Proc. NATO Defense Research Group Seminar on Robotics in the Battlefield, Paris, France, March 1991.

VISUAL NAVIGATION USING OPTICAL FLOW

Martin Herman and Tsai-Hong Hong

Sensory Intelligence Group
Robot Systems Division
National Institute of Standards and Technology
Gaithersburg, MD 20899

ABSTRACT

Optical flow is the flow of intensity information across a camera image plane as the camera moves through the environment. This paper describes how optical flow from a camera mounted on a vehicle can be used to perform real-time navigation, during both teleoperated low data rate driving and autonomous driving. In particular, we describe (1) how to extract optical flow from complex outdoor imagery, (2) how the flow values can be used to extract range, (3) how to use this information for road following, obstacle avoidance, and other basic navigation behaviors, and (4) how to use obstacle and surface roughness information as part of a system for teleoperated low data rate driving of a vehicle.

1. Introduction

Visual navigation involves using visual cues to achieve behaviors such as following a road or curvilinear feature, traversing open field terrain, obstacle avoidance, local path planning, approaching an object without collision, pursuing a moving object, etc.

Optical flow is the flow of intensity information across a camera image plane as the camera moves through the environment. The optical flow in a camera mounted on a moving vehicle represents the relative motion of the surrounding scene past the camera. Objects that are close to the camera will appear to flow faster than objects that are distant. This idea is the basis for using flow to obtain range. If we can extract the magnitude of flow at every point in the image, then we can convert this value to a range value.

In this paper, we describe how optical flow in a camera mounted on a vehicle can be used to perform real-time navigation, during both teleoperated low data rate driving and autonomous driving.

The differences between our approach and previous approaches to visual navigation are the following:

1. Most previous approaches have involved *static* image analysis of camera imagery to obtain information for navigation. That is, each image of the scene is processed as if it were a static image. This is one reason why these approaches are slow. However, a fundamental property of mobility is that the vehicle, and camera, is in motion. Our approach therefore involves performing *dynamic* image analysis to obtain information about the environment.
2. Many previous approaches have used *active* sensors to obtain 3-D environmental information. Our approach involves *passive* sensing and dense range extraction using optical flow. An advantage of this approach for battlefield scenarios is that the vehicle does not emit radiation during sensing, thus minimizing detection by the enemy.
3. Most existing road following algorithms convert the information extracted from images into a 3-D, vehicle-centered cartesian coordinate system aligned with the ground plane. Mobility and steering decisions (such as whether to turn left or right, how much to turn, whether to speed up or slow down, and by how much) are then determined in this coordinate system. A 3-D reconstruction is therefore performed before mobility decisions are made. In our approach, many of the algorithms will involve generating mobility and steering commands directly using image information represented in the 2-D image coordinate system. For example, steering commands can be generated directly from the optical flow readings at various points in the image. There are two main advantages to such approaches. The first is that since 3-D reconstruction results in inferred quantities (i.e., recovering the third dimension), control that is based on this method is not as robust as control based on directly observable 2-D image quantities [1]. Second, control based on 2-D image quantities is simpler and requires much less computation; it is therefore much faster.
4. Previous approaches have used cameras that are fixed to the vehicle and which have limited fields of view. In such configurations, only objects and terrain directly ahead of the camera are visible. In unstructured environments, other directions need to be seen so that obstacles, terrain features, and landmarks that are not directly ahead of the camera can be identified. In our approach, this problem is avoided by using a very wide field of view camera with foveal-peripheral image resolution. The gaze of the camera must then be intelligently controlled so that the high resolution fovea is servoed to points of interest.

2. Real-Time Optical Flow Extraction

Optical flow is the flow of intensity information across the image plane due to relative motion between the camera and objects in the environment. It may be represented in the form of an image, where each pixel has associated with it an instantaneous velocity vector representing image motion at that point. In practice, optical flow is extracted by processing a time sequence of at least two images. The goal of our work in this area has been to extract highly accurate optical flow in real time. This can be done if the flow direction at every point is known ahead of time.

2.1. Predicting the Flow Field

Our method of flow extraction assumes that (a) the camera is moving in a stationary world and (b) the camera motion is known. These assumptions lead to two conclusions. First, the optical flow field in the image (i.e., the flow direction at every point) can be predicted. Second, once optical flow has been extracted, the flow vectors can easily be converted to range values. To see why, we will consider three types of camera motion -- pure translation, pure rotation, and a combination of translation and rotation (also see [3]).

Consider a camera undergoing pure translation. Figure 1 is an illustration from Gibson [2] showing the optical flow induced by this motion in a stationary environment. The arrows represent angular velocities of flow vectors formed by spherical projection of the environment onto a sphere centered at the camera focal point. Note that the flow vectors are zero directly ahead of and behind the moving object. The point directly ahead is called the *focus of expansion* (FOE), and all points appear to flow outward from this point. All flow vectors lie along great circles on the sphere as shown in the figure. If a camera with a planar imaging element were located at the center of the sphere, then the great circles, when centrally projected into the camera, would appear as straight lines in the image plane. Furthermore, if the FOE were also projected into the camera, then all motion would appear to flow radially outward in straight lines from this point. Therefore, if the camera motion is known, the flow direction at each image point can be predicted.

To see how the flow vectors can be converted into range values, consider the spherical coordinate system shown in Figure 2. A point P in the scene is projected onto the sphere by intersecting the ray from the camera center to P with the surface of the sphere. In Figure 2, suppose that the positive y-axis is defined by the camera velocity vector. Then the set of great circles that intersect the y-axis represent optical flow lines. Consider a point of interest P' on the sphere. Let A be the angle from the positive y-axis to the ray from the camera center to P' and let B be the angle from the positive z-axis to the plane of the great circle containing P'. If v is the camera velocity and dA/dt is the value of the optical flow on the sphere at angle A, then the following formula is used to calculate the range r to the point in the scene corresponding to the point at A [3]:

$$r = \frac{v \sin A}{dA/dt}. \tag{1}$$

For a camera undergoing pure rotation, all optical flow on the sphere will be along small circles perpendicular to the axis of rotation. For known camera motion, not only the flow direction but the flow magnitude at each image point can be predicted. However, camera rotation provides no information for calculating range.

For a camera undergoing both rotation and translation, the optical flow at each image point is given by the vector sum of the rotational and translational components of flow. Figure 3 shows the situation where the translation vector is perpendicular to the rotational axis. At point P' on the sphere, the direction and magnitude of optical flow due to the known rotation can be predicted along the small circle as shown. The direction of flow due to the known translation can be predicted along the great circle as shown. Since the magnitude of the flow vector due to translation cannot be predicted from the camera motion, the resultant flow vector at point P' lies within a family of vectors as shown in the figure. One technique that can be used to extract and analyze the optical flow is to subtract out the flow due to rotation. This can be accomplished by warping each image so as to eliminate all grey scale displacements due to rotation. This is easily accomplished for known camera rotation. The result is a temporal sequence of images with no rotational flow components. For these images, the flow direction can be predicted, extracted, and converted to range values as described above for pure translational motion.

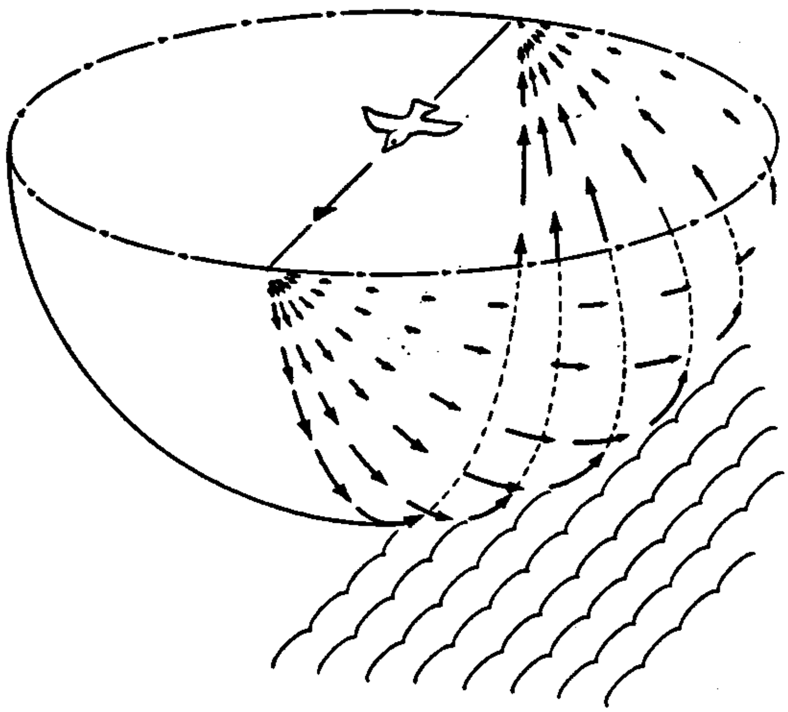


Figure 1. Optical flow induced by camera motion. (Reprinted from Gibson [2], figure 9.3)

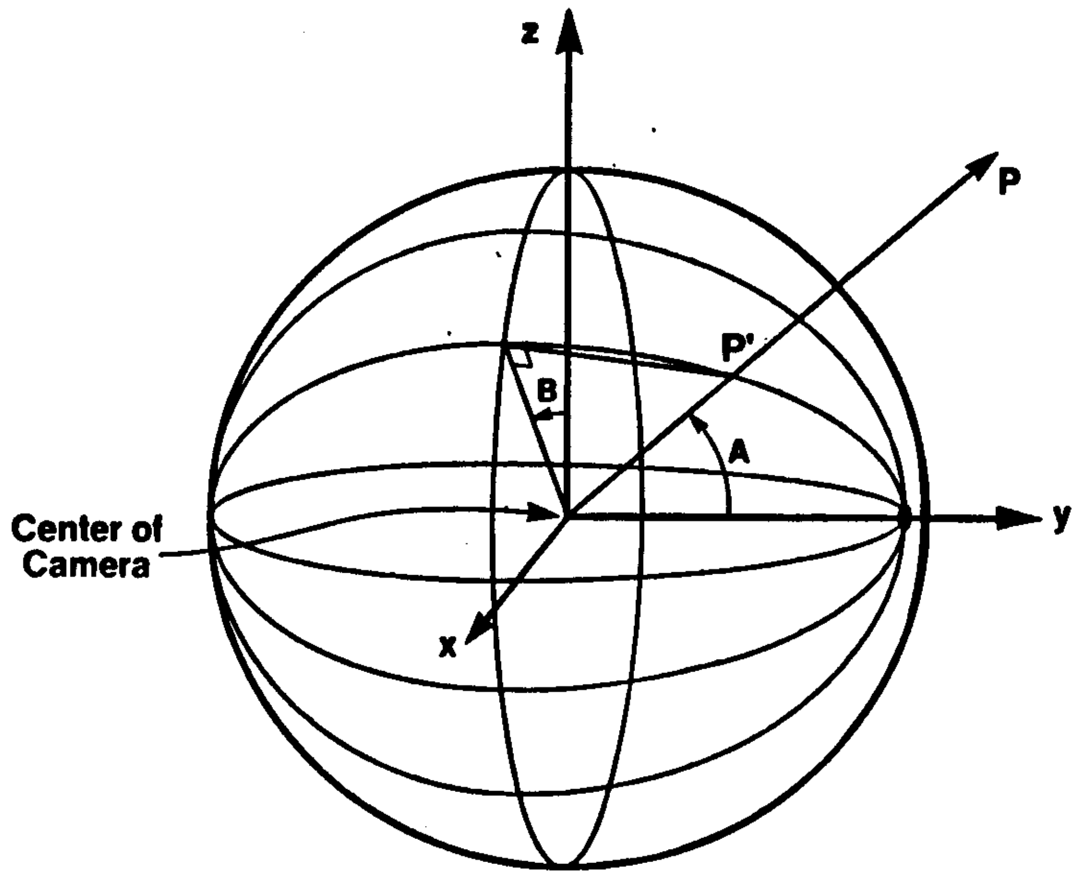


Figure 2. Spherical coordinate system for optical flow.

OPTIC FLOW DUE TO COMBINED ROTATION AND TRANSLATION

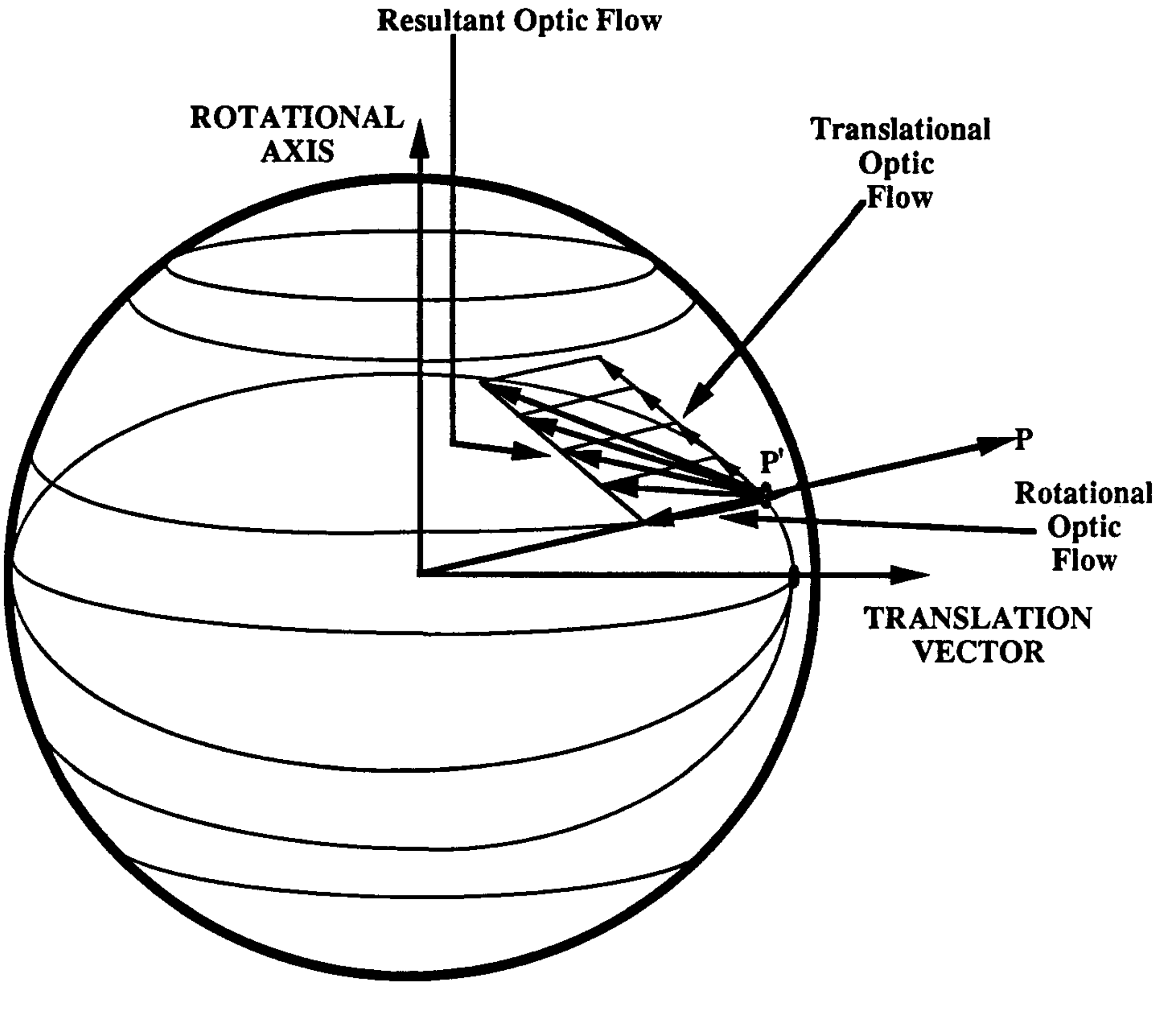


Figure 3.

2.2. Accuracy and Real-Time Performance

The methods described above show how the flow field can be predicted. In this section, we describe how knowledge of this flow field can lead to extraction of highly accurate optical flow and to real-time performance.

When extracting the flow vectors, knowledge of the vector directions eliminates the aperture problem [4]. The aperture problem states that if image motion is detected using a spatially local operator, then only the component of motion parallel to the local brightness gradient (i.e., perpendicular to the local edge) can be computed. This component of motion is called normal flow. Typically, the problem of converting the normal flow vectors to true flow vectors is handled with a costly constraint satisfaction algorithm that combines the normal flow components over an extended region of the image. Because our approach assumes that the true flow vector directions are already known, only the magnitudes of these flow vectors need to be computed. This leads to real-time performance. Another factor leading to real-time performance is that by eliminating the aperture problem, local image operators can be used to extract the flow, and these local operators can run in parallel at all points in the image. Our approach should also lead to greater accuracy of extracted flow vectors since we already know the vector direction accurately.

2.3. Flow Extraction Techniques

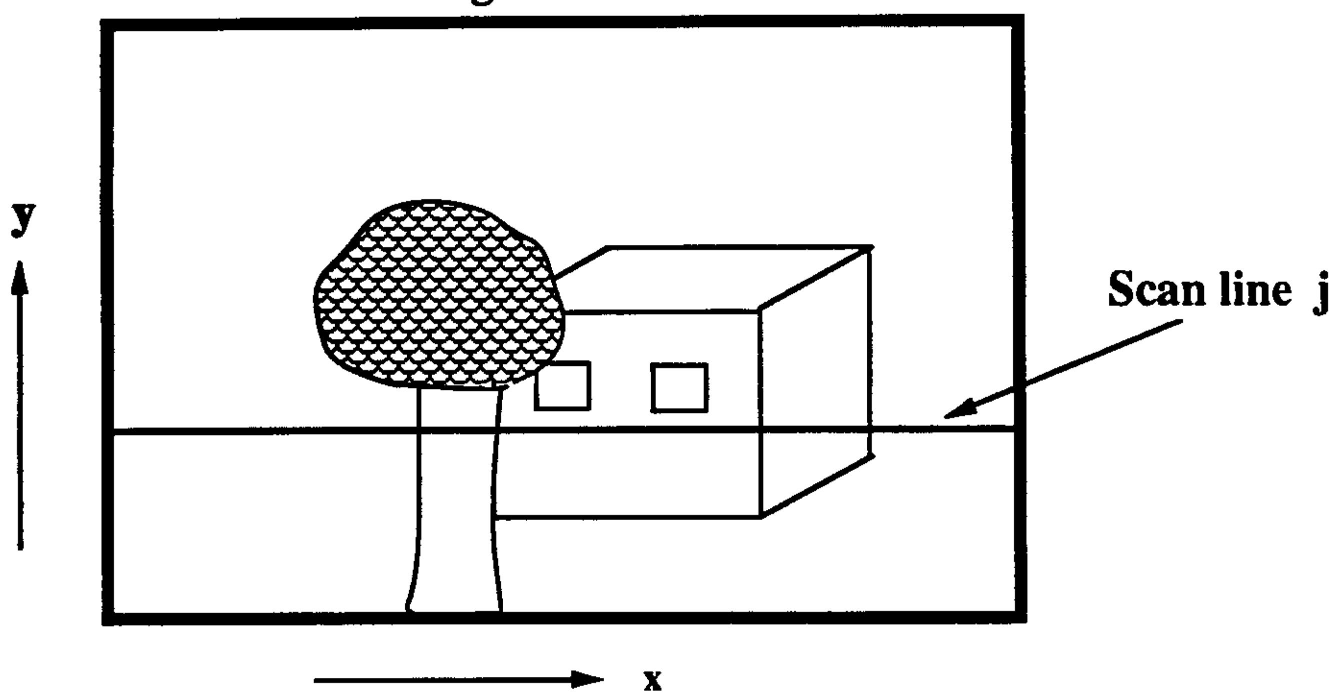
This section briefly describes three optical flow extraction techniques that we have implemented for our experiments. These are (1) spatial and temporal derivatives, (2) spatio-temporal imagery and (3) temporal cross correlation. The technique of spatial and temporal derivatives involves calculating both spatial and temporal derivatives at each point in the image [5]. Because a spatially local image operator is involved, this method normally computes only the flow component parallel to the spatial gradient (as described above). However we make use of the known flow direction to compute the true flow as the ratio of the temporal derivative to the directional spatial derivative along the flow line. This technique has been implemented to run in real-time on the PIPE image processing machine [6] and is described in more detail in [7].

Another technique we have implemented is the method of spatio-temporal image analysis (or epipolar-plane image analysis [8]). Suppose that the camera is moving in the horizontal direction, along its scan lines, and consider scan line j in the image (Figure 4a). We can associate a spatio-temporal image with this scan line (Figure 4b) in which the x-axis represents the x-direction along the scan line and the y-axis represents time. Each row of the spatio-temporal image will show the given scan line at a different point in time. At each instant of time, the current rows in the image are shifted up by one row and the new version of scan line j is inserted at the bottom of the image. Each feature on the original scan line will generate a line, or *streak,* in the spatio-temporal image. The slope of the streak is proportional to the optical flow of the feature; the smaller the slope, the greater the flow velocity. This algorithm has been implemented to run in real time on PIPE. We have also implemented edge finding techniques to extract the slope of the streaks. The resulting slope images are thresholded to perform range segmentation. Further details are provided in [9]. Notice that this technique requires advance knowledge of the flow direction. In principle, if the flow direction is radially outward from the FOE rather than along a scan line, motion along the radial line as a function of time can be plotted in the spatio-temporal image.

Another technique that we have implemented is temporal cross correlation. This is shown in Figure 5. For some scan line j in the image, we obtain the temporal brightness profile (i.e., intensity vs. time) for each of two successive pixels i and i+1. These pixels may be adjacent to one

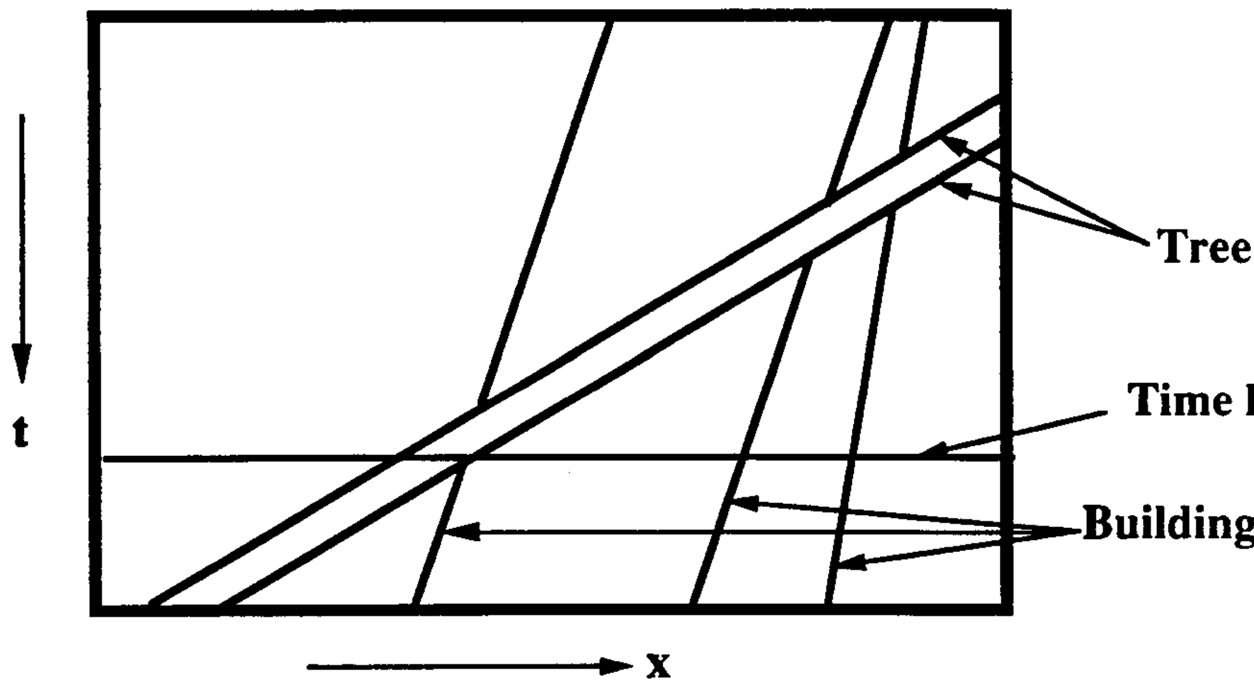
SPATIO-TEMPORAL IMAGERY

Image at time k



Motion along scan lines.
(a)

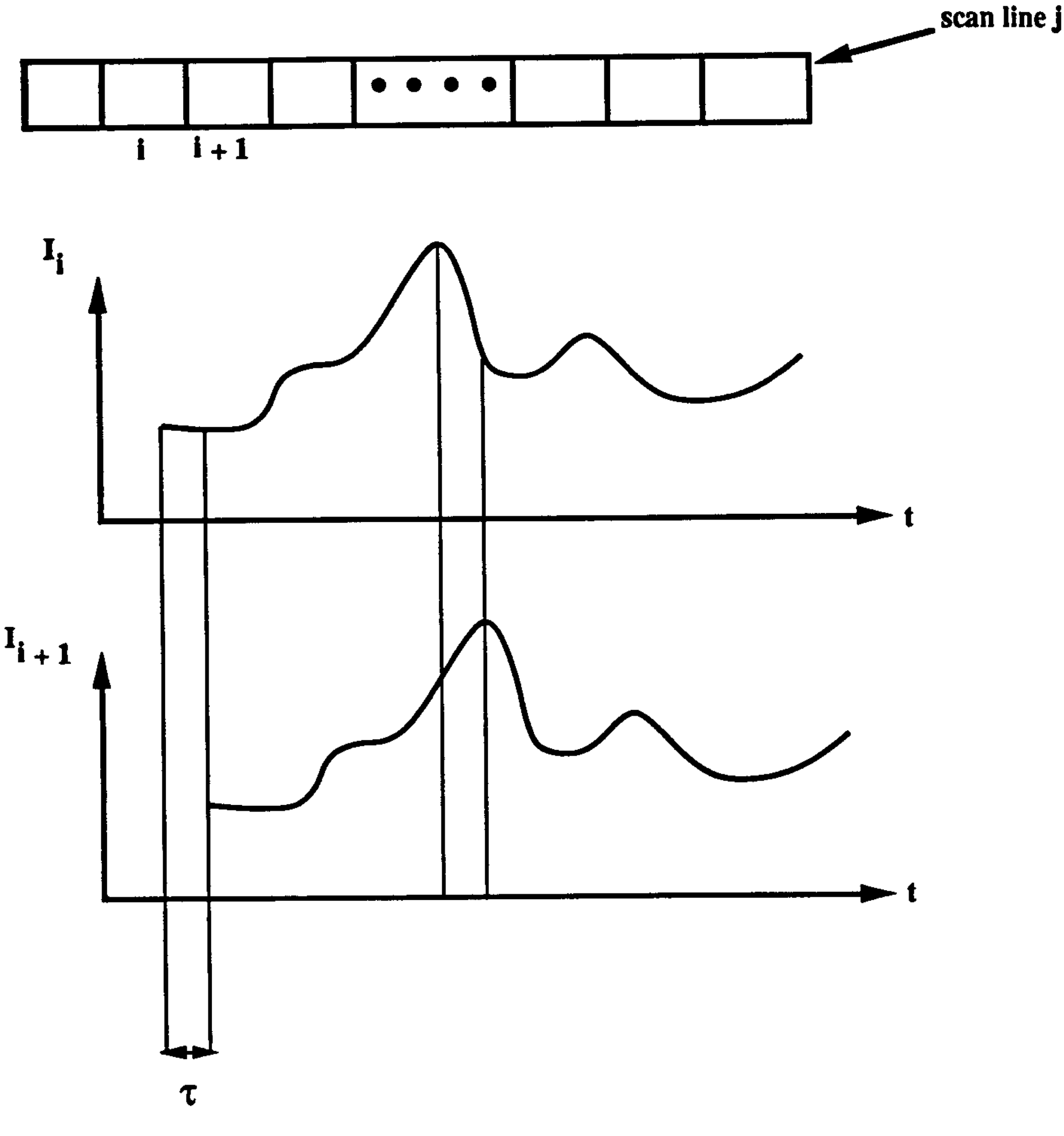
Image for scan line j



Slope of streak inversely proportional to flow velocity
(b)

Figure 4.

TEMPORAL CROSS CORRELATION



$$\Phi(\tau) = \int^{\tau} I(t) \bullet I(t+\tau)\, dt$$

Figure 5.

another or may be separated by some distance. By correlating these profiles, we can determine the amount of time τ for brightness values to travel from one pixel to the next. Hardware for performing this is described in [3].

3. Using Optical Flow for Unmanned Vehicles

In order to use optical flow for unmanned vehicles, we classify the flow into two categories: (1) the global flow field induced by vehicle and camera motion, and (2) the local flow fields induced by motion of objects in the environment. The flow due to the first category is useful for vehicle navigation. It consists of all flow arising from the stationary objects in the environment. The flow due to the second category is useful for acquiring, identifying, and tracking moving targets or objects. To make use of the global flow field, we need an inertial navigation system that gives us vehicle and camera motion. Once we know the camera motion, we can convert the optical flow image into a range image. The range image can then be analyzed to determine positions of objects, to describe the local terrain, and to determine the roughness of the surface areas ahead (from bumps, ruts, potholes, gullies, etc.). Objects that are nearby are potential obstacles. Other objects may be landmarks, and a recognition algorithm may be able to recognize them. Also, the optical flow can be used directly to determine steering commands for the vehicle, as in road following operations and other basic navigation behaviors. For teleoperated low data rate driving, areas with the potential obstacles or rough surfaces can be transmitted to the remote operator and displayed on his console to help him in making steering decisions. (This will be described below.) To make all of this work in real time, the camera should be stabilized so that optical flow due to camera jitter is eliminated.

4. Basic Navigation Behaviors

This section describes how some basic navigation behaviors can be achieved using optical flow and other image information to directly determine steering commands. The topic of road following will be discussed in the next section.

1. Steering toward a specific object or location.

   -- When the vehicle is undergoing forward translation in a static environment, the flow field appears to flow radially outward from the FOE, as explained above. The FOE is also the direction of vehicle motion (i.e., the 3-D vector from the camera focal point to the FOE in the image plane represents the direction of motion). As the moving vehicle changes direction, the FOE in the image shifts position (assuming the camera orientation relative to the vehicle is fixed). Therefore to maintain movement in the direction of an object, the vehicle should be steered so as to servo the FOE to keep it as near as possible to the region in the image that represents the object.

2. Approaching an object without collision.

   -- Imminent collision with an object will appear in the image sequence as a rapid expansion of the region that represents the object. By limiting this rate of expansion (which can be measured using optic flow), the object can be approached without collision.

3. Following or pursuing a moving object.

   -- If the vehicle is pursuing a moving object and the size of the object in the image is increasing, then the vehicle is gaining on the object; if the size is decreasing, the vehicle is losing ground; otherwise the vehicle's range to the object is remaining steady. In order to follow a moving object at a steady distance, a constant object size in the image should be maintained. The FOE should be servoed (by steering the vehicle) so as to maintain it near the region in the image that represents the object.

4. Avoiding obstacles.

   -- To avoid an obstacle, the FOE must be servoed so as to maintain it in a region in the image outside the boundaries of the region that represents the obstacle.

5. Vision-Based Road Following

Our approach to the road following problem builds upon the theoretical framework of the recently developed visual field theory [10, 11]. This theory provides quantitative relationships between a stationary 3-D environment and a moving camera. The theory involves pre-computing the expected instantaneous optical flow values in the camera imagery arising from every point in 3-D space. This is accomplished by generating the structure of a field in 3-D space consisting of contours and surfaces surrounding the moving camera (Figure 6). If static objects are placed anywhere in the surrounding space, the optical flow produced by these objects in the camera is predicted by the field theory. The field is always centered at the camera pinhole point and moves with the camera. The structure of the field changes as a function of the instantaneous camera motion. The theory provides a theoretical and scientific basis for optical flow-based road following algorithms.

Using this theory, we have shown that, in principle, the only road feature necessary for following curved, convex roads is the position of the tangent point on the road edge (i.e., the point on the road edge lying on an imaginary line tangent to the road edge and passing through the camera, Figure 7) and its optical flow. This is in contrast to most current systems which attempt to find as much about the road as possible. Existing systems often ignore the tangent point when making steering decisions, and usually are not concerned with the optical flow values of points on the road. In practice, larger portions of the road may have to be extracted in order to reliably find the tangent point. We have also shown that fast, simple control approaches are possible that directly use this image information. For more details on this approach to road following, see [12].

6. Low Data Rate Remote Vehicle Driving

Remotely driving a ground vehicle involves an operator who sits at a remote control center and views video images that are transmitted from one or more cameras mounted on the vehicle. While observing these images, the operator drives the vehicle by means of driving controls involving steering, brakes, throttle, and transmission. These controls generate appropriate driving actuator signals which are transmitted to the vehicle [15, 16].

In order for the operator to effectively drive the vehicle, the video images must be of sufficient quality and must be updated as frequently as possible. Full rate video transmission from the vehicle to the operator requires about 60 megabits per second for 512 x 512 images with 8 bits per pixel at 30 frames per second. However there are several problems with using the wide communication bandwidth required for such transmission. First, wide bandwidth radio communication requires direct line of sight between the transmitter and receiver. This is not feasible in realistic outdoor scenarios where vehicles are likely to be driven behind hills and mountains and therefore hidden from direct view by the operator station. Second, wide bandwidth links are relatively expensive. Finally, full rate video uses up a large part of the bandwidth allocations. This could present a problem if there are many vehicles being operated simultaneously, since wide bandwidth links for the vehicles could combine to exceed the entire communication spectrum. Fiber optic tethers, which

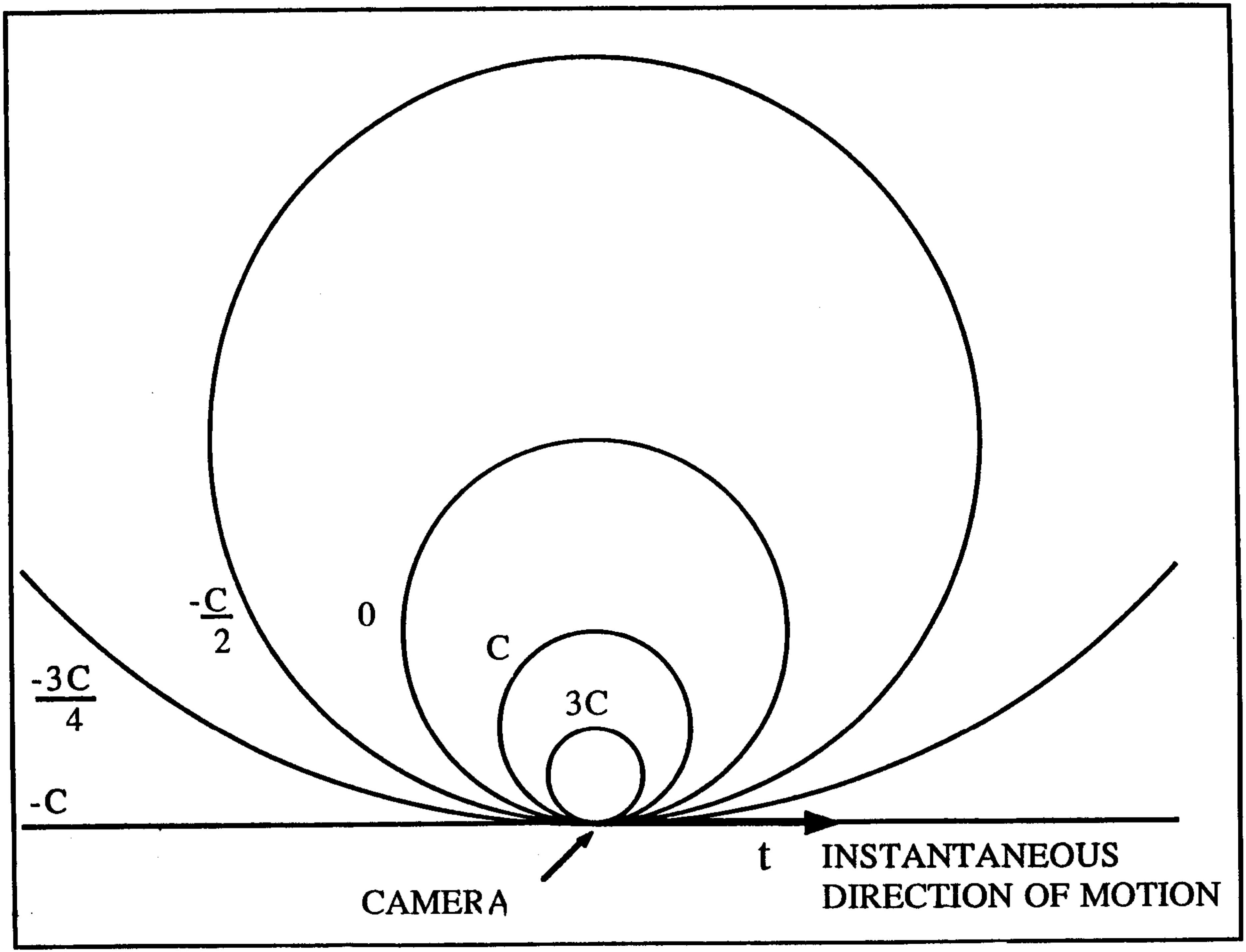


Figure 6. Optical Flow Values Due to Camera Motion

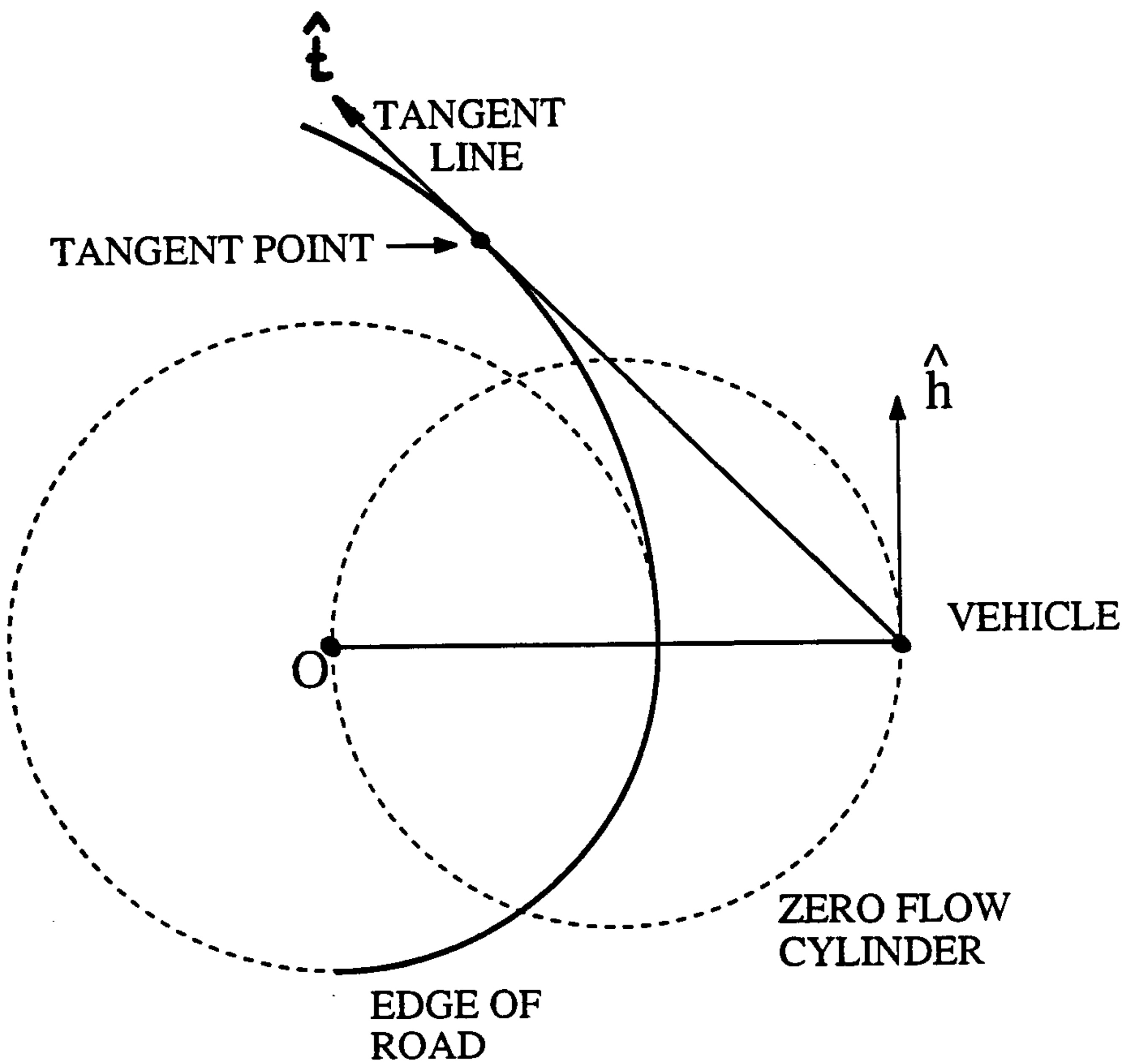


Figure 7. Road Following

have wide bandwidth communications capabilities, also have several problems, including limited ruggedness, difficulties in deployment and retrieval, and the problem of repairs.

Many of these difficulties can be overcome by utilizing narrow band radio links which have communication bandwidths on the order of 60 kilobits per second or less. Since the full video desired for teleoperation cannot be transmitted over narrow band links, efficient and effective techniques of real-time video compression offering compression ratios of 1000:1 or greater must be developed.

To solve the problem of video compression for remote driving, we have proposed a hybrid method which combines image processing compression (i.e., techniques whose input is an image and whose output is a compressed image), discrete cosine transform compression, temporal frame rate reduction (i.e., transmitting much less than 30 frames per second), and intelligent on-board processing of visual data to detect obstacles, surface roughness, and road edges. The sequence of events as they would occur in the system is as follows. Images are obtained from one or more cameras mounted on-board the vehicle. These images then undergo compression using the hybrid technique. After the compressed code is transmitted over a communication link to the operator station, it is decompressed so as to result in a sequence of full resolution images. Obstacles and road edges transmitted to the remote operator are then used to enhance his video imagery.

Several algorithms for the image processing compression portion of this method have been implemented to run in real time on PIPE and have been used in actual field experiments for teleoperated control. The algorithms implemented on PIPE include grey level quantization, non-maximal edge suppression, foveal-peripheral simulation, image differencing, histogram slicing, binning, Laplacian pyramid decomposition and reconstruction, decimation and Poisson interpolation, and linear predictive coding. Details describing these algorithms and their implementation on PIPE are presented in [13], while results of field experiments are described in [14].

Video compression techniques that provide compression ratios of 1000:1 or greater in real-time will suffer from degraded imagery at the operator's console. The degradation might be in the form of limited image resolution, limited contrast between obstacles and background, or limited frame rate. These limitations will decrease the ability of the operator to (1) detect, recognize, and track obstacles in real time, (2) detect, recognize, and characterize surface roughness immediately ahead of the vehicle, and (3) detect and track road edges during road following operations. Information transmitted to the operator using intelligent on-board processing of visual data can be used to enhance his video imagery so as to overcome some of the problems due to the degraded imagery. The intelligent on-board processing can be done using optical flow analysis and road tracking analysis.

For example, if obstacles are detected using optical flow analysis, then the position in the image of each obstacle, as well as its extent in the image, can be transmitted to the operator station. This information can then be overlayed on top of the imagery viewed by the operator (perhaps in the form of a flashing box enclosing the obstacle) to alert the operator. If rough surface areas are detected using image flow analysis and transmitted to the operator, these areas can be highlighted to alert the operator so that he avoids these areas or slows down. Similarly, if road edges are detected using on-board procesing, then the image positions of the road edges can be transmitted to the operator and overlayed on top of the imagery viewed by him.

7. Conclusion

This paper has presented several ways in which optical flow can be used to perform visual navigation of ground vehicles. A primary difference between our approach and most previous approaches to visual navigation is that we perform dynamic image analysis as opposed to static image analysis to obtain information about the environment. This allows us to achieve real-time results more easily.

We have implemented several optical flow extraction algorithms on the PIPE real-time image processing computer. We have also described how optical flow can be used for basic navigation behaviors, including road following, obstacle avoidance, pursuing a moving target, and steering towards and approaching a stationary target.

Finally, we have shown how obstacle and surface roughness information obtained using optical flow can be used as part of a system for teleoperated low data rate driving of a vehicle.

Acknowledgements

This work has been supported by the U.S. Army Laboratory Command, Human Engineering Laboratory and Harry Diamond Laboratories, and by the Defense Advanced Research Projects Agency (DARPA), Tactical Technology Office. Special thanks go to Mr. Charles M. Shoemaker, formerly of the Human Engineering Laboratory, to Dr. Philip Emmerman of Harry Diamond Laboratories, and to Dr. Jasper Lupo of DARPA for their direction and guidance.

References

1. Aloimonos, J. "Purposive and Qualitative Active Vision." *Proc. DARPA Image Understanding Workshop,* September 1990.
2. Gibson, J.J. *The Senses Considered as Perceptual Systems.* Houghton Mifflin, Boston, 1966.
3. Albus, J.S. and Hong, T.-H. "Motion, Depth, and Image Flow." *Proc. 1990 IEEE International Conference on Robotics and Automation,* Cincinnati, Ohio, May 1990.
4. Marr, D. and Ullman, S. "Directional Selectivity and Its Use in Early Visual Processing." *Proc. R. Soc. Lond. B,* Vol. 211, 151-180, 1981.
5. Horn, B.K.P. and Schunck, B.G. "Determining Optical Flow." *Artificial Intelligence,* 17, 1981, 185-203.
6. Kent, E.W., Shneier, M.O. and Lumia, R. "PIPE (Pipelined Image Processing Engine)." *J. Parallel and Distributed Computing,* 2, 50-78, 1985.
7. Rangachar, R., Hong, T.-H., Herman, M. and Lupo, J. "Real-Time Implementation of a Differential Range Finder." *Proc. SPIE Real-Time Image Processing II: Algorithms, Architectures, and Applications,* Vol. 1295, Orlando, Florida, April 1990, 211-222.
8. Bolles, R.C. and Baker, H.H. "Epipolar-Plane Image Analysis: A Technique for Analyzing Motion Sequences." *Proc. DARPA Image Understanding Workshop,* Miami Beach, Florida, December 1985, 137-148.
9. Rangachar, R., Hong, T.-H., Herman, M., Luck, R. and Lupo, J. "Real-Time Differential Range Estimation Based on Time-Space Imagery Using PIPE." *Proc. SPIE Real-Time Image Processing II: Algorithms, Architectures, and Applications,* Vol. 1295, Orlando, Florida, April 1990, 247-256.

10. Raviv, D. "A Quantitative Approach to Camera Fixation." Internal Report NISTIR 4324, National Institute of Standards and Technology, May 1990.

11. Raviv, D. and Herman, M. "Towards an Understanding of Camera Fixation." *Proc. 1990 IEEE International Conference on Robotics and Automation,* Cincinnati, Ohio, May 1990, 28-33.

12. Raviv, D. and Herman, M. "A New Approach to Vision and Control for Road Following." Internal Report NISTIR 4476, National Institute of Standards and Technology, January 1991.

13. Herman, M., Chaconas, K., Nashman, M. and Hong, T.-H. "Low Data Rate Remote Vehicle Driving." *Proc. IEEE International Symposium on Intelligent Control 1988,* Arlington, VA, August 1988, 376-381.

14. Herman, M., Chaconas, K., Nashman, M. and Hong, T.-H. "Video Compression for Remote Vehicle Driving." *Proc. SPIE Advances in Intelligent Robotics Systems: Mobile Robots III,* Vol. 1007, Cambridge, MA, November 1988, 136-143.

15. Corbett, G.K., Killough, S.M., Richardson, B.S., and Herman, M. "A Human Factors Testbed for Ground-Vehicle Telerobotics Research." *Proc. IEEE Southeastcon '90,* New Orleans, Louisiana, April 1990, 618-620.

16. Corbett, G.K., Killough, S.M., Richardson, B.S., and Herman, M. "A Human Factors Research Environment for Ground-Vehicle Telerobotics Research." *Proc. Third International Symposium on Robotics and Manufacturing,* Vancouver, B.C., Canada, July 1990.