Smart Cars and Smart Roads

J. Malik, J.Weber, Q.-T. Luong and D. Koller[1]
Computer Science, University of California at Berkeley, CA 94720.
[1]ECRC, Arabellastrasse 17, D-91925 Munich, Germany

Abstract

This paper describes two projects applying computer vision to Intelligent Vehicle Highway Systems. The first project has resulted in the development of a system for monitoring traffic scenes using video information. The objective is to estimate traffic parameters such as flow rates, speeds and link travel times, as well as to detect quickly disruptive incidents such as stalled vehicles and accidents. The second project is aimed at developing vision as a sensor technology for vehicle control. The novel feature of this project, compared to most previous approaches, is the extensive use of binocular stereopsis. First, it provides information for obstacle detection, grouping, and range estimation which is directly used for longitudinal control. Secondly, the obstacle–ground separation enables robust localization of partially occluded lane boundaries as well as the dynamic update of camera rig parameters to deal with vibrations and vertical road curvature.

1 Introduction

Traffic congestion is a serious problem in industrialized societies. In the state of California in the US, congestion is projected to triple by 2005 with expected peak hour freeway speeds dropping to 11 mph. Congestion and safety are related: accident rates increase under congestion and half of all congestion is caused by accidents and other incidents.

Obviously, this problem has to be attacked on multiple fronts. This paper concentrates on two areas where computer vision technology can help significantly.

In the short run, traffic surveillance using video cameras could be substantially automated. Some of the major uses of such a data collection system would be

- Fast incident detection without human monitoring of multiple video signals.
- Estimation of travel times between various points. This could be used in conjunction with variable message signs for flow control.
- Detailed traffic condition information for public use.

Unlike conventional loop detectors, which are buried underneath highways to count vehicles, video monitoring systems are less disruptive and less costly to install. They also have greater range and allow for more detailed descriptions

BMVC 1995 doi:10.5244/C.9.37

of traffic situations. In this paper, we describe a prototype robust, vision-based traffic surveillance system [16, 17].

In the long run, platooning has considerable promise as a way of increasing freeway capacity without building new lanes. A number of different system architectures are being studied, but common to all of them is the idea that vehicle control would be at least partially automated to permit vehicles to be driven safely with small inter-vehicle distances. That necessitates sensing of the car's position with respect to the lane markers and other vehicles. While a variety of different sensing modalities including laser range finders, radar or magnetic sensors could be used for this purpose, computer vision offers some advantages. It has a large field of view, is passive, and is rather well adapted to the current traffic infrastructure which was designed from the point of human drivers using vision.

Prominent work in this area is due to Dickmanns and collaborators[5] in Germany and Pomerleau and collaborators[22] at CMU. Our approach [18] at UC Berkeley, is distinctive in making considerable use of binocular stereopsis.

2 Traffic surveillance using video information

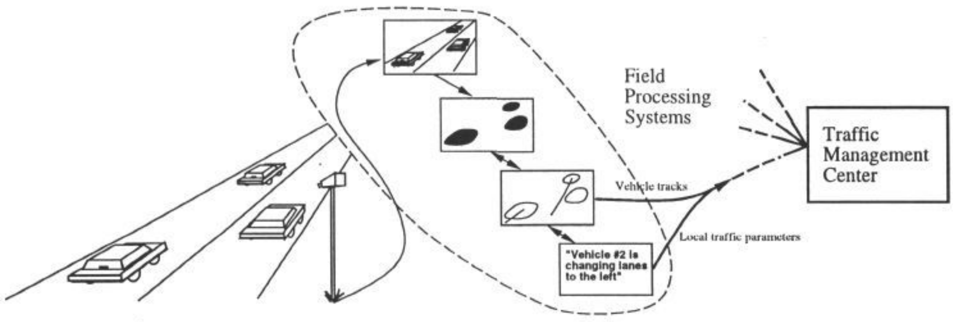


Figure 1: Our system concept for video traffic surveillance.

Our system concept is illustrated in Figure 1. The core idea is to have video cameras mounted on poles or other tall structures looking down at the traffic scene. Video is captured, digitized, and processed by onsite computers, and then transmitted in summary form to a Traffic Management Center (TMC) for collation and computation of multi-site statistics such as link travel times. Processing occurs in four stages:

1. Optical-flow segmentation (very slow traffic) or background image differencing (normal traffic conditions) to detect and group blobs corresponding to individual vehicles as they come into the field of view. At this stage, a bounding box can be fitted to the blob to estimate shape parameters for vehicle classification.

2. Tracking each individual vehicle to refine and update its position and velocity in 3D world coordinates, as well as the shape parameters, until it leaves the

tracking zone.

3. Reasoning from the track data in order to infer local traffic flow parameters including vehicle counts per lane, average speeds, incidents, lane change frequencies, etc. These parameters, together with track information (timestamp, vehicle type, color, shape, X–Y position), are communicated to the TMC at regular intervals.

4. At the TMC, local traffic parameters from each site are collated and displayed as desired, and/or used in controlling signals, message displays, and other traffic control devices. In the case of incidents, the TMC can also request transmission of stored images from the incident site. Computers at the TMC also process the track information from neighboring camera sites to compute long-distance parameters such as link times and origin–destination counts.

Commercial video-based traffic surveillance systems have aimed for simple techniques for measuring traffic flow that could be implemented in real time. For instance, the first generation of systems such as the Autoscope[21] attempted to count numbers of vehicles crossing a row of pixels by looking for the associated temporal change in brightness values. Such systems can be easily fooled e.g. by shadows linking cars across lanes. The next generation of commercial systems are based on tracking ideas. Model-based vehicle tracking systems have previously been investigated by several research groups, the most prominent being the groups at Karlsruhe [14, 15] and at the University of Reading[1, 24]. The emphasis is on recovering trajectories and models with high accuracy for a small number of vehicles.

The challenges of designing video traffic surveillance systems are that of identifying vehicles despite imprecise video data and changing lighting conditions, tracking individual vehicles despite their overlapping with one another, and efficiently providing high-level descriptions based on evidence accumulated over time. The major tradeoff in the design of such a system, indeed of most realtime computer vision or image processing systems, is the tradeoff between robustness/accuracy and speed. More sophisticated, robust algorithms require more computation; however they will continue to operate in conditions such as shadows, dense traffic, and day-to-night transitions when the naive algorithms will fail.

After three years of this research, we have largely succeeded in this goal [9, 16, 17]. We have a prototype system that operates at 15 frames per second in detecting and tracking vehicles in a wide variety of traffic, weather, and lighting conditions. From the track information, the system computes individual and average vehicle speeds, lane flow rates, lane change counts. Other parameters such as headways and queue lengths are easily obtained from track information. The system also currently detects stalled vehicles (correctly distinguishing them from vehicles stopped in traffic). Some results are shown in Figures 2 and 3.

3 Surveillance system modules

Our traffic surveillance system is based on the block diagram shown in Figure 4.

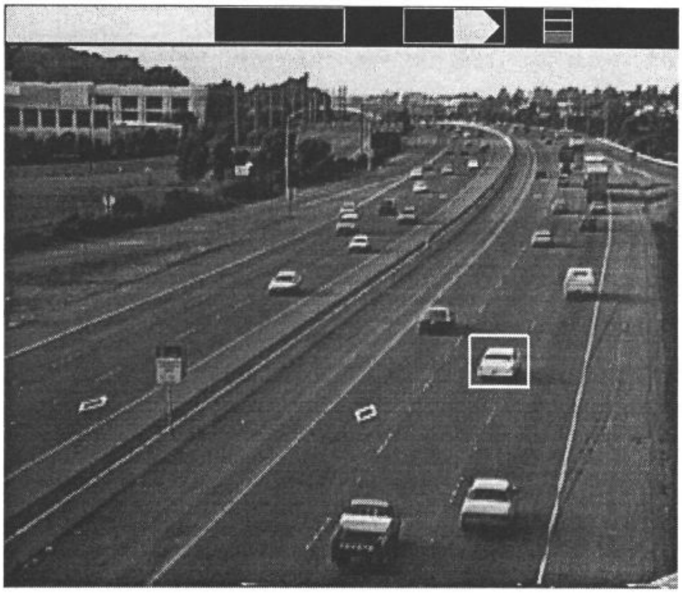


Figure 2: Highway scene (I-580 in Richmond, CA) showing tracking information for the car outlined. The bar at the top shows the vehicle's speed. The arrow at top is the lane change indicator. The indicator at top right is green to show that the vehicle is operating normally (i.e., not stalled).

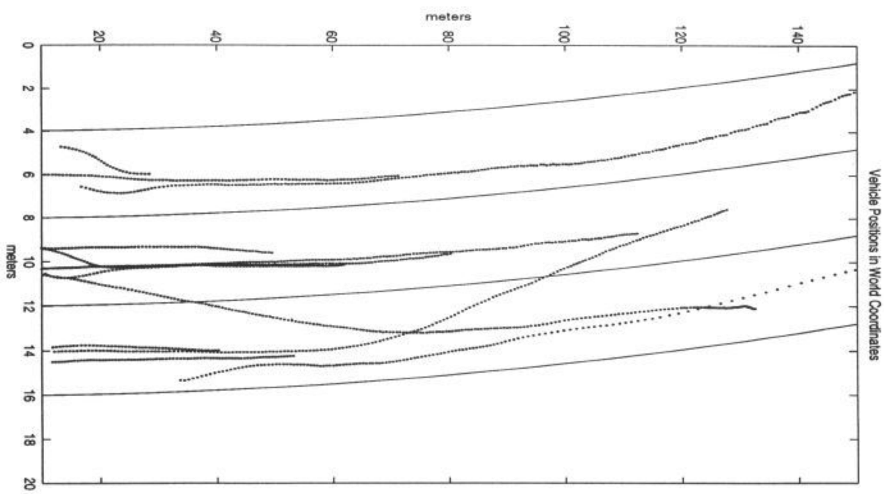


Figure 3: Computed bird's-eye view of road segment shown in Fig. 2 with tracks overlaid.

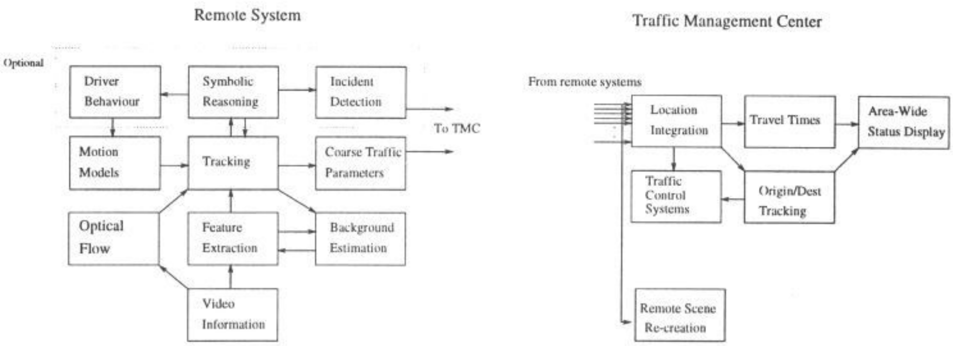


Figure 4: Block diagram of the complete traffic surveillance system.

The analysis generally proceeds from low-level processing of road traffic images to high-level descriptions of the traffic situation. The following three phases can be identified:

1. Image processing to detect vehicles as they enter the scene, and to estimate shape parameters.
2. Tracking each individual vehicle to estimate its position and velocity. Reasoning from the track data in order to infer local traffic parameters.
3. At the TMC: Collation of local parameters and computation of long-distance parameters from multi-camera track data.

3.1 Vehicle detection, track and shape initialization

A surveillance system initiates vehicle identification and tracking by determining what parts of each image belong to moving objects and what parts belong to the background. In normal traffic, this is accomplished by examining the difference in pixel intensities between each new frame and an estimate of the stationary background. Reliable background estimation, which is critical for accurate identification of moving "blobs", is made more difficult as lighting conditions change. We perform this *initialization* step by using a Kalman filter-based adaptive background model[10, 12]. This allows the background estimate to evolve as the weather and time of day affect lighting conditions.

The regions of the images identified as not being part of the background and not currently being tracked are used to initialize new tracks. To form a new track, the motion of the vehicle must be initialized. The motion of the region between two frames forms the initial velocity estimate. Shape parameters for the vehicle are also initialized at this point. This is performed by using the bounding outline of the moving region to form an initial estimate of vehicle shape.

Shadows (especially long shadows at dawn and dusk) are a significant problem for vision-based systems. We have derived a new method for shadow removal. The core idea is that the boundary of a car is defined both as a brightness boundary and as a texture boundary. While the brightness boundary separating a car from its shadow may be difficult to detect, the texture boundary serves as a robust separator and can be calculated efficiently.

The image-differencing approach has been shown to work extremely well with normal traffic flows. With very slow-moving and stop-and-go traffic, however, the background model will begin to average in the cars themselves as a significant component of the intensity. Fortunately, these conditions are exactly those suited for optical flow calculations, because objects do not move very many pixels between frames. The track initialization and shape information from either image-differencing or optical-flow methods can be used equally well in the tracking phase.

After identifying moving blobs, the vision system attempts to disambiguate individual vehicles and estimate their shapes. This helps with associating data over a sequence of images and with obtaining accurate vehicle trajectories. Our system performs these tasks by developing a correlation mask over time. This mask conforms to the estimated appearance of the vehicle in the image.

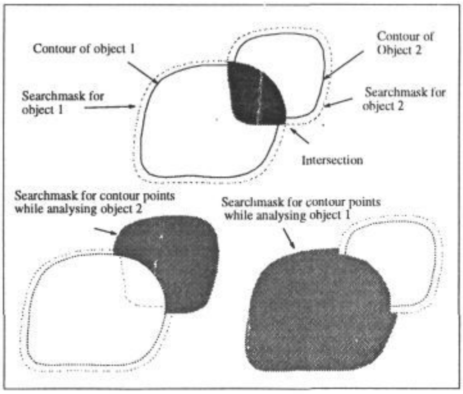
3.2 Robust multiple vehicle tracking

Two primary factors that complicate this task are noisy measurements and vehicle occlusions, which make it more difficult to identify and disambiguate vehicles.

To address the problem of noisy measurements, we employ the Kalman filter[7] formalism to provide most likely estimates of the state of a vehicle, $\vec{X}_t = (x, y, \dot{x}, \dot{y})$ based on accumulated observations. The tracking is performed in a *world* coordinate system. This is accomplished by projecting the points on the image onto the road plane. Since the road can be assumed flat for the range of the image, this transformation only requires a simple linear transformation in homogeneous coordinates. The advantage of tracking in world coordinates is that physical constraints of a vehicles motion model can be used to guide tracking. For example, the knowledge that vehicles have finite acceleration will limit the range of motion a vehicle can have in the image from frame to frame.

At each time frame we measure the position of the center of the vehicle in the image. This position is translated into world coordinates and used as our state measurement, $\vec{Z}_t$. The measurement noise is found by taking the known measurement variance in image coordinates and transforming it into world coordinates. In this way we can use the fact that as a vehicle becomes more distant, its apparent size becomes smaller and the uncertainty in its position increases. This fact is often not used in systems which track purely in image coordinates.

Because vehicles often overlap with each other in the road images, the extracted contours of vehicles will become distorted for some frames. This can cause artificial shifts in vehicle trajectories, since tracks are obtained by connecting centers of contours along the image sequence. To avoid these artificial shifts and to obtain reasonable tracks, we employ an explicit occlusion reasoning algorithm, which compensates for overlapping vehicles. The basic idea is to exploit the known traffic scene geometry and the fact that motion is assumed to be constrained to the ground plane [16]. This knowledge makes it possible to determine a depth ordering among the objects in the scene, and this depth ordering defines the order in which objects are able to occlude each other(see Fig. 5).



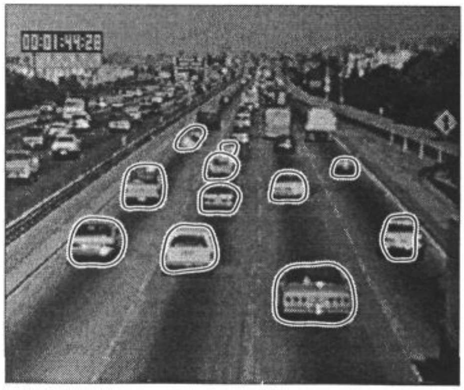


Figure 5: Vehicle occlusion can be predicted from the depth ordering of vehicles. Information in the occluded region must be extrapolated for the more distant vehicle. The figure on the right contains a partially occluded vehicle.

The tracking process provides the instantaneous positions and velocities of all tracked vehicles. This information can be used to computer the local traffic parameters, such as flow rate, average vehicle speed, lane changes, queue length, average spatial headway etc. This summary information can then be communicated, say every 1 second, to the Traffic Management Center.

3.3 Information processing at the TMC

We are currently implementing vehicle handoff between consecutive cameras, using track-matching, so that a single vehicle can be tracked over a long distance. With this information, link travel times and origin–destination counts can be computed for roads covered by overlapping or nearly-overlapping fields of view. We have also demonstrated the computation of bounding boxes to give basic size and shape information for individual vehicles. Shape and color information can then be used to assist in long-distance tracking of vehicles through non-contiguous surveillance regions (which we expect to be the norm except in critical regions of the freeway network). The most promising technique here seems to be matching of within-lane vehicle classification sequences.

For incident detection, we have demonstrated the use of probabilistic reasoning techniques based on the formalism of dynamic belief networks. Details may be found in [9, 17].

We'll now move on to describe the second project.

4 Stereo-based approach to vehicle control

We have developed a system[18] for vision based longitudinal and lateral vehicle control which makes extensive use of binocular stereopsis. Previous work on autonomous vehicle guidance by Dickmanns's group[5] and Pomerleau[22] has concentrated mostly on road following. In crowded traffic scenes, the presence of other vehicles causes two problems. First, they are potential obstacles, which are to be detected. This problem has been addressed using optical flow interpretation [6], stereopsis [26, 20], or a combination of both [3]. These approaches are often computationally expensive. Dickmanns and collaborators have used approaches based on finding symmetric objects which are computationally less expensive, but are not likely to be as general or robust. In addition to being potential obstacles, vehicles can also occlude significant fragments of lane markers, causing problems for algorithms that do not explicitly take occlusion into account.

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. 5.

At each new instant, we first compute the stereo disparity using an efficient algorithm based on the Helmholtz shear. 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. 6). 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 by a specialized feature detector. From the image positions of the lane markers, we can update the geometric parameters of the stereo rig. 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. 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. 6.

For more details on our approach, please see [18].

5 The geometrical model

5.1 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 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

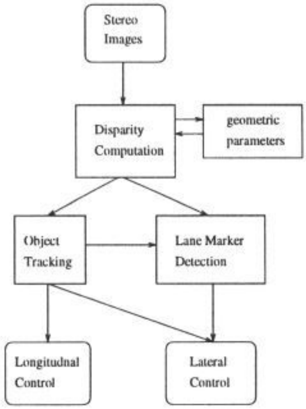


Figure 6: Block diagram of the stereo-based vehicle control system. Arrows indicate flow of information.

the case of two cameras, we see 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:

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

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

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

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 $\mathbf{n}$ is the unit normal vector of the plane and d the distance of the plane to the origin.

5.2 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 [8] 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:

$$\begin{cases} u' = u + h_{12}v + h_{13} \\ v' = v \end{cases} \tag{2}$$

From this expression, and comparing with Eqn 1, one can show 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, and the component of the normal $\mathbf{n}$ along the X-axis is zero.

In such a situation, the stereo rig is entirely characterized by the intrinsic parameters of the first camera $\mathbf{A}$, and the 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:

$$\begin{cases} h_{12} = b/d \sin\alpha \\ h_{13} = b/d \cos\alpha \end{cases} \tag{3}$$

Ideas related to the Helmholtz shear have been used previously for obstacle detection by mobile robots [19, 26].

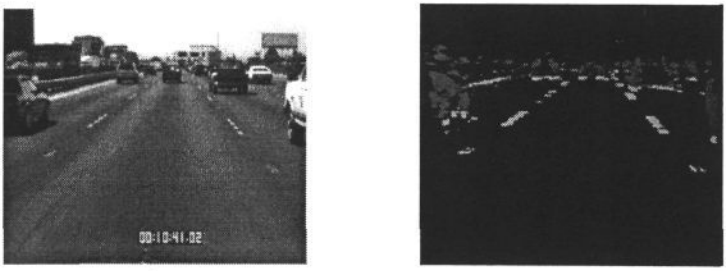
6 Dynamic Stereopsis

The oft-quoted criticism of stereopsis for use in vehicle navigation is that it is computationally very 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. 5. After applying this 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. 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 7 shows the result on a single frame.

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 temporal integration of information 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 are interested in connected, rigid objects. This allows 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



a) b)

Figure 7: 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 indicates regions where the disparity could not be accurately recovered.

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 9 gives an example of two objects which are connected in this image/depth $3D$ space.

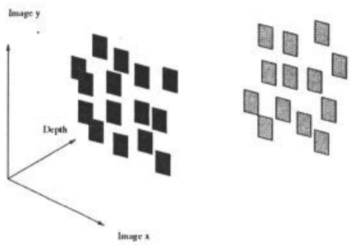


Figure 8: Connected components in image/depth space consist of those pixels which are nearest neighbors in image coordinates as well as having depth differences less than depth uncertainty.

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 9 shows the objects found by this method. Note the tendency of the scheme to oversegment. This can happen when the connecting region between two parts of an object lacks features of sufficient correlation strength.

7 Updating lateral position and stereo rig parameters

For lateral control, the most crucial variables to be sensed are the position and orientation of the vehicle relative to the lane markers. In addition to these parameters, sensing of road curvature is very useful as it facilitates a smoother trajectory.

A number of different approaches have been used for this problem[4, 11, 13, 5]. Our approach follows the spirit of [5] in using a parametrized curve model of the

Figure 9: Objects identified as being in the same lanes of traffic as the test vehicle. On the right side of the image is a bird's-eye-view from above the road surface showing the relative position of the tracked objects with respect to the test vehicle.

lane markers which is estimated and then dynamically updated. Details may be found in [18]. The most significant difference is that we can use the results of the obstacle detection stage to exclude features that might lie in the search window (based on previous lane marker position), but belong to vehicles instead of to lane markers. This prevents us from getting spurious fits in the presence of occlusion of lane markers. We expect this to be particularly useful in congested scenes.

We also need to dynamically update the camera rig geometry with respect to the road, characterized by the two parameters: inclination angle α and camera height, h. These change due to car vibrations and change in vertical road curvature. It is crucial to estimate these accurately, since it is known [4] 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.

To update the camera geometry relative to the ground plane, we use the following simple heuristic: The points-of-interest which exhibit small residual disparities are assumed to lie on the ground plane. We attribute the residual disparities not to a global movement of the ground plane but instead to error in our estimate of inclination angle α and height h. The idea then is to minimize with respect to α 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 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 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 10.

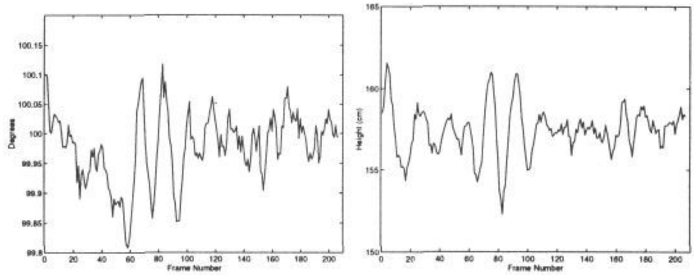


Figure 10: Camera inclination angle and camera height estimated from ground plane disparities for a freeway driving sequence.

8 Conclusion

This paper describes two projects applying computer vision to intelligent vehicle highway systems. The first project has resulted in the development of a realtime system for monitoring traffic scenes using video information. The second project is developing a system for vision based longitudinal and lateral vehicle control. The vision module provides the following information to be used by the vehicle control system:

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

The originality of our approach is in the extensive use of binocular stereopsis for this purpose.

Acknowledgements

This research was supported by California Department of Transportation through the PATH program. The incident detection module of the video surveillance system based on dynamic Bayes nets was developed by T. Huang, S. Russell, and B. Rao.

References

[1] K.D. Baker and G.D. Sullivan. Performance assessment of model-based tracking. In *Proc. of the IEEE Workshop on Applications of Computer Vision*, pages 28-35, Palm Springs, CA, 1992.

[2] Yaakov Bar-Shalom and Thomas E. Fortmann. *Tracking and Data Association*. Academic Press, New York, 1988.

[3] 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.

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

[5] E.D. Dickmanns and B.D. Mysliwetz. Recursive 3-d road and relative ego-state recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14:199–213, 1992.

[6] W. Enkelmann. Obstacle detection by evaluation of optical flow fields from image sequences. In *Proc. First European Conference on Computer Vision*, pages 134–138. Antibes, France, Apr. 23-26, 1990.

[7] Arthur Gelb, editor. *Applied Optimal Estimation.* The MIT Press, Cambridge, MA., 1974.

[8] H.v. Helmholtz. *Treatise on Physiological Optics* (translated by J.P.C. Southall), volume 1–3. Dover, NY, 1925.

[9] T. Huang, D. Koller, J. Malik, G. Ogasawara, B. Rao, S. Russell, and J.Weber. Automatic symbolic traffic scene analysis using belief networks. In *Proceedings of the 12th National Conference on Artificial Intelligence*, pages 966–972, Seattle, WA, July 31–Aug. 4, 1994.

[10] Klaus-Peter Karmann and Achim von Brandt. Moving object recognition using an adaptive background memory. In V Cappellini, editor, *Time-Varying Image Processing and Moving Object Recognition, 2.* Elsevier, Amsterdam, The Netherlands, 1990.

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

[12] Michael Kilger. A shadow handler in a video-based real-time traffic monitoring system. In *IEEE Workshop on Applications of Computer Vision*, pages 1060–1066, Palm Springs, CA, 1992.

[13] K. Kluge and C. Thorpe. Representation and recovery of road geometry in yarf. In *Proc. Intelligent vehicles symposium*, pages 114–119, 1992.

[14] D. Koller, K. Daniilidis, and H.-H. Nagel. Model-based Object Tracking in Monocular Image Sequences of Road Traffic Scenes. *International Journal of Computer Vision*, 10: 257-281,1993.

[15] D. Koller, N. Heinze, and H.-H. Nagel. Algorithmic characterization of vehicle trajectories from image sequences by motion verbs. In *IEEE Conf. on Computer Vision and Pattern Recognition*, pages 90–95, Lahaina, Maui, Hawaii, June 3-6, 1991.

[16] D. Koller, J. Weber, and J. Malik. Robust multiple car tracking with occlusion reasoning. In *Proc. Third European Conference on Computer Vision*, pages 189–196, Stockholm, Sweden, May 2-6, 1994, J.-O. Eklundh (ed.),Lecture Notes in Computer Science **800-801**, Springer-Verlag, Berlin, Heidelberg, New York, 1994.

[17] D. Koller, J. Weber, T. Huang, J. Malik, G. Ogasawara, B. Rao, and S. Russell. Towards robust automatic traffic scene analysis in real-time. In *Proceedings of the International Conference on Pattern Recognition*, Israel, November 1994.

[18] Q.T. Luong, J. Weber, D. Koller, and J. Malik. An integrated stereo-based approach to automatic vehicle guidance. In *Proceedings of the Fifth International Conference on Computer Vision*, Cambridge, MA, June 1994, pp. 52-57.

[19] 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.

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

[21] P. Michalopoulos. Vehicle detection video through image processing: the Autoscope system. *IEEE Trans. on Vehicular Technology*, 40:21-29, 1991.

[22] 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.

[23] D. Raviv and M. Herman. A new approach to vision and control for road following. In *Conference on Computer Vision and Pattern Recognition*, pages 217–225, Lahaina, Maui, Hawaii, June 3-6, 1991.

[24] G.D.Sullivan. Visual interpretation of known objects in constrained scenes. In *Phil. Trans. Roy. Soc (B)*, 337: 361-370, 1992.

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

[26] 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.