantial segments of the lane boundaries may be occluded (d) on-line updating of the external camera calibration with respect to the road. The vision system is designed to interface in a modular fashion with the use of non-visual sensors such as magnetic sensors for lateral position measurement and active range sensors (e.g. Doppler radar) for an integrated approach to vehicle control such as that being investigated in the California PATH project.

Longitudinal control — i.e. maintaining a safe, constant distance from the vehicle in front — is supported by detecting and measuring the distances to leading vehicles using binocular stereopsis. A known camera geometry with respect to the locally planar road is used to map the images of the road plane in the two camera views into alignment. Any significant residual image disparity then indicates an object not lying in the road plane and hence a potential obstacle [MBLB91]. This approach allows us to separate image features into those lying in the road plane, e.g. lane markers, and those due to other objects. In the absence of this separation, image features due to vehicles which happen to lie in the search zone for lane markers would corrupt the estimation of the road boundary contours.

The features which lie on the road are stationary in the scene and appear

to move only because of the egomotion of the vehicle. Measurements on these features are used for dynamic update of (a) the lateral position of the vehicle with respect to the lane markers (b) the camera parameters in the presence of camera vibration and changes in road slope. Lane markers are detected and used for lateral control, i.e. following the road while maintaining a constant lateral distance to the road boundary [DM92, Tho90, Pom92, RH91, APR92]. For that purpose we model the road and hence the shape of the lane markers as clothoidal curves. The system gives a precise estimate of all the positional and dynamical parameters of the lane markers. In particular, the flow of the lane markers is computed (time derivative of the position and orientation of the lane in the mobile car's reference frame) and the important look-ahead information consisting of road curvature is also obtained. The lane is sensed with a look ahead distance of 60 meters, and even at this distance, all the road parameters are estimated.

The binocular measurement of the lane markers enables us to perform an on-line updating of the external geometric parameters of the stereo rig with respect to the road. We can therefore deal with change in slope of the road, as well as variations in inclination of the car caused by bumps or accelerations and decelerations. This allows us to maintain a high accuracy of measurement within our domain-specific stereo algorithms.

Outline of our approach

The idea behind our approach is to build a reliable and efficient system by exploiting a number of geometric constraints which arise from the configuration of our stereo rig, and from the fact that the road can be modeled locally as a plane. These geometric constraints are detailed in Sec. 3.2.

At each new instant, we first compute the stereo disparity using an efficient algorithm based on the Helmholtz shear (Sec. **3.3).** The disparity map is used in two ways. First, a 3D obstacle map is dynamically updated over time by tracking identified vehicles and introducing new vehicles which appear. (Sec. 2.4). This provides the information needed for longitudinal control, ie measuring the distances to leading vehicles. Second, the areas of the image belonging to the ground plane are identified. This ensures that the search area for lane markers (which is defined using the parametric description of the lane markers which was found at the previous instant) is not corrupted by occlusions. Within this area, the lane markers are localized

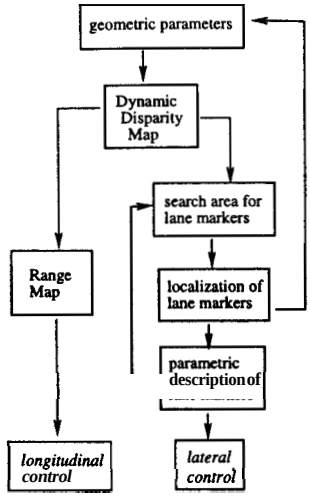


Figure 2.1: The flow of information in our integrated approach

by a specialized feature detector (Sec. 2.5). From the image positions of the lane markers, we can update the geometric parameters of the stereo rig (Sec. 2.6) The new parameters will be used to compute the stereo disparity at the next instant, and to map the lane markers to the ground plane, where a parametric description is obtained for them (Sec. 2.5). This parametric description provides the information needed for lateral control, ie maintaining a constant distance to the road boundary. The flow of information that we just described is summarized in Fig. 2.1.

2.2 The geometrical model

2.2.1 A stereo rig viewing a plane

In our application, the vision system consists of a binocular stereo rig. The road surface plays an important role, since it contains the lane markers to be tracked for lateral control, and since every object which lies above it is to be considered as a potential obstacle. Our key assumption is that this surface can be locally modeled as a plane.

The camera is modeled as a pinhole camera using the projective linear

model. There is a one-to-one correspondence between the image plane $\mathcal{R}_1$ and a given plane Π, and that this correspondence is given by the homography:

$$\mathbf{m}_1 = \mathbf{H}_{\Pi 1}\mathbf{M}_\Pi$$

where $\mathbf{m}_1$ (resp $\mathbf{M}_\Pi$) are the projective coordinates of a point of $\mathcal{R}_1$ (resp Π). In the case of two cameras, we get that the two images $\mathbf{m}_1$ and $\mathbf{m}_2$ of a point $\mathbf{M}_\Pi$ on a given plane Π are related by the homographic relation:

$$\mathbf{m}_2 = \mathbf{H}_{12}\mathbf{m}_1$$

2.2.2 The Helmholtz shear

In a particular case, this relation reduces to what we call the Helmholtz shear, a configuration where the process of computing the stereo disparity is tremendously simplified. We have chosen this term to acknowledge the fact that this insight is due to Helmholtz [Hel25] more than a hundred years ago. He observed that objectively vertical lines in the left and the right view perceptually appear slightly rotated. This led him to the hypothesis that the human brain performs a shear of the retinal images in order to map the ground plane to zero disparity. Then, any object above the ground plane will have non-zero disparity. This is very convenient because the human visual system is most sensitive around the operating point of zero disparity.

In the most general situations where the Helmholtz shear applies, the correspondence between two views of a point of the road plane can therefore be described by the relation:

$$\begin{cases} u' = \mathrm{u} + h_{12}v + h_{13} \\ v' = \mathrm{v} \end{cases} \tag{2.1}$$

which means that the matrix $\mathbf{H}_{12}$ takes the special form:

$$\mathbf{H}_{12} = \mathbf{I}_3 + \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} 0 & h_{12} & h_{13} \end{bmatrix}$$

It is known that the expression for the general homography is:

$$\mathbf{H}_{12} = \mathbf{A}'(\mathbf{R} + \frac{1}{d}\mathbf{T}\mathbf{n}^T)\mathbf{A}^{-1} \tag{2.2}$$

In this expression, $\mathbf{A}$ (resp $\mathbf{A'}$) is the matrix of intrinsic parameters of the first (resp. second) camera. The motion parameters $\mathbf{R}$ and $\mathbf{T}$ describe the displacement between the two cameras. The equation of plane Π is $\mathbf{n}^T\mathbf{M} = d$, where n is the unit normal vector of the plane and d the distance of the plane to the origin. From this expression, one can easily see that the correspondence $\mathbf{H}_{12}$ is a *Helmholtz shear*, if and only if

- the intrinsic parameters $\mathbf{A}$ and $\mathbf{A'}$ are the same'
- the rotation $\mathbf{R}$ is the identity
- the translation $\mathbf{T}$ has only a component along the X-axis
- the component of the normal n along the X-axis is zero

In such a situation, the stereo rig is entirely characterized by the following parameters:

- intrinsic parameters of the first camera $\mathbf{A}$
- baseline b

The position of the plane with respect to the stereo rig can be described by two parameters (that we will call the geometric parameters), for instance:

- the height of the stereo rig with respect to the road plane d
- the angle of tilt of the stereo rig with respect to the road plane α

They are related to the coefficients of the Helmholtz shear by:

$$\begin{cases} h_{12} = b/d \sin \alpha \\ h_{13} = b/d \cos \alpha \end{cases} \tag{2.3}$$

'Actually with the exception of the parameters u_0 and u'_0 which may differ. In the sequel, we will just suppose that $\mathbf{A}$ and $\mathbf{A'}$ are totally identical.

2.3 Binocular Stereopsis

Although proposed algorithms in the literature for computing binocular stereopsis are quite computationally expensive we are able to reduce the complexity considerably by using region-of-interest processing and exploitation of domain constraints.

The process of computing the stereo disparity is tremendously simplified by using the *Helmholtz shear* described in Sec. **3.2.** After applying this very simple transformation to the image, obstacles get mapped to points of non-zero disparity making them very easy to detect.

The disparity is found by computing the normalized correlation between small horizontal windows in the two images at the locations of the points-of-interest. The normalized correlation for disparity shift τ, at horizontal image location x is:

$$\rho_x(\tau) = \frac{\sigma^2_{l,r}(x+\tau, x)}{\sigma_{l,l}(x+\tau, x+\tau)\sigma_{r,r}(x,x)} \tag{2.4}$$

where the correlations $\sigma_{i,j}(x, \mathrm{y})$ are approximated by summations

$$\sigma^2_{i,j}(x,y) = \sum_{u=-W/2}^{+W/2} g_i(x+u)g_j(y+u) \tag{2.5}$$

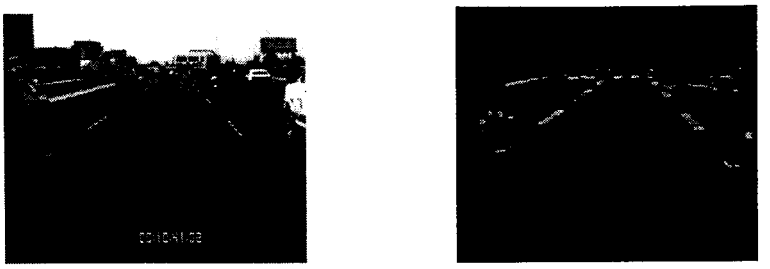
which are calculated over a window of size W.

Points-of-interest are those locations in the right image where the value of $\sigma^2_{r,r}$ is above a threshold. The normalized correlation function is calculated only in those regions. Sub-pixel disparities are obtained by quadratic interpolation of the function about the maximum τ.

Residual disparities — which appear in the image after the ground plane disparity has been mapped to zero — indicate objects which appear above the ground plane. A simple threshold is used to distinguish between features lying on the ground plane (e.g. lane markers or other stuff painted on the road) and features due to objects lying above the ground plane (which may become future obstacles). Figure 2.2 shows the result on a single frame.

2.4 Temporal integration

Computing depth from just a pair of images is known to be sensitive to noise. One can improve the accuracy of the depth estimation by exploiting the



a) b)

Figure **2.2:** a) left image and b) light indicates objects were detected to be on the road surface, dark indicates objects are above the road surface, black indicate regions where the disparity could not be accurately recovered.

temporal integration of information with time using the expected dynamics of the scene via Kalman filters. Objects of interest will be assumed to be either other vehicles on the road or stationary objects connected to the road plane. In addition we can exploit the physical constraints of the environment. We can assume we are interested in connected, rigid objects. This would allow us to use spatial coherence in identifying objects from the depth map.

We utilize the spatial coherence of objects in order to segment the depth map into objects of interest. First, connected components are found in a ***3D*** space consisting of the two image dimensions plus the depth dimension. In the two image dimensions, points are connected if they are one of the 4 nearest neighbors. In the depth dimension they are connected if the difference in depth is less than the expected noise in the depth estimates. Figure 2.4 gives an example of two objects which are connected in this image/depth **3 0** space.

These connected components form the basis of potential objects which are to be tracked with time. If the same object appears in two consecutive frames, we can initialize a Kalman filter to track its position and velocity with respect to our vehicle. Figure 2.4 show the objects found by this method.

2.5 Lane marker detection and tracking

To avoid the computational expenses and the sensitivity to the vertical vibration induced by the car suspension which plague the methods based on

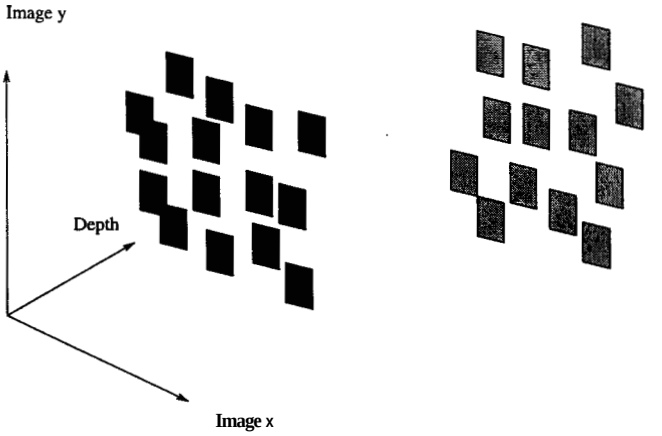


Figure **2.3:** Connected components in image/depth space consist of those pixels which are nearest neighbors in image coordinates as well as having depth differences less than depth uncertainty.

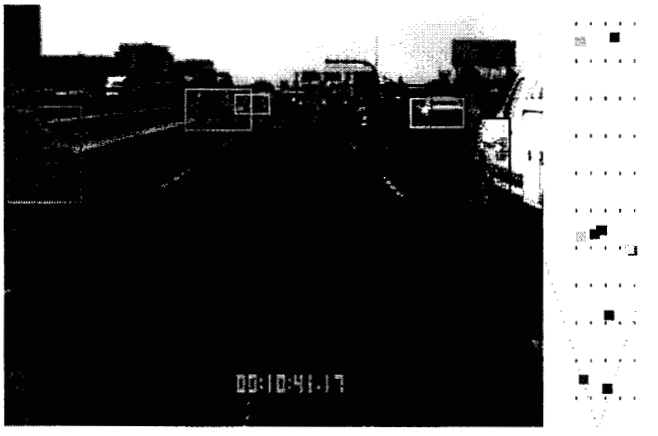


Figure 2.4: Objects identified as being in the same lanes of traffic as the test vehicle. On the right side of the image is a "birds-eye-view" from above the road surface showing the relative position of the tracked objects with respect to the test vehicle.

optical flow, our approach to lateral control relies on the evaluation of the global relative motion of the lane markers. The goal of lane marker detection and tracking is to provide us with the lane flow information. We need therefore to compute, in the coordinate system linked to the car:

- the horizontal offset of the lane marker
- the orientation of the lane marker

In addition to these parameters, we also take into account road curvature that we want to detect as far ahead as possible.

When the system is started (or reinitialized in case of detected unconsistencies), we assume that we are in the most usual configuration where the car is on a straight portion of the road. We search for the position of potential line markers as straight lines, using the fact that they are parallel. The algorithm, which uses only simple voting and clustering operations, is based on the image of Gaussian derivatives, which was already computed to find the features during disparity computation. An example is shown Fig. 2.5.

In the steady state mode, the tracking is based on a predict and verify procedure. The following loop continually executes:

- *Predict* new parameters for each lane marker.
- Define horizontals *search* bands in the image (an example is given Fig. 2.5) based on the predicted parameters, their uncertainties, and the previous homography matrices. The information provided by stereopsis is used here to exclude points which are images of obstacles.
- *Localize* within the search zone the center of line markers, using a bright bar model. This is done in two stages. First, we localize the center of the marker by performing a local convolution with an approximation of the elongated 2D filter consisting of the product of a second derivative of a Gaussian (perpendicular direction) by a Gaussian (direction parallel to the lane marker). Then, to select the final position, we use a model-based approach. Within the expected interval of width, we compare the brightness value and variance of the bar and of the background. An example of the points found is shown Fig. 2.5.

- Backproject the points just found to the ground plane using the updated homography matrices and the predicted parameters. *Fit* a new clothoid model to the backprojected points. To avoid large-scale non-linear minimization, specific techniques are used. First, the offset and orientation are computed by determining the tangent at the origin of the curve through a LMedS line fitting. Then to reduce the search space for the curvature parameters, we take into account the fact that they can only evolve in a limited number of ways, depending on the type of the portion of the road where the vehicle is (successively: straight, curve entry, middle of a curve, end of a curve). An example of a portion of clothoid found is shown in Fig. **2.6.**

- *Update* the model parameters

The road model that we use is based on the the actual road layouts widely used in civil engineering to produce high-speed roads. Each of the line markers detected is modeled as a plane curve which is characterized by the **5** parameters:

- O: lateral offset of the line in the car's coordinate system

- θ: angle between the direction of the line at the closest position and the line of sight of the car

- C_1, C_2, C_3 curvature at the beginning, middle, and end of the observation zone.

The clothoid model, which is used in road designing, consists in assuming that the curvature along the road is a continuous function of arc length **,** with a piecewise constant variation a;: $C(s) = C_1 + a_i s$. The constant ¡changes only at the middle of the observation zone. The model can represent accurately straight lines, arc of circles, and the transitions between them. In spite of the small number of parameters, and of these simplifications, the model is more accurate than assuming just zero curvature [TMGM88, Cri90, Ken89] or parabolic sections [KT92].

The computation of the actual 3D points from the values of the model parameters is quite expensive, since curvature is a second-order quantity, and thus to have a fast access to the curve from its parameters we use a look-up-table which is precomputed.

On a typical example, the results of curvature estimation are fairly consistent with what could be expected, as can be seen in Fig. 2.6. **As** shown in Fig. 2.7, occlusion of lane markers might lead to spurious values of the curvature parameters. Removing the occluded points using using the residual disparity with respect to the ground plane enables us to find that the lane being tracked is straight.

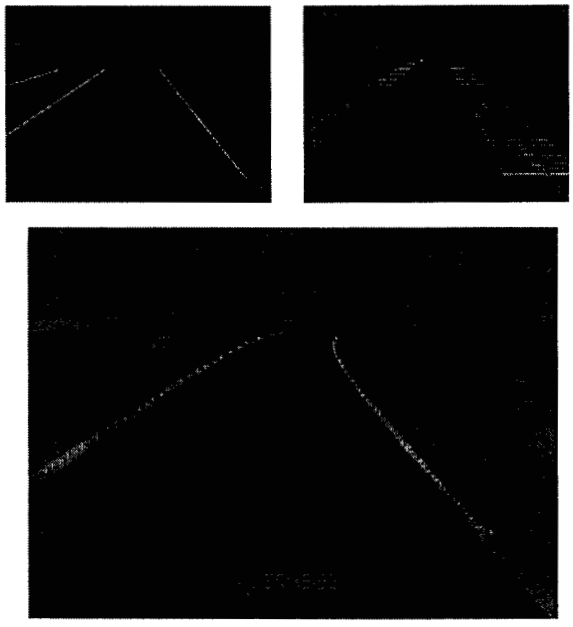


Figure 2.5: The initialization of the algorithm is done by detection of portions of straight lines of common orientation (top left). Within the search zone predicted (top right), a precise localization of line markers points if performed (bottom).

2.6 Determining the camera geometry using residual disparity

So far, we have supposed that the camera geometry with respect to the road is known, and fixed. However, the movements of the car's suspension and the change in vertical road curvature can affect this geometry. Indeed, it

parameter	left	right
lateral offset (m)	0.73	4.57
orientation (degrees)	90.6	90.0
initial radius of curvature $\frac{1}{C_1}$ (m)	2074	3910
final radius of curvature $\frac{1}{C_2}$ (m)	309	297

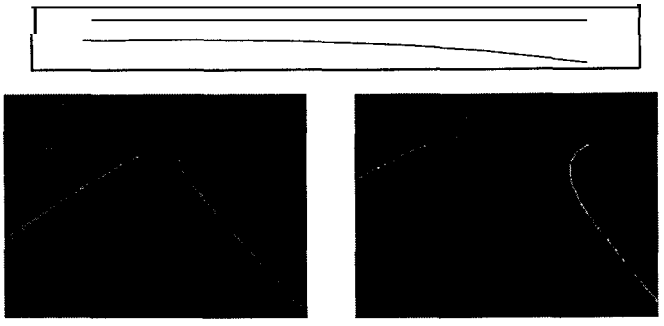


Figure **2.6:** Estimated parameters, top view, and reprojected view of the fitted clothoids. The zoom shows that the fit is quite precise, in spite of the large distance and curvature variation.

has been reported [DD90] that a small difference in the assumed and actual camera tilt angle with respect to the ground affects the **3D** reconstruction significantly. Moreover, the operation of mapping the ground plane disparity to zero is very sensitive to this parameter, as a small error in the inclination angle will cause a significant error on the localization of the ground plane. Therefore, we need a way to update the camera geometry relative to the ground plane. The idea is to use the measurements of the image of the road to compute the external parameters.

The two geometric parameters in our model are the inclination angle $_{cy}$ and camera height, h. The points-of-interest which exhibit small residual disparities are assumed to lie on the ground plane. This residual disparity can be used to update the inclination angle $_{cy}$ and height h. The idea is to minimize with respect to $_{cy}$ and h the sum of squares of differences between these measured disparities and the disparity under the ground plane assumption. The values of α and h are then continuously updated over time using a linear Kalman Filter based on the dynamics of α and h. For example, the height h is modeled as a damped harmonic oscillator driven by noise. This

parameter	all points	occlusion removal
lateral offset (m)	2.14	2.10
orientation (degrees)	-0.67	-0.45
$\frac{1}{C_1}$ (m)	10000	10000
$\frac{1}{C_3}$ (m)	250	10000

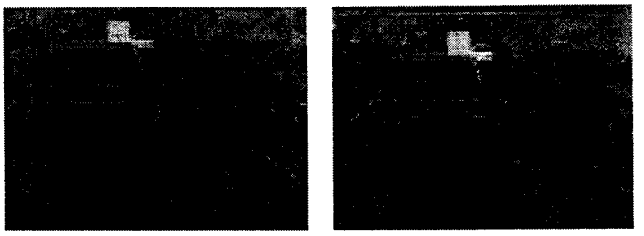


Figure 2.7: Estimated parameters and zoom showing the feature points (green) and the fitted lane markers (red). Left; all the feature points are used. Right: the spurious feature points (purple) are removed using the residual disparity with respect to the ground plane

is a model consistent with the suspension system of the car.

There are essentially two origins for variations in α and h: a short term variation due to camera vibrations, which requires a large process noise, and a long term variation caused by a change in the slope of the road, which can be captured using a small process noise. An example of the results obtained from a sequence of 210 frames recorded during 7 seconds of freeway driving is shown in figure 2.8.

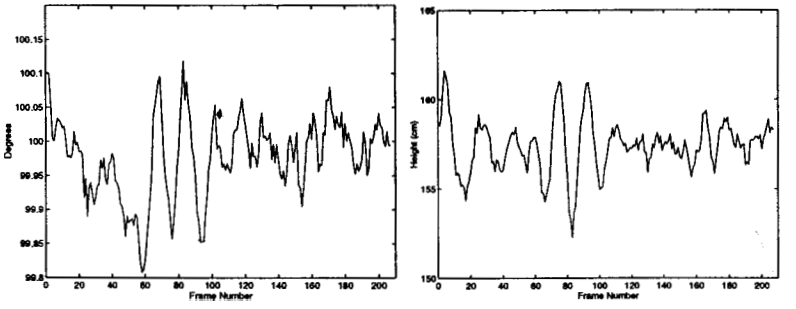


Figure 2.8: Camera inclination angle and camera height estimated from ground plane disparities for a freeway driving sequence.

Chapter 3

A Real-Time Approach to Stereopsis and Lane-Finding

3.1 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 3.3 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 3.4 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.

3.2 Description of imaging setup

Figure 3.1 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

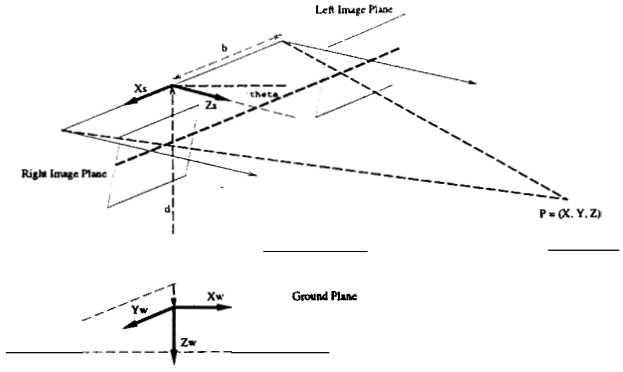


Figure **3.1:** Stereo configuration used in our experiments.

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 **3.1** through **3.4** define the relationship between the coordinates of a point with respect to the stereo rigs frame of reference and the coordinates of its projection in the left and right images.

$$u^l = s_x((X_s + b)/Z_s) + c_x^l \tag{3.1}$$
$$v^l = s_y(Y_s/Z_s) + c_y^l \tag{3.2}$$
$$u^r = s_x((X_s - b)/Z_s) + c_x^r \tag{3.3}$$
$$v^r = s_y(Y_s/Z_s) + c_y^r \tag{3.4}$$
$$\tag{3.5}$$

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

$$\begin{pmatrix} X_s \\ Y_s \\ Z_s \\ 1 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -\sin\theta & 0 & \cos 8 & (d\cos\theta) \\ \cos 8 & 0 & \sin 8 & (d\sin\theta) \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{pmatrix} \tag{3.6}$$

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.

$$(u^l - u^r) = \frac{2bs_x \cos\theta}{ds_y}(v^l - h^l) + (c_x^l - c_x^r) \tag{3.7}$$

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.

3.3 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 **3.1.**

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.

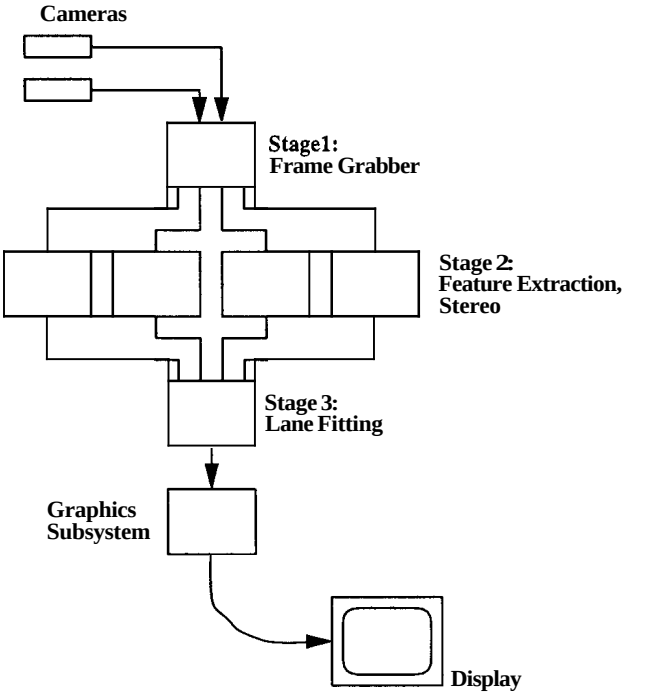


Figure 3.2: Network of TMS320C40s used to implement the real-time lane tracking and stereo systems.

We have chosen to implement our stereo algorithm on a network of TMS320C40 Digital Signal Processors which are arranged in a network as

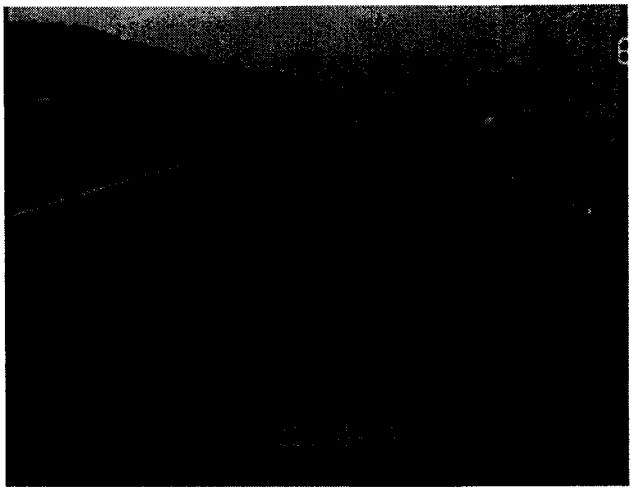
shown in Figure 3.2. The processors are arranged in a pipeline to improve system throughput. The processor at the head of the pipeline is equipped with a frame grabber which it uses to capture pairs of images from the left and right cameras. It then sends the image data on to a set of four processors 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 3.3. 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.

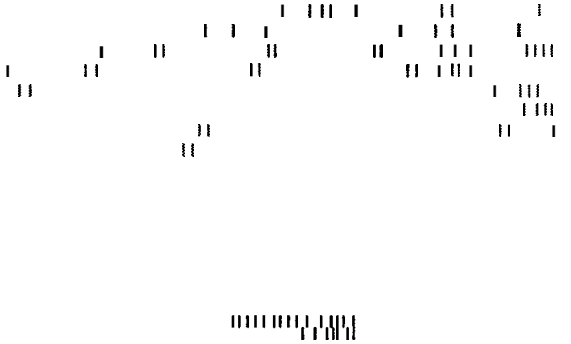
3.4 Lane Recognition

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.

Figure 3.4 shows the relationship between the vehicle and a section of the highway lane some distance ahead while Figure 3.5 shows how that portion



(a)



(b)

Figure **3.3:** (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

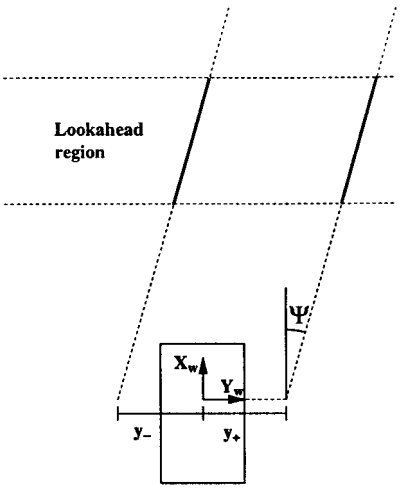


Figure **3.4:** Figure showing the relationship between the vehicle and the highway lane

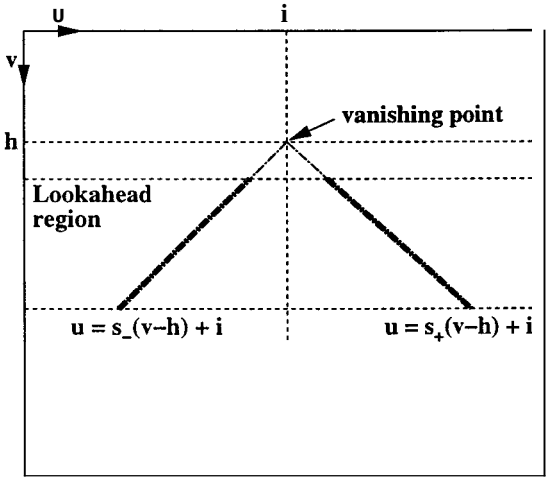


Figure *3.5:* The appearance of the roadway in the image plane

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 **(3.9)** through **(3.11)** describe the well known relationship between these image parameters and the parameters of the road and camera models.

$$h^l = c^l_y - s_y \tan\theta \tag{3.8}$$

$$i^l = c^l_x - s_x \frac{\tan\psi}{\cos\theta} \tag{3.9}$$

$$s^l_+ = \frac{s_x}{s_y}\{\cos\theta(\frac{y_+ + b}{d}) - \sin\theta\tan\psi\} \tag{3.10}$$

$$s^l_+ = \frac{s_x}{s_y}\{\cos\theta(\frac{y_- + b}{d}) - \sin\theta\tan\psi\} \tag{3.11}$$

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.

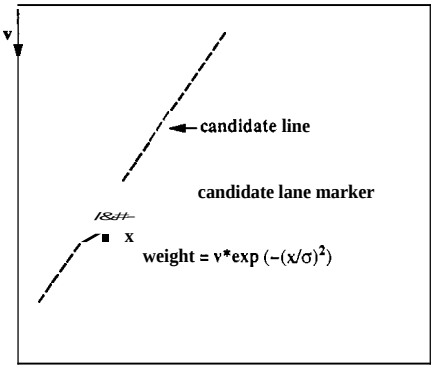


Figure **3.6:** Figure showing how a straight line is ranked by a particular image measurement

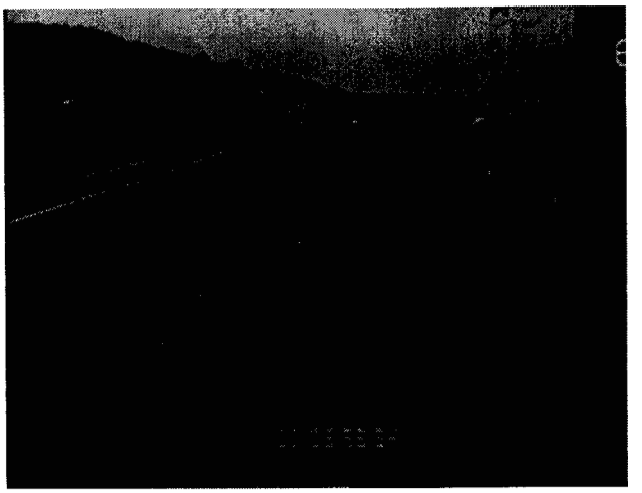
For each candidate line, the contribution of a given image measurement is based on the lateral distance between that measurement and the line and is weighted by the position of the feature in the image as shown in Figure 3.6. Feature points that are lower in the image (and hence closer to the vehicle) 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.



(a)

(b)

Figure *3.7:* (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

Chapter 4

Conclusion

We have proposed an integrated approach for vision based longitudinal and lateral vehicle control. In this approach, the vision module provides the following information to be used for any control system:

- detection of the leading vehicles and measurement of their distance,
- estimation of the flow of lane markers and of road curvature at a distance

The main originality of our approach is the extensive use of binocular stereopsis, and its integration with lane marker detection.

We have presented a new stereo algorithm which exploits domain constraints to achieve efficiency in computing the instantaneous disparity map, as well as a new method to update dynamically a **3-D** obstacle map, which is well suited to the nature of our problem. Our results illustrate the fact that by combining these two algorithms, it is possible to obtain a reliable and relatively dense obstacle map at a small computational cost.

We have then shown that the lane tracking task, which is traditionally carried on monocularly, can benefit significantly from a stereo based-approach. In particular, we are able to deal with crowded traffic scenes where substantial segments of the lane markers may be occluded. The binocular measurement of the lane markers enables us to perform an on-line updating of the external geometric parameters of the stereo rig with respect to the road. We can therefore deal with change in slope of the road, as well as variations in inclination of the car caused by bumps or accelerations and decelerations.

In Chapter 3 we discussed our real-time implementations of stereo-vision and lane-tracking. 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 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 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 under MOU 257. We have implemented lateral control control algorithms based on the output of our vision system which have been tested at speeds of up to 65mph. We plan to continue to improve and refine our control strategies over the coming months.

Bibliography

[APR92] O.D. Altan, H.K. Patnaik, and R.P. Roesser. Computer architecture and implementation of vision-based real-time lane sensing. In *Proc. of the Intelligent Vehicles '92 Symposium*, pages 202–206, 1992.

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

[CE90] S. Carlsson and J.-O. Eklundh. Object detection using model based prediction and motion parallax. In *Proc. First European Conference on Computer Vision*, pages 297–306, Apr 1990.

[CMT91] S. Chandrashekar, A. Meygret, and M. Thonnat. Temporal analysis of stereo image sequences of traffic scenes. In *Proc. Vehicle Navigation and Information Systems Conference*, pages 203–212, 1991.

[Cri90] J. Crisman. *Color Visionfor the Detection of unstructured Roads and Intersection.* PhD thesis, Carnegie Mellon University, 1990.

[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.

[DD90] D. DeMenthon and L.S. Davis. Reconstruction of a road by local image matches and global 3d optimiztion. In *IEEE Int. Conf. on Robotics and Automation*, pages 1337–1342, 1990.

[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.

[Enk90] W. Enkelmann. Obstacle detection by evaluation of optical flow fields from image sequences. In *Proc. First European Conference on Computer Vision,* pages 23–26, April 1990.

[Hel25] H.v. Helmholtz. *Treatise on Physiological Optics* (translated by J.P.C. Southall). Dover, NY, 1925.

[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.

[KYO+96] Takeo Kanade, Atsushi Yoshida, Kazuo Oda, Hiroshi Kano, and Masaya Tanaka. A stereo machine for video-rate dense depth mapping and its new applications. In *Proc. IEEE Conf. on Comp. Vision and Putt. Recog.,* pages 196–202, 1996.

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

[Mat92] Larry Matthies. Stereo vision for planetary rovers: Stochastic modeling to near real-time implementation. *International Journal of Computer Vision,* 8(1):71–91, 1992.

[MBLB91] H.A. Mallot, H.H. Bulthoff, J.J. Little, and S. Bohrer. Inverse perspective mapping simplifies optical flow computation and obstacle detection. *Biological cybernetics,* 64(3):177–185, 1991.

[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.

[OOS+90] M. Ohzora, T. Ozaki, S. Sasaki, M. Yoshida, and Y. Hiratsuka. Video-rate image processing system for an autonomous personal vehicle system. In *IAPR Workshop on Machine Vision Application*, pages 389–392, Nov 1990.

[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.

[Pom92] D.A. Pomerleau. Progress in neural network-based vision for autonomous robot driving. In *Proc. of the Intelligent Vehicles '92 Symposium*, pages 391–396, 1992.

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

[RH91] D. Raviv and M. Herman. A new approach to vision and control for road following. In *Proc. IEEE Conf. on Comp. Vision and Patt. Recog.*, pages 217–225, June 1991.

[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.

[Tho90] C. Thorpe. *Vision and Navigation: The Carnegie-Mellon Navlab.* Kluwer Academic Publishers, Norwell, Mass, 1990.

[TMGM88] M.A. Turk, D.G. Morgenthaler, K.D. Gremban, and M. Marra. Vits — a vision system for autonomous land vehicle navigation. *IEEE Trans. Pattern Anal. Machine Intell.*, 10:342–361, 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.

[ZJB+90] Y. Zheng, D.G. Jones, S.A. Billings, J.E.W. Mayhew, and J.P. Frisby. Switcher: a stereo algorithm for ground plane obstacle detection. *Image and Vision Computing*, 8:57–62, 1990.