CAR-TR-881
CS-TR-3882

N00014-95-1-0521
March 1998

Estimating Relative Vehicle Motions in Traffic Scenes

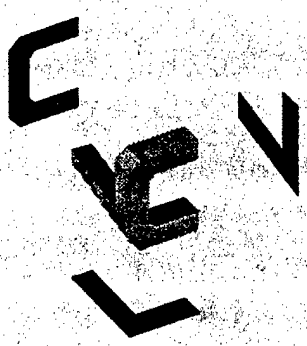
Zoran Duric[1,2], Roman Goldenberg[3], Ehud Rivlin[1,3], Azriel Rosenfeld[1]

[1]Center for Automation Research
University of Maryland
College Park, MD 20742-3275

[2] Computer Science Department
George Mason University
Fairfax, VA 22030-4444

[3]Department of Computer Science
Technion - Israel Institute of Technology
Haifa, Israel 32000

COMPUTER VISION LABORATORY



CENTER FOR AUTOMATION RESEARCH

UNIVERSITY OF MARYLAND
COLLEGE PARK, MARYLAND
20742-3275

CAR-TR-881
CS-TR-3882

N00014-95-1-0521
March 1998

Estimating Relative Vehicle Motions in Traffic Scenes

Zoran Duric[1,2], Roman Goldenberg[3], Ehud Rivlin[1,3], Azriel Rosenfeld[1]

[1]Center for Automation Research
University of Maryland
College Park, MD 20742-3275

[2] Computer Science Department
George Mason University
Fairfax, VA 22030-4444

[3]Department of Computer Science
Technion – Israel Institute of Technology
Haifa, Israel 32000

Abstract

Autonomous operation of a vehicle on a road calls for understanding of various events involving the motions of the vehicles in its vicinity. In this paper we show how a moving vehicle which is carrying a camera can estimate the relative motions of nearby vehicles. We present a model for the motion of the observing vehicle, and show how to "stabilize" it, i.e. to correct the image sequence so that transient motions resulting from bumps, etc. are removed and the sequence corresponds more closely to the sequence that would have been collected if the motion had been smooth. We also model the motions of nearby vehicles and show how to detect their motions relative to the observing vehicle. We present results for several road image sequences which demonstrate the effectiveness of our approach.

The support of the Office of Naval Research under Grant N00014-95-1-0521 is gratefully acknowledged.

1 Introduction

Autonomous operation of a vehicle on a road calls for understanding of various events involving the motions of the vehicles in its vicinity. In normal traffic flow, most of the vehicles on a road move in the same direction without major changes in their distances and relative speeds. When a nearby vehicle deviates from this norm (e.g. when it passes or changes lanes), or when it is on a collision course, some action may need to be taken. In this paper we show how a vehicle carrying a camera can estimate the relative motions of nearby vehicles.

Understanding the relative motions of vehicles requires modeling both the motion of the observing vehicle and the motions of the other vehicles. We represent the motions of vehicles using a *Darboux motion* model that corresponds to the motion of an object moving along a smooth curve that lies on a smooth surface. We show that deviations from Darboux motion correspond primarily to small, rapid rotations around the axes of the vehicle. These rotations arise from the vehicle's suspension elements in response to unevenness of the road. We estimate bounds on both the smooth rotations due to Darboux motion (from highway design principles) and the non-smooth rotations due to the suspension. We show that both types of rotational motion, as well as the non-smooth translational component of the motion (bounce), are small relative to the smooth (Darboux) translational motion of the vehicle.

This analysis is used to model the motions of both the observer and observed vehicles. We use the analysis to show that only the rotational velocity components of the observer vehicle are important. On the other hand, the rotational velocity components of an observed vehicle are negligible compared to its translational velocity. As a consequence we need to estimate the rotational velocity components only for the observing vehicle. This is the case even when an observed vehicle is changing its direction of motion relative to the observing vehicle (turning or changing lanes); the turn shows up as a gradual change in the direction of the relative translational velocity.

An important consequence of the Darboux motion model is that for a fixed forward-looking camera mounted on the observer vehicle the direction of translation (and therefore the position of the focus of expansion (FOE)) remains the same in the images obtained by the camera. We use this fact to estimate the observing vehicle's rotational velocity components; this is done by finding the rotational flow which, when subtracted from the observed flow, leaves a radial flow pattern (radiating from the FOE) of minimal magnitude.

We describe the motion field using full perspective projection, estimate its rotational components, and derotate the field. The flow fields of nearby vehicles are then, under the Darboux motion model, pure translational fields. We analyze the motions of the other vehicles under weak perspective projection, and derive their motion parameters. We present results for several road image sequences obtained from cameras carried by moving vehicles. The results demonstrate the effectiveness of our approach.

In the next section we present motion models for road vehicles, discuss ideal and real vehicle motion, and analyze the relative sizes of the smooth and non-smooth velocity components. In Section 3 we discuss the image motion and describe a way to estimate the necessary derotation.

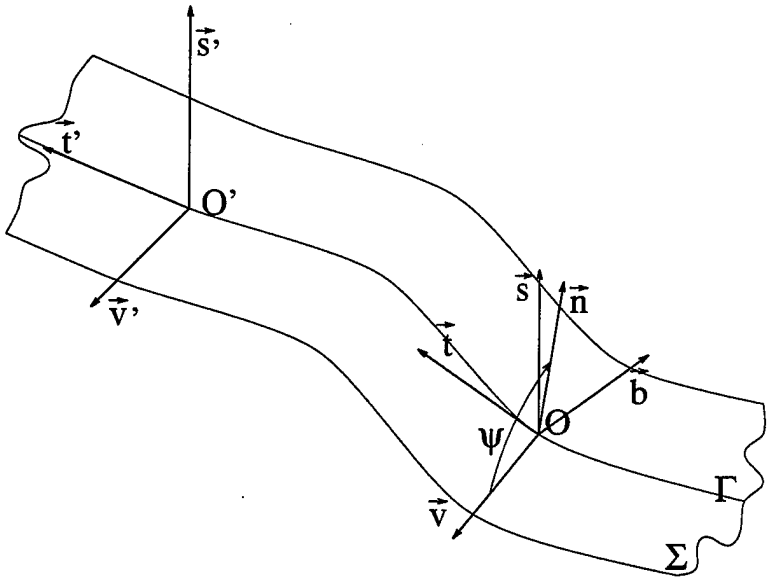


Figure 1: The Darboux frame moves along the path Γ which lies on the surface Σ.

Section 4 describes methods of estimating nearby vehicle motions from the normal flow field. Section 5 presents experimental results for several sequences taken at different locations. In Section 6 we review some prior work on various topics that are related to this paper, involving road detection, vehicle detection, and motion analysis, and compare them with the approach presented in this paper.

2 Motion Models of Highway Vehicles

The ideal motion of a ground vehicle does not have six degrees of freedom. If the motion is (approximately) smooth it can be described as motion along a smooth trajectory Γ lying on a smooth surface Σ. Moreover, we shall assume that the axes of the vehicle (the fore/aft, crosswise, and up/down axes) are respectively parallel to the axes of the *Darboux frame* defined by Γ and Σ. These axes are defined by the tangent $\vec{\mathbf{t}}$ to Γ (and Σ), the second tangent $\vec{\mathbf{v}}$ to Σ (orthogonal to $\vec{\mathbf{t}}$), and the normal $\vec{\mathbf{s}}$ to Σ (see Figure 1). Our assumption about the axes is reasonable for the ordinary motions of standard types of ground vehicles; in particular, we are assuming that the first two vehicle axes are parallel to the surface and that the vehicle's motion is parallel to its fore/aft axis (the vehicle is not skidding).

Consider a point O moving along a (space) curve Γ. There is a natural coordinate system $Otnb$ associated with Γ, defined by the tangent $\vec{\mathbf{t}}$, normal $\vec{\mathbf{n}}$, and binormal $\vec{\mathbf{b}}$ of Γ. The triple $(\vec{\mathbf{t}}, \vec{\mathbf{n}}, \vec{\mathbf{b}})$ is called the *moving trihedron* or *Frenet-Serret coordinate frame.* We have the Frenet-Serret formulas [22]

$$\vec{\mathbf{t}}' = \kappa\vec{\mathbf{n}}, \quad \vec{\mathbf{n}}' = -\kappa\vec{\mathbf{t}} + \tau\vec{\mathbf{b}}, \quad \vec{\mathbf{b}}' = -\tau\vec{\mathbf{n}} \tag{1}$$

where κ is the curvature and τ the torsion of Γ.

When the curve Γ lies on a smooth surface Σ, it is more appropriate to use the *Darboux*

frame ($\vec{\mathbf{t}}$, $\vec{\mathbf{v}}$, $\vec{\mathbf{s}}$) [22]. We take the first unit vector of the frame to be the tangent $\vec{\mathbf{t}}$ of Γ and the surface normal $\vec{\mathbf{s}}$ to be the third frame vector; finally we obtain the second frame vector as $\vec{\mathbf{v}} = \vec{\mathbf{s}} \times \vec{\mathbf{t}}$ (see Figure 1). Note that $\vec{\mathbf{t}}$ and $\vec{\mathbf{v}}$ lie in the tangent plane of Σ. Since the vector $\vec{\mathbf{t}}$ belongs to both the *Otnb* and *Otvs* frames, they differ only by a rotation around $\vec{\mathbf{t}}$, say through an angle $\psi \equiv \psi(s)$. We thus have

$$\begin{pmatrix} \vec{\mathbf{v}} \\ \vec{\mathbf{s}} \end{pmatrix} = \begin{pmatrix} \cos\psi & \sin\psi \\ -\sin\psi & \cos\psi \end{pmatrix} \begin{pmatrix} \vec{\mathbf{n}} \\ \vec{\mathbf{b}} \end{pmatrix}. \tag{2}$$

The derivatives of $\vec{\mathbf{t}}, \vec{\mathbf{v}}, \vec{\mathbf{s}}$ with respect to arc length along Γ can be found from (1) and (2):

$$\vec{\mathbf{t}}' = \kappa_g \vec{\mathbf{v}} - \kappa_n \vec{\mathbf{s}}, \quad \vec{\mathbf{v}}' = -\kappa_g \vec{\mathbf{t}} + \tau_g \vec{\mathbf{s}}, \quad \vec{\mathbf{s}}' = \kappa_n \vec{\mathbf{t}} - \tau_g \vec{\mathbf{v}} \tag{3}$$

where

$$\kappa_g \equiv \kappa \cos\psi, \quad \kappa_n \equiv \kappa \sin\psi, \quad \tau_g \equiv \tau + \frac{d\psi}{ds};$$

κ_g is called the *geodesic curvature*, κ_n is called the *normal curvature*, and τ_g is called the *(geodesic) twist.*

It is well known that the instantaneous motion of a moving frame is determined by its rotational velocity $\vec{\omega}$ and the translational velocity $\vec{T}$ of the reference point of the frame. The translational velocity $\vec{T}$ of O is just $\vec{\mathbf{t}}$ and the rotational velocity of the *Otvs* frame is given by the vector

$$\vec{\omega}_d = \tau_g \vec{\mathbf{t}} + \kappa_n \vec{\mathbf{v}} + \kappa_g \vec{\mathbf{s}}.$$

Hence the derivative of any vector in the *Otvs* frame is given by the vector product of $\vec{\omega}_d$ and that vector. It can be seen that the rate of rotation around $\vec{\mathbf{t}}$ is just τ_g, the rate of rotation around $\vec{\mathbf{v}}$ is just κ_n, and the rate of rotation around $\vec{\mathbf{s}}$ is just κ_g.

If, instead of using the arc length s as a parameter, the time t is used, the rotational velocity $\vec{\omega}_d$ and translational velocity $\vec{T}$ are scaled by the speed $v(t) = ds/dt$ of O along Γ.

2.1 Real Vehicle Motion

We will use two coordinate frames to describe vehicle motion. The "real" vehicle frame $C\xi\eta\zeta$ (which moves non-smoothly, in general) is defined by its origin C, which is the center of mass of the vehicle, and its axes: $C\xi$ (fore/aft), $C\eta$ (crosswise), and $C\zeta$ (up/down); and the ideal vehicle frame *Otvs* (the Darboux frame), which corresponds to the smooth motion of the vehicle.

The motion of the vehicle can be decomposed into the motion of the *Otvs* frame and the motion of the $C\xi\eta\zeta$ frame relative to the *Otvs* frame. As we have just seen, the rotational velocity of the *Otvs* (Darboux) frame is $v\vec{\omega}_d = v(\tau_g \vec{\mathbf{t}} + \kappa_n \vec{\mathbf{v}} + \kappa_g \vec{\mathbf{s}})$ and its translational velocity is $v\vec{\mathbf{t}}$. We denote the rotational velocity of the $C\xi\eta\zeta$ (vehicle) frame by $\vec{\omega}_v$ and its translational velocity by $\vec{T}_v$.

The position of the $C\xi\eta\zeta$ frame relative to the *Otvs* frame is given by the displacement vector $\vec{d}_{v/d}$ between C and O, and the relative orientation of the frames is given by an orthogonal

rotational matrix (matrix of direction cosines) which we denote by $R_{v/d}$. The translational velocity of the vehicle (the velocity of C) is the sum of three terms: (i) the translational velocity of the Darboux frame $v\vec{\mathbf{t}}$, (ii) the translational velocity $\vec{T}_{v/d} \equiv \dot{\vec{d}}_{v/d}$, and (iii) the displacement $v\vec{\omega}_d \times \vec{d}_{v/d}$ due to rotation of C in the $Otvs$ frame. The translational velocity of the vehicle expressed in the $Otvs$ frame is thus $v\vec{\omega}_d \times \vec{d}_{v/d} + v\vec{\mathbf{t}} + \dot{\vec{d}}_{v/d}$; its translational velocity in the $C\xi\eta\zeta$ frame is

$$\vec{T}_v = R_{v/d}^T(v\vec{\omega}_d \times \vec{d}_{v/d} + v\vec{\mathbf{t}} + \dot{\vec{d}}_{v/d}). \tag{4}$$

Similarly, the rotational velocity of $C\xi\eta\zeta$ is the sum of two terms: (i) the rotational velocity $vR_{v/d}^T\vec{\omega}_d$ of the $Otvs$ frame, and (ii) the rotational velocity $\vec{\omega}_{v/d}$, which corresponds to the skew matrix $\Omega_{v/d} = R_{v/d}^T\dot{R}_{v/d}$. The rotational velocity of the $C\xi\eta\zeta$ frame expressed in the $Otvs$ frame is thus $v\vec{\omega}_d + R_{v/d}\vec{\omega}_{v/d}$; the corresponding expression in the $C\xi\eta\zeta$ frame is

$$\vec{\omega}_v = vR_{v/d}^T\vec{\omega}_d + \vec{\omega}_{v/d}. \tag{5}$$

Rotations around the fore/aft, sideways, and up/down axes of a vehicle are called *roll*, *pitch*, and *yaw*, respectively. In terms of our choice of the real vehicle coordinate system, these are rotations around the ξ, η, and ζ axes.

2.2 Departures of Vehicle Motion from Smoothness

The motion of a ground vehicle depends on many factors: the type of intended motion; the speed of the vehicle; the skill of the driver; the size, height and weight of the vehicle; the type and size of the tires (or tractor treads), and the nature of the suspension mechanism, if any; and the nature of the surface on which the vehicle is being driven. These factors tend to remain constant; they undergo abrupt changes only occasionally, e.g. if a tire blows out, or the vehicle suddenly brakes or swerves to avoid an obstacle, or the type of surface changes. Such events may produce impulsive changes in the vehicle's motion, but the effects of these changes will rapidly be damped out. In addition to these occasional events, "steady-state" non-smoothness of a ground vehicle's motion may result from roughness of the surface.

A ground vehicle drives over roads that have varying degrees of roughness [2]. The roughness consists primarily of small irregularities in the road surface. In discussing the effects of the roughness of the road on the motion of a ground vehicle we will assume, for simplicity, an ordinary, well balanced four-wheeled vehicle moving on a planar surface that is smooth except for occasional small bumps (protrusions). The bumps are assumed to be "small" relative to the size of the wheels, so that the effect of a wheel passing over a bump is impulsive. (We could also allow the surface to have small depressions, but a large wheel cannot deeply penetrate a small depression, so the depressions have much smaller effects than the bumps.)

As the vehicle moves over road surface, each wheel hits bumps repeatedly. We assume that the vehicle has a suspension mechanism which integrates and damps the impulsive effects of the bumps. Each suspension element is modeled by a spring with damping; its characteristic function is a sine function multiplied by an exponential damping function (see [24, 34]).We

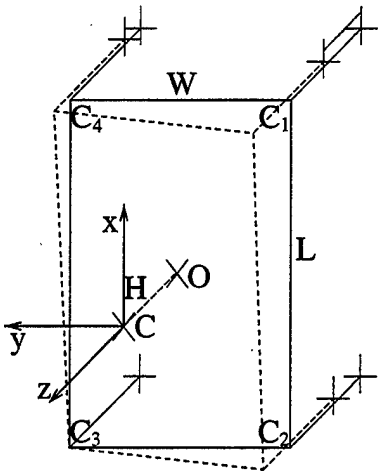


Figure 2: A possible motion of the base of a vehicle as the vehicle hits a bump.

assume that the suspension elements associated with the four wheels are independent of each other and are parallel to the vertical axis of the vehicle.

The vehicle may hit new bumps while the effects of the previous bumps are still being felt. Each hit forces the suspension and adds to the accumulated energy in the spring; thus it can happen that the suspension is constantly oscillating, which has the effect of moving the corners of the vehicle up and down. The period of oscillation is typically on the order of 0.5 sec (see [24, 34]). In general, it takes several periods to damp out the spring; for example, the damping ratio provided by shock absorbers of passenger cars is in the range $0.2 - 0.4$. The maximum velocity of the oscillation is typically on the order of 0.1 m/sec.

Consider the coordinate system $Cxyz$ with origin at the center of mass C of the vehicle (see Figure 2). Let v_i be the velocity corner C_i of the vehicle, and let the length and width of the vehicle be L and W. From the v_i's we can compute the angular velocity matrix

$$\tilde{\Omega} \equiv \begin{pmatrix} 0 & -\omega_z & \omega_y \\ \omega_z & 0 & -\omega_x \\ -\omega_y & \omega_x & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & (v_3 - v_4)/L \\ 0 & 0 & (v_2 - v_3)/W \\ (v_4 - v_3)/L & (v_3 - v_2)/W & 0 \end{pmatrix}. \tag{6}$$

Note that any of the v_is can be positive or negative. Multiplication by $\tilde{\Omega}$ can be replaced by the vector product with the angular velocity vector $\vec{\omega} = \vec{\imath}(v_3 - v_2)/W + \vec{\jmath}(v_3 - v_4)/L$ where the rate of rotation around the x axis (the roll velocity) is $\omega_x = (v_3 - v_2)/W$ and the rate of rotation around the y axis (the pitch velocity) is $\omega_y = (v_3 - v_4)/L$. As noted above, we typically have $|v_i| < 0.1\,\text{m/sec}$. If we assume that $W > 1\,\text{m}$ and $L > 2\,\text{m}$ we have $|\omega_x| < 0.2\,\text{rad/sec}$ and $|\omega_y| < 0.1\,\text{rad/sec}$. The yaw velocity component is $|\omega_z| = \mathcal{O}(v_i^2/\sqrt{W^2 + L^2})$ which is $|\omega_z| = \mathcal{O}(0.01)\,\text{rad/sec}$. (For a complete derivation see [15].)

The translational velocity vector of the center C of the vehicle is obtained by using the velocities $v_1 - v_3$, $v_2 - v_3$, 0, and $v_4 - v_3$ for the corners and adding v_3 to the velocity in the

direction of the z-axis. We thus have

$$\vec{T} = \begin{pmatrix} t_x \\ t_y \\ t_z \end{pmatrix} = \begin{pmatrix} -(v_4 - v_3)H/L \\ (v_2 - v_3)H/W \\ (v_2 + v_4)/2 \end{pmatrix}. \tag{7}$$

If we assume that $H < 0.5\,\mathrm{m}$ ($< W/2$) we have $|t_x| < 0.05\,\mathrm{m/sec}$, $|t_y| < 0.1\,\mathrm{m/sec}$, and $|t_z| < 0.1\,\mathrm{m/sec}$.

We can draw several conclusions from this discussion: (i) The effects of small bumps are of short duration, i.e. they can be considered to be impulsive. The suspension elements integrate and damp these effects, resulting in a set of out-of-phase oscillatory motions. (ii) The yaw component of rotation due to the effect of a bump is very small compared to the roll and pitch components. (iii) The translational effects of a bump are proportional to the velocities (or displacements) of the suspension elements and the dimensions of the vehicle and are quite small.

2.3 The Sizes of the Smooth and Non-smooth Velocity Components

We now compare the sizes of the velocity components which are due to the ideal motion of the vehicle — i.e., the velocity components of the Darboux frame (Section 2) — to the sizes of the velocity components which are due to departures of the vehicle frame from the Darboux frame (Section 2.2).

The translational velocity of the Darboux frame is just $v\vec{\mathbf{t}}$; thus the magnitude of the translational velocity is just v. If $v = 10\,\mathrm{m/sec}$ ($= 36\,\mathrm{km/hr} \approx 22\,\mathrm{mi/hr}$) this velocity is much larger than the velocities which are due to departures of the vehicle from the Darboux frame, which, as we have just seen, are on the order of 0.1m/sec or less.

The rotational velocity of the Darboux frame is $v\vec{\omega}_d = v(\tau_g\vec{\mathbf{t}} + \kappa_n\vec{\mathbf{v}} + \kappa_g\vec{\mathbf{s}})$; thus the magnitude of the rotational velocity is $v\sqrt{\tau_g^2 + \kappa_n^2 + \kappa_g^2}$. In this section we will estimate bounds on τ_g, κ_n, and κ_g. Our analysis is based on the analyses in [2, 17, 34] and on the highway design recommendations published by the American Association of State Highway Officials [1].

Good highway design allows a driver to make turns at constant angular velocities, and to follow spiral arcs in transitioning in and out of turns, in order to reduce undesirable acceleration effects on the vehicle. A well-designed highway turn has also a transverse slope, with the outside higher than the inside, to counterbalance the centrifugal force on the turning vehicles. Thus the ideal (smooth) motion of a vehicle has piecewise constant translational and rotational velocity components, with smooth transitions between them. Note that the translational components are constant in the vehicle coordinate frame even when the vehicle is turning, unless it slows down to make the turn.

To illustrate the typical sizes of these components, consider a ground vehicle moving with velocity v along a plane curve Γ on the surface Σ. If Σ is a plane and Γ is a circular arc with radius of curvature $\rho_g = |\kappa_g|^{-1}$ (i.e., the vehicle is turning with a constant steering angle), the

angular velocity of the vehicle is $v\vec{\omega}_d = v\kappa_g\vec{s}$ and there is a centripetal acceleration $\vec{a}_c = v^2\kappa_g\vec{v}$ at the vehicle's center [33]. As a result there is a centrifugal force on the vehicle proportional to $\|\vec{a}_c\|$ and the mass of the vehicle. If skidding is to be avoided the limit on $\|\vec{a}_c\|$ (see [2]) is given by

$$\|\vec{a}_c\| = v^2\kappa_g \leq g(\tan\alpha + \mu_a) \tag{8}$$

where g is the gravitational acceleration, α is the transverse slope, and μ_a is the coefficient of adhesion between the wheels and the surface. [Typical values of μ_a range from $0.8 - 0.9$ for dry asphalt and concrete to 0.1 for ice (see [34], page 26).] From (8) we have either a lower bound on ρ_g for a given v or an upper bound on v for a given ρ_g. For example, if $v = 30\,\mathrm{m/sec}$ ($\approx 108\,\mathrm{km/hr}$), $\alpha = 0.05\,\mathrm{rad}$, and $\mu_a = 0.2$ from $v^2/\rho_g < 0.25g$ we have $\rho_g > 367\,\mathrm{m}$. This yields an upper bound on the yaw angular velocity of $< v|\kappa_g| = v/\rho_g \approx 0.08\,\mathrm{rad/sec}$, which is somewhat larger than the yaw angular velocity arising from the departures from Darboux motion.

Other dynamic constraints on a vehicle such as the limits on torques and forces can be used to obtain constraints on τ_g and κ_n. (These and other considerations such as safety and comfort were used in [1] to make recommendations for highway design; for a summary of these recommendations see [15].) For both vertical curves (crossing a hill) and turning curves the (recommended lower bound on the) radius of curvature $\rho_{\min}$ grows with the square of the design velocity v_d. However, the resulting (design) yaw and pitch angular velocities are limited by $v_d/\rho_{\min}$. Thus for smaller velocities v the vehicle can negotiate tighter vertical and turning curves and thus have even larger values of the yaw and pitch angular velocities. Typical values of the roll and pitch angular velocities are given in [15].

For realistic vehicle speeds we can conclude the following about the impulsive and smooth translational and rotational velocity components of the vehicle [15]. The impulsive effects on the translational velocity are approximately two orders of magnitude smaller than the smooth velocity components themselves. Impulsive effects on the yaw angular velocity are somewhat smaller than the smooth yaw component arising from worst-case turns of the road; for moderate turns the impulsive effects are comparable in size to the smooth yaw velocity. Impulsive effects on the roll angular velocity are approximately an order of magnitude larger than the smooth roll component arising from worst-case twists (and turns) of the road; for gentler twists the smooth roll velocity is even smaller. Similarly, impulsive effects on the pitch angular velocity are approximately an order of magnitude larger than the smooth pitch velocity arising from worst-case changes of vertical slope (i.e., vertical curves) of the road; for gentler vertical curves the smooth pitch angular is even smaller. (The impulsive effects are not significantly affected by turns, twists, or vertical slope.) We can thus conclude that impulsive effects on the roll and pitch angular velocities are significant and larger than the corresponding smooth velocities, and that impulsive effects on the yaw angular velocity are on the order of the smooth yaw velocity.

2.4 Camera Motion

Assume that a camera is mounted on the vehicle; let $\vec{d}_c$ be the position vector of the mass center of the vehicle relative to the nodal point of the camera The orientation of the vehicle coordinate

system $C\xi\eta\zeta$ relative to the camera is given by an orthogonal rotational matrix (a matrix of the direction cosines) which we denote by R_c. The columns of R_c are the unit vectors of the vehicle coordinate system expressed in the camera coordinate system. We will assume that the position and orientation of the vehicle relative to the camera coordinate system do not change as the vehicle moves. Thus we will assume that R_c and $\vec{d}_c$ are constant and known.

Given the position $\vec{p}_e$ of a scene point E in the vehicle coordinate system $C\xi\eta\zeta$, its position $\vec{r}_e$ in the camera coordinate system is given by

$$\vec{r}_e = R_c\vec{p}_e + \vec{d}_c$$

Since R_c and $\vec{d}_c$ are constant we have $\dot{\vec{r}}_e = R_c\dot{\vec{p}}_e$. The velocity of E is given by

$$\dot{\vec{r}}_e = -\vec{\omega} \times \vec{r}_e - \vec{T}. \tag{9}$$

In this expression, the rotational velocity is $\vec{\omega} = R_c\vec{\omega}_v$ (see (5)), and the translational velocity is $\vec{T} = R_c\vec{T}_v + \vec{\omega} \times \vec{d}_c$ (see (4)), where $\vec{\omega}_v$ and $\vec{T}_v$ are the rotational and translational velocities of the vehicle coordinate system.

We saw in Section 2.3 that both $\|v\vec{\omega}_d\|$ and $\|\vec{\omega}_{v/d}\|$ are $\mathcal{O}(0.1)$rad/sec. The factors R_c and $R^T_{v/d}$ do not affect the magnitude of either $\vec{\omega}$ or $\vec{T}$. Thus the two components of rotational velocity have comparable magnitudes.

As regards the translational components, note that for normal speeds of the vehicle ($v >$ 10m/sec$\approx$ 22mi/hr), typical suspension elements, and the camera mounted on the vehicle close to the center of mass we have (see Section 2.3) $\|\vec{d}_{v/d}\| = \mathcal{O}(0.025)$m/sec, $\|\dot{\vec{d}}_{v/d}\| = \mathcal{O}(0.1)$m/sec, and $\|\vec{d}_c\| = \mathcal{O}(1)$m/sec. The magnitudes of the translational velocity components are thus $\|v\vec{\omega}_d \times \vec{d}_{v/d}\| \leq v\|\vec{\omega}_d\|\,\|\vec{d}_{v/d}\| = \mathcal{O}(0.0025)$m/sec; $\|v\vec{\mathbf{t}}\| = v = \mathcal{O}(10)$m/sec; and $\|\vec{\omega} \times \vec{d}_c\| \leq \|\vec{\omega}\|\,\|\vec{d}_c\| = \mathcal{O}(0.1)$m/sec. Therefore, the dominant term in the expression for $\vec{T}$ is $v\vec{\mathbf{t}}$ since it is two orders of magnitude larger then any of the other three terms of $\vec{T}$.

2.5 Independently Moving Vehicles

We are interested in other vehicles that are moving nearby. We assume the other vehicles are all moving in the same direction. To facilitate the derivation of the motion equations of a rigid body $\mathcal{B}$ we use two rectangular coordinate frames, one ($Oxyz$) fixed in space, the other ($Cx_1y_1z_1$) fixed in the body and moving with it. The position of the moving frame at any instant is given by the position $\vec{d}_1$ of the origin C_1, and by the nine direction cosines of the axes of the moving frame with respect to the fixed frame. For a given position $\vec{p}$ of P in $Cx_1y_1z_1$ we have the position $\vec{r}_p$ of P in $Oxyz$

$$\vec{r}_p = R\vec{p} + \vec{d}_1 \tag{10}$$

where R is the matrix of the direction cosines (the frames are taken as right-handed so that $\det R = 1$). The velocity $\dot{\vec{r}}_p$ of P in $Oxyz$ is given by

$$\dot{\vec{r}}_p = \vec{\omega}_1 \times (\vec{r}_p - \vec{d}_1) + \dot{\vec{d}}_1. \tag{11}$$

where $\dot{\vec{d}}_1$ is the translational velocity vector and $\vec{\omega}_1 = (\omega_x \ \omega_y \ \omega_z)^T$ is the rotational velocity vector.

From Section 2.3 we have for a typical vehicle $\|\vec{\omega}_1\| = \mathcal{O}(0.1)$rad/sec and $\|\vec{r}_p - \vec{d}_1\| = \mathcal{O}(1)$m. We thus have $\|\vec{\omega}_1 \times (\vec{r}_p - \vec{d}_1)\| \leq \|\vec{\omega}_1\| \, \|\vec{r}_p - \vec{d}_1\| = \mathcal{O}(0.1)\mathcal{O}(1) = \mathcal{O}(0.1)$m/sec. For the translational velocity we have $\|\dot{\vec{d}}_1\| = v$; for normal speeds of the vehicle $v > 10$m/sec $\approx$ 22m/sec so that $\|\dot{\vec{d}}_1\| = \mathcal{O}(10)$m/sec. We can conclude that for any point P on the vehicle the translational velocity is two orders of magnitude larger than the rotational velocity.

If we make the fixed frame $Oxyz$ correspond to the camera frame at time t we have from (9) and (11) that the velocity of a point P on the vehicle expressed in the camera frame is given by

$$\dot{\vec{r}}_p = -\vec{\omega} \times \vec{r}_p + \vec{\omega}_1 \times (\vec{r}_p - \vec{d}_1) - \dot{\vec{T}} + \dot{\vec{d}}_1. \tag{12}$$

In (12) the vector $-\dot{\vec{T}} + \dot{\vec{d}}_1$ corresponds to the relative translational velocity between the camera and the independently moving vehicle. Regarding the first and the second terms on the r.h.s. of (12) we can see that for comparable rotational velocities $\vec{\omega}$ and $\vec{\omega}_1$ the first term will dominate the second term since usually $\|\vec{r}_p - \vec{d}_1\| \ll \|\vec{r}_p\|$. We will use this observation later.

3 Image Motion

3.1 The Imaging Models

Let (X, Y, Z) denote the Cartesian coordinates of a scene point with respect to the fixed camera frame (see Figure 3), and let (x, y) denote the corresponding coordinates in the image plane. The equation of the image plane is $Z = f$, where f is the focal length of the camera. The perspective projection onto this plane is given by

$$x = \frac{fX}{Z}, \quad y = \frac{fY}{Z}. \tag{13}$$

For weak perspective projection we need a reference point (X_c, Y_c, Z_c). A scene point (X, Y, Z) is first projected onto the point (X, Y, Z_c); then, through plane perspective projection, the point (X, Y, Z_c) is projected onto the image point (x, y). The projection equations are then given by

$$x = \frac{X}{Z_c} f, \quad y = \frac{Y}{Z_c} f. \tag{14}$$

3.2 The Image Motion Field and the Optical Flow Field

The instantaneous velocity of the image point (x, y) under perspective projection is obtained by taking the derivatives of (13) and using (9):

$$\dot{x} = \frac{\dot{X}Z - X\dot{Z}}{Z^2} = \frac{-Uf + xW}{Z} + \omega_x \frac{xy}{f} - \omega_y \left(\frac{x^2}{f} + f \right) + \omega_z y, \tag{15}$$

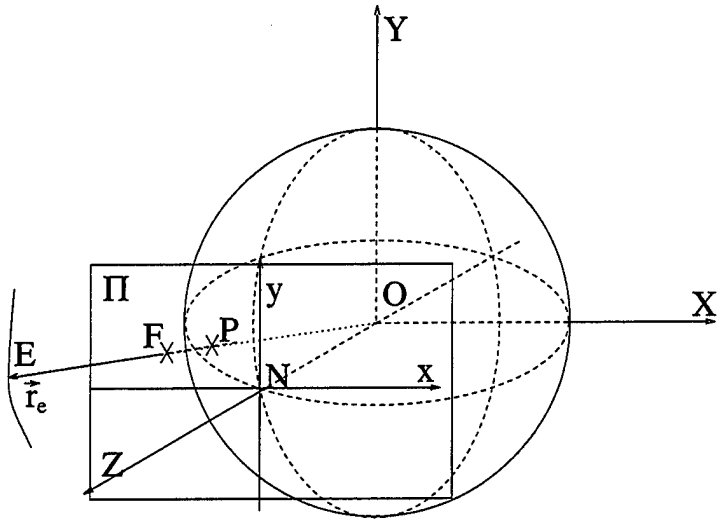


Figure 3: The plane perspective projection image of P is $F = f(X/Z, Y/Z, 1)$; the weak perspective projection image of P is obtained through the plane perspective projection of the intermediate point $P_1 = (X, Y, Z_c)$ and is given by $G = f(X/Z_c, Y/Z_c, 1)$.

$$\dot{y} = \frac{\dot{Y}Z - Y\dot{Z}}{Z^2} = \frac{-Vf + yW}{Z} + \omega_x\left(\frac{y^2}{f} + f\right) - \omega_y\frac{xy}{f} - \omega_z x. \tag{16}$$

The instantaneous velocity of the image point (x, y) under weak perspective projection can be obtained by taking derivatives of (14) with respect to time and using (9):

$$\dot{x} = f\frac{\dot{X}Z_c - X\dot{Z}_c}{Z_c^2} = \frac{-Uf + xW}{Z_c} - f\omega_y\frac{Z}{Z_c} + \omega_z y, \tag{17}$$

$$\dot{y} = f\frac{\dot{Y}Z_c - Y\dot{Z}_c}{Z_c^2} = \frac{-Vf + yW}{Z_c} + f\omega_x\frac{Z}{Z_c} - \omega_z x. \tag{18}$$

Let $\vec{\imath}$ and $\vec{\jmath}$ be the unit vectors in the x and y directions, respectively; $\dot{\vec{r}} = \dot{x}\vec{\imath} + \dot{y}\vec{\jmath}$ is the projected motion field at the point $\vec{r} = x\vec{\imath} + y\vec{\jmath}$. If we choose a unit direction vector $\vec{n}_r$ at the image point $\vec{r}$ and call it the normal direction, then the *normal motion field* at $\vec{r}$ is $\dot{\vec{r}}_n = (\dot{\vec{r}} \cdot \vec{n}_r)\vec{n}_r$. $\vec{n}_r$ can be chosen in various ways; the usual choice (as we shall now see) is the direction of the image intensity gradient.

Let $I(x, y, t)$ be the image intensity function. The time derivative of I can be written as

$$\frac{dI}{dt} = \frac{\partial I}{\partial x}\frac{dx}{dt} + \frac{\partial I}{\partial y}\frac{dy}{dt} + \frac{\partial I}{\partial t} = (I_x\vec{\imath} + I_y\vec{\jmath}) \cdot (\dot{x}\vec{\imath} + \dot{y}\vec{\jmath}) + I_t = \nabla I \cdot \dot{\vec{r}} + I_t$$

where ∇I is the image gradient and the subscripts denote partial derivatives.

If we assume $dI/dt = 0$, i.e. that the image intensity does not vary with time, then we have $\nabla I \cdot \vec{u} + I_t = 0$. The vector field $\vec{u}$ in this expression is called the *optical flow.* If we choose the

normal direction $\vec{n}_r$ to be the image gradient direction, i.e. $\vec{n}_r \equiv \nabla I/\|\nabla I\|$, we then have

$$\vec{u}_n = (\vec{u}\cdot\vec{n}_r)\vec{n}_r = \frac{-I_t\nabla I}{\|\nabla I\|^2} \tag{19}$$

where $\vec{u}_n$ is called the *normal flow*.

It was shown in [31] that the magnitude of the difference between $\vec{u}_n$ and the normal motion field $\dot{\vec{r}}_n$ is inversely proportional to the magnitude of the image gradient. Hence $\dot{\vec{r}}_n \approx \vec{u}_n$ when $\|\nabla I\|$ is large. Equation (19) thus provides an approximate relationship between the 3-D motion and the image derivatives. We will use this approximation later in this paper.

3.3 Estimation of Rotation

We now describe our algorithm for estimating rotation. In this section we give only a brief description of the algorithm. A full description and a proof of correctness will be given in a forthcoming paper. We shall use the following notation: Let I be the image intensity at $\vec{r}$, and let $\vec{n}_r = n_x\vec{\imath} + n_y\vec{\jmath} = \nabla I/\|\nabla I\|$ be the direction of the image intensity gradient at $\vec{r}$. The *normal motion field* at $\vec{r}$ is the projection of the image motion field onto the gradient direction $\vec{n}_r$ and is given by $\dot{\vec{r}}_n = (\dot{\vec{r}}\cdot\vec{n}_r)\vec{n}_r$. From (15-16) we have

$$\begin{aligned}\dot{\vec{r}}_n\cdot\vec{n}_r &= n_x\dot{x} + n_y\dot{y} = \frac{1}{Z}\left[n_x(-Uf + xW) + n_y(-Vf + yW)\right] \\ &+ \left[n_x\frac{xy}{f} + n_y\left(\frac{y^2}{f} + f\right)\right]\omega_x - \left[n_x\left(\frac{x^2}{f} + f\right) + n_y\frac{xy}{f}\right]\omega_y + (n_x y - n_y x)\omega_z\end{aligned} \tag{20}$$

The first term on the r.h.s. of (20) is the translational normal motion $\dot{\vec{r}}_t\cdot\vec{n}_r$ and the remaining three terms are the rotational normal motion $\dot{\vec{r}}_\omega\cdot\vec{n}_r$. From now on we will assume that the camera is forward-looking, i.e. that the *focus of expansion* (*FOE*) is in the image.

The *normal flow* at $\vec{r}$ is defined as $-I_t/\|\nabla I\|$. From [31] we know that the magnitude of the difference between the normal flow field and the normal motion field is inversely proportional to the gradient magnitude; we can thus write

$$\dot{\vec{r}}_n\cdot\vec{n}_r = \dot{\vec{r}}_t\cdot\vec{n}_r + \dot{\vec{r}}_\omega\cdot\vec{n}_r = -\frac{I_t}{\|\nabla I\|} + \mathcal{O}(\|\nabla I\|^{-1}) = \vec{u}_n\cdot\vec{n}_r + \mathcal{O}(\|\nabla I\|^{-1}). \tag{21}$$

If the camera motion is a pure translation the image motion field is a radial pattern; the magnitude of each image motion vector is proportional to the distance of the image point from the *focus of expansion* (*FOE*) and inversely proportional to the depth of the corresponding scene point. If the position $\vec{r}_0 = \vec{\imath}x_0 + \vec{\jmath}y_0$ of the *FOE* is known the translational motion field can be obtained from the translational normal motion field by multiplying each of $\dot{\vec{r}}_t\cdot\vec{n}_r$ by a vector whose direction is $(\vec{r} - \vec{r}_0)$ and whose magnitude is inversely proportional to the angle between the normal flow and the normal motion vector. The translational motion field is then given by

$$\dot{\vec{r}}_t = (\dot{\vec{r}}_t\cdot\vec{n}_r)\frac{\vec{r} - \vec{r}_0}{(\vec{r} - \vec{r}_0)\cdot\vec{n}_r}. \tag{22}$$

Note that if we knew $\vec{\omega}$ we could compute the rotational motion field $\dot{\vec{r}}_\omega$ and the rotational normal motion $\dot{\vec{r}}_\omega \cdot \vec{n}_r$ and use (21) to obtain

$$\dot{\vec{r}}_t \cdot \vec{n}_r \approx \vec{u} \cdot \vec{n}_r - \dot{\vec{r}}_\omega \cdot \vec{n}_r. \tag{23}$$

If we combine (22) and (23) we have

$$\dot{\vec{r}}_t \approx \left(\vec{u} \cdot \vec{n}_r - \dot{\vec{r}}_\omega \cdot \vec{n}_r\right) \frac{(\vec{r} - \vec{r}_0)}{(\vec{r} - \vec{r}_0) \cdot \vec{n}_r}. \tag{24}$$

If the *FOE* is known or can be estimated, we can use (24) to estimate the rotational velocity vector $\vec{\omega}$ by minimizing $\sum_{\vec{r}} \|\dot{\vec{r}}_t\|^2$. Indeed, the image motion field in the neighborhood of the *FOE* is composed of the translational image motion and the rotational image motion. The roll component of the rotational image motion is orthogonal to the translational image motion, so that it increases the magnitude of the image motion field and the normal motion field. The yaw and the pitch components of the rotational image motion are approximately constant in the neighborhood of the *FOE* and just shift the position of the singular (zero) point of the flow field [14]. Furthermore, the rotational normal motion accounts for most of the image motion field at the distant image points [15]. Therefore, if we subtract the rotational image motion, the sum of the magnitudes of the resulting (translational) flow field will be minimal. Using (20) and (24) we then have

$$\vec{\omega} = \arg\min_{\vec{\omega}} \sum_{\vec{r}} \|\vec{u} \cdot \vec{n}_r - \dot{\vec{r}}_\omega \cdot \vec{n}_r\|^2 \frac{\|(\vec{r} - \vec{r}_0)\|^2}{\|(\vec{r} - \vec{r}_0) \cdot \vec{n}_r\|^2}. \tag{25}$$

In matrix form this problem corresponds to minimizing $\|A\vec{\omega} - \vec{b}\|$ (see [29]) where the rows $\vec{a}_i$ of A are given by

$$\vec{a}_i = \frac{\|(\vec{r} - \vec{r}_0)\|}{\|(\vec{r} - \vec{r}_0) \cdot \vec{n}_r\|} \left[n_x \frac{xy}{f} + n_y \left(\frac{y^2}{f} + f\right); \quad -n_x \left(\frac{x^2}{f} + f\right) - n_y \frac{xy}{f}; \quad n_x y - n_y x \right] \tag{26}$$

and the elements b_i of $\vec{b}$ are given by

$$b_i = \vec{u} \cdot \vec{n}_r \frac{\|(\vec{r} - \vec{r}_0)\|}{\|(\vec{r} - \vec{r}_0) \cdot \vec{n}_r\|}.$$

The solution is given by

$$\vec{\omega} = A^+ \vec{b}$$

where A^+ is the generalized inverse of A (see [29]).

3.4 Estimating the FOE

In the case of a forward looking camera and an unknown *FOE* it is possible to simultaneously estimate the *FOE* and $\vec{\omega}$. Based on our analysis in Section 2 we observe that, in the case of a camera rigidly connected to the vehicle and a scene without independently moving objects, the

direction of the translational velocity of the vehicle remains approximately constant throughout the image sequence. The rotational velocity $\vec{\omega}$, however, depends on both the geometry of the road and the activity of the suspension elements; hence, $\vec{\omega}$ changes as the vehicle moves. It is possible to choose a part of the road without horizontal and/or vertical curves and/or changes of lateral slope so that the vehicle rotations correspond to activities of the suspension elements. In these cases the rotational velocity components change in an almost periodic manner. If the scenes are chosen so that significant parts of images correspond to distant scene points the following algorithm can be used to reliably estimate the *FOE*.

Given N successive image frames taken by a forward looking camera mounted on a moving vehicle we form the following function

$$\varphi = \sum_{k=1}^{N} \sum_{i=1}^{N_k} \|\vec{u}_i \cdot \vec{n}_r - \dot{\vec{r}}_{\omega_k} \cdot \vec{n}_r\|^2 \frac{\|(\vec{r}_i - \vec{r}_0)\|^2}{\|(\vec{r}_i - \vec{r}_0) \cdot \vec{n}_r\|^2} \tag{27}$$

where the inside sum is over all normal flow vectors in the kth frame, and the outside sum is over all frames; the position of the *FOE* does not change throughout the sequence (N frames), but $\vec{\omega}$ changes at every frame (hence the index k). We estimate $\vec{\omega}_k, k = 1, \ldots, N$, and $\vec{r}_0$ by minimizing φ.

A straightforward method for minimizing φ corresponds to a nonlinear least squares problem. It can be observed from (27) that φ is linear in the $\vec{\omega}_k$s and nonlinear in $\vec{r}_0$. In the numerical analysis literature such problems are called "separable nonlinear least squares problems" [26]. Several algorithms for solving such problems have been proposed; a unifying feature of these algorithms is that they have better performance than standard nonlinear least squares algorithms [26] since they are based on solving problems of smaller dimensionality that the unseparated problem.

In this paper we use a simple version of the separable algorithm. We choose an initial $\vec{r}_0$ and solve for the $\vec{\omega}_k$ vectors. This problem is equivalent to solving N linear least squares problems as described in Section 3.3. We then have $\vec{\omega}_k = A_k^+ \vec{b}_k$ with the A_ks and the $\vec{b}_k$s appropriately defined. Once the $\vec{\omega}_k$s are estimated we have a nonlinear least squares problem in the two components of $\vec{r}_0$: we need to minimize φ for the fixed $\vec{\omega}_k$s. From (27) we then have

$$\vec{r}_0 = \arg\min_{\vec{r}_0} \sum_{k=1}^{N} \sum_{i=1}^{N_k} \alpha_{k,i}^2 \frac{\|(\vec{r}_i - \vec{r}_0)\|^2}{\|(\vec{r}_i - \vec{r}_0) \cdot \vec{n}_r\|^2} \tag{28}$$

where $\alpha_{k,i} = \|\vec{u}_i \cdot \vec{n}_r - \dot{\vec{r}}_{\omega_k} \cdot \vec{n}_r\|$. We use the Gauss-Newton algorithm to solve this problem. The partial derivatives $\partial\varphi/\partial x_0$ and $\partial\varphi/\partial y_0$ that are required by the Gauss-Newton algorithm are readily obtained from (27-28).

After solving for $\vec{r}_0$ we substitute $\vec{r}_0$ in (27) and solve for the $\vec{\omega}_k$s; we then use these values to solve for $\vec{r}_0$ again. The process is repeated until $\vec{r}_0$ converges. A simple test for convergence is $\|\vec{r}_0(s) - \vec{r}_0(s+1)\| \leq 1$, where s is the iteration number.

3.5 The Image Motion Field of an Observed Vehicle

From (17–18) we obtain the (approximate) equations of projected motion for points on a vehicle under weak perspective:

$$\dot{x} = \frac{Uf - xW}{Z_c}, \tag{29}$$

$$\dot{y} = \frac{Vf - yW}{Z_c}. \tag{30}$$

Equations (29-30) relate the image (projected) motion field to the scaled translational velocity $Z_c^{-1}\vec{T} = Z_c^{-1}(U\ V\ W)^T$.

Given the point $\vec{r} = x\vec{\imath} + y\vec{\jmath}$ and the normal direction $n_x\vec{\imath} + n_y\vec{\jmath}$, from (29-30) the normal motion field $\dot{\vec{r}}_n \cdot \vec{n} = n_x\dot{x} + n_y\dot{y}$ is given by

$$\dot{\vec{r}}_n \cdot \vec{n} = n_x f U Z_c^{-1} + n_y f V Z_c^{-1} - (n_x x + n_y y) W Z_c^{-1} \tag{31}$$

Let

$$\mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} \equiv \begin{pmatrix} n_x f \\ n_y f \\ -n_x x - n_y y \end{pmatrix}, \quad \mathbf{c} = \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} \equiv \begin{pmatrix} U Z_c^{-1} \\ V Z_c^{-1} \\ W Z_c^{-1} \end{pmatrix}. \tag{32}$$

Using (32) we can write (31) as $\dot{\vec{r}}_n \cdot \vec{n} = \mathbf{a}^T\mathbf{c}$. The column vector $\mathbf{a}$ is formed of observable quantities only, while each element of the column vector $\mathbf{c}$ contains quantities which are not directly observable from the images. To estimate $\mathbf{c}$ we need estimates of $\dot{\vec{r}}_n \cdot \vec{n}$ at three or more image points.

3.6 Estimating Vehicle Motion from Normal Flow

As in Section 3.5 we use linear least squares to estimate parameter vector $\mathbf{c}$ from the normal flow.

In the case of a moving vehicle the parameters of interest are the vehicle's trajectory and its *rate of approach*. The rate of approach

$$\nu = \frac{W}{Z_c}$$

(measured in sec^{-1}) is equivalent to the inverse of the *time to collision* and corresponds to the rate with which an object is approaching the camera (or receding from it). The rate $\nu = 0.1/\text{sec}$ means that every second the object travels 0.1 of the distance between the observer and its current position. A negative rate of approach means that the object is going away from the camera.

4 Experiments

In Sections 4.1 and 4.2 we give examples illustrating road detection, stabilization, and vehicle detection. In Section 4.3 we present results for several sequences showing vehicles in motion.

4.1 Road and Vehicle Detection

We detect the road region by finding road edges and lane markers. A Canny edge detector is applied and Hough-like voting is used to detect dominant straight lines in the image. Each line is parameterized by its normal angle α and its displacement d from the center of the image. For all possible values of α and d the image is scanned along the corresponding line. The number of edge points that are found within a strip along the line, and whose gradient direction is orthogonal to the line direction, is taken as the vote for the corresponding point in the α - d plane. Among those lines, only the ones with a certain weight and orientation are chosen to be road line candidates.

If several lines with close α and d values are candidates, only the best representative of these lines is chosen. The other lines are eliminated by applying local maximum suppression in the α - d plane. Note that if the lines represent the two edges of a lane marker, the gradient directions will have opposite signs for the two edges.

Due to perspectivity, the road boundaries and lane marker lines should converge to a single point. Candidate lines that do not converge to a single point are not identified as a road or lane boundaries. The identified lines are the maximal subsets of candidate lines that all intersect at or close to one point (all the intersection points of every pair of the lines are located in a small region).

Figure 4 presents some road detection and vehicle detection results for four different sequences (collected in three different countries).

4.2 Stabilization

As was shown in [15] the impulsive effects introduce significant changes in the roll and pitch angular velocities (see Figure 5). Figure 6 shows three examples of image sequence stabilization by compensating the rotational effects of road bumps. The estimated rotational normal flow component (column c) is subtracted from the total normal flow (column b), which yields the translational normal flow component (column d) One can see that the translational normal flow components at distant points are close to zero.

The vector of rotation is estimated by using the method based on FOE calculation, as described in Section 3.3, or alternatively, by estimating the amount of rotation from the apparent shifts of distant points. Two examples of distant point identification, using horizon points, are shown in Figure 7.

4.3 Relative Motions of Vehicles

After derotating and detecting moving vehicles, we can analyze their motions using the algorithm for motion estimation under weak perspective.

In the first experiment we used an image sequence taken in Haifa, Israel, from a vehicle

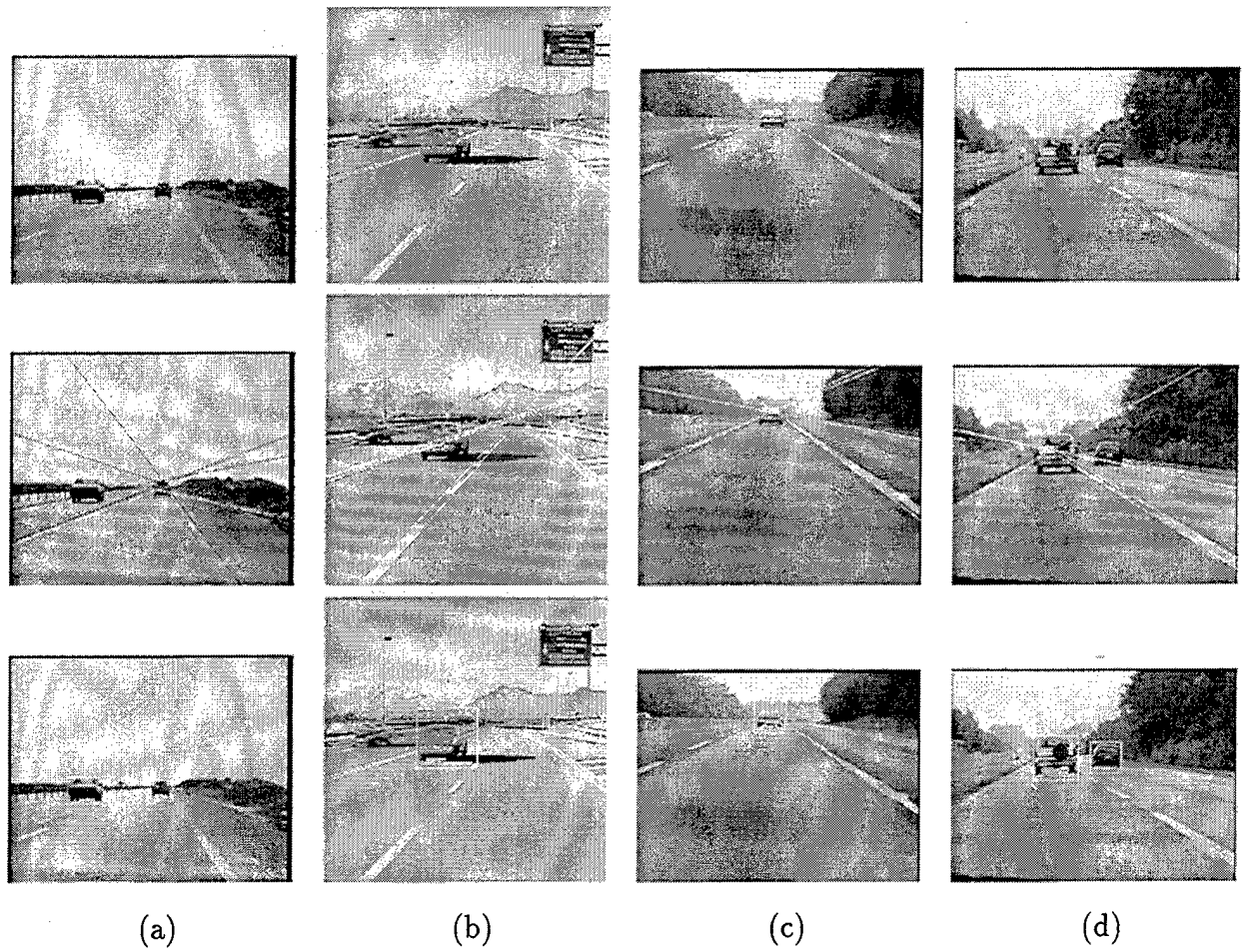
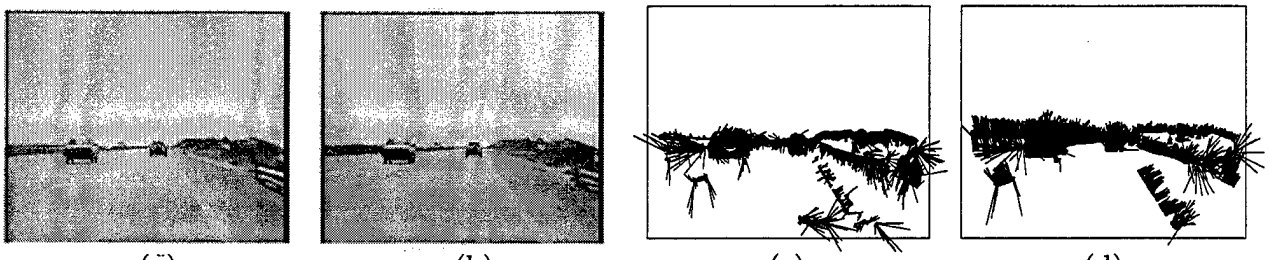


Figure 4: (a-d) A selected frame from each of four sequences. Top: the input images. Middle: results of road detection. Bottom: results of vehicle detection.

following two other accelerating vehicles. The sequence consisted of 90 frames (slightly less than three seconds). Figure 8 shows frames 0, 30 and 60, and the corresponding normal flow on the vehicles. Figure 9 shows estimated values of UZ_c^{-1}, VZ_c^{-1}, and WZ_c^{-1} for the central (closest) vehicle. These values correspond to the translations of the vehicles relative to the vehicle carrying the camera (i.e., in the observer coordinate system). Because of our choice of coordinate system the rate of approach ν is equal to the negative of WZ_c^{-1}.

The graphs show that the motion components have a simple behavior; before they reach their extremal values they can be approximated by straight lines, indicating constant relative accelerations.

In the second experiment we used an image sequence of a van, taken in France, from another vehicle following the van [10, 16]. The sequence consisted of 56 frames (slightly less than two seconds). Figure 10 shows frames 5, 15, 25, and 35 as well as the corresponding normal flow. Figure 11 shows estimated values of UZ_c^{-1}, VZ_c^{-1}, and WZ_c^{-1}. The graph shows that there is an impending collision (rate of approach greater than 1 sec^{-1}). Around the 20th frame the



(a) (b) (c) (d)

Figure 5: (a-b) Two images taken 1/15th of a second apart; (c-d) their normal flows. One can see the effects of bumps. In the first frame, the flow vectors point downward; in the second, they point upward.

rate of approach becomes zero (as do all the velocity components) and after that it becomes negative because the van starts pulling away from the vehicle carrying the camera. A similar image sequence was used in [10] in studies of vehicle convoy behavior.

The third sequence (taken from the IEN Galileo Ferrari Vision Image Library) consisted of 26 frames. Figure 12 shows frames 1, 14 and 26, as well as the corresponding normal flow. Figure 13 shows estimated values of UZ_c^{-1}, VZ_c^{-1}, and WZ_c^{-1}. The graph shows that the W component of the translational velocity is dominant over the U and V components, which is correct for a vehicle that overtakes the observer vehicle and does not change lanes; the two vehicles are moving on parallel courses.

Figure 14 shows frames 1, 26 and 47 of another 48-frame sequence, taken in Haifa, Israel; as well as the corresponding normal flow. Figure 15 shows UZ_c^{-1}, VZ_c^{-1}, and WZ_c^{-1} graphs for the left (overtaking) and central vehicles. One can see that the graphs differ mainly in the values of the W component, since the relative speed of approach for the left vehicle is greater than that for the central one. The U and V components are relatively small; all three vehicles are moving in the same direction.

5 Related Work

5.1 Understanding Object Motion

Some work has been done on understanding object motion, but this work has almost always assumed a stationary viewpoint. Understanding object motion is based on extracting the object's motion parameters from an image sequence. Broida and Chellappa [6] proposed a framework for motion estimation of a vehicle using Kalman filtering. Weng et al. [32] assumed an object that possesses an axis of symmetry, and a constant angular momentum model which constrained the motion over a local frame subsequence to be a superposition of precession and translation. The trajectory of the center of rotation can be approximated by a vector polynomial. Changing the parameters of the model with time allows adaptation to long-term changes in the motion characteristics.

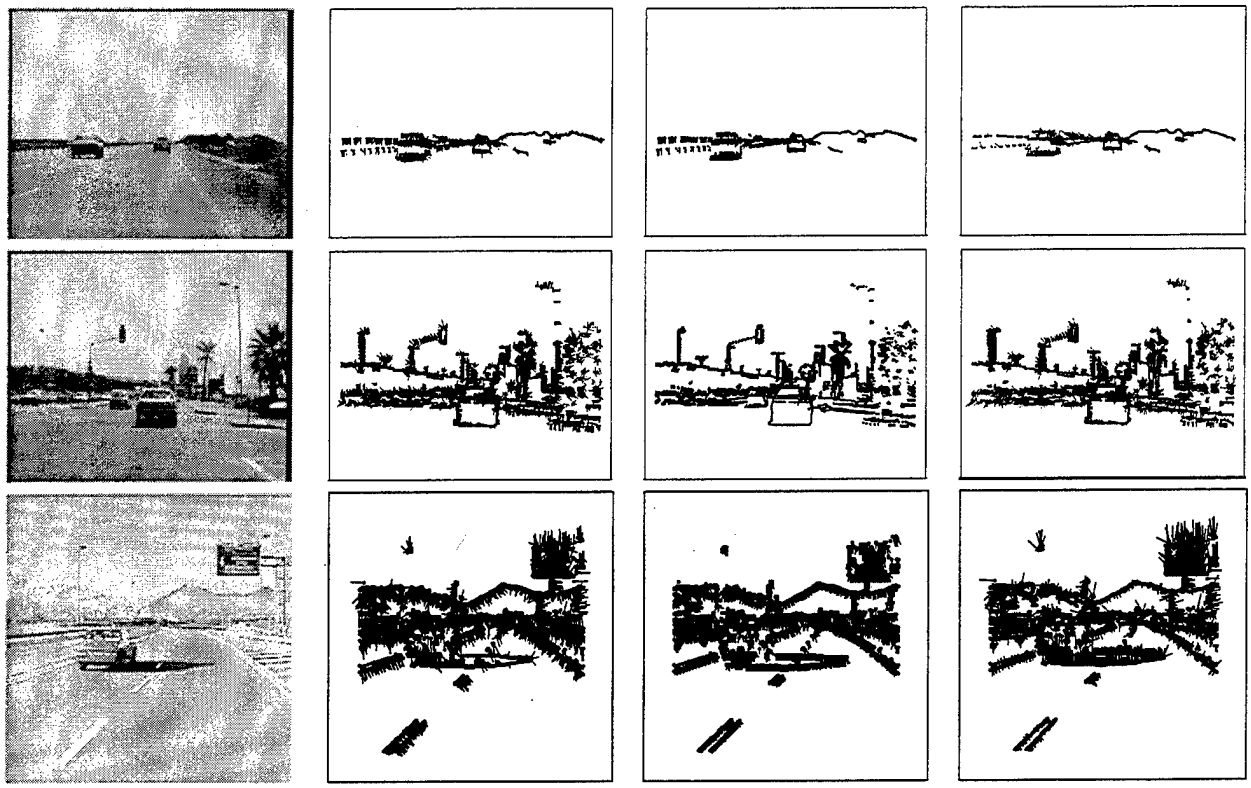
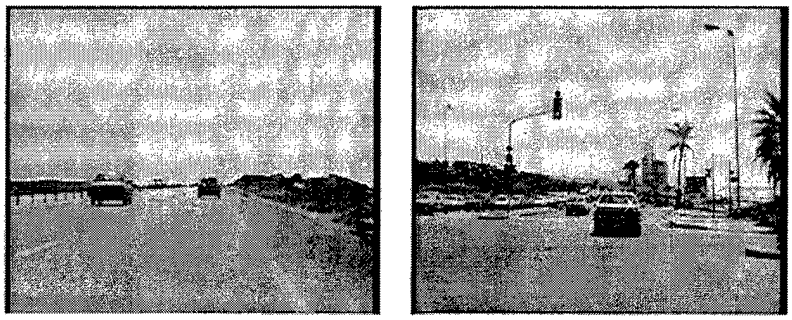


Figure 6: Stabilization results for one frame from each of three sequences: (a) Input frame. (b) Normal flow. (c) Rotational normal flow. (d) Translational normal flow

In [16] Duric et al. tried to understand the motions of objects such as tools and vehicles, based on the fact that the natural axes of the object tend to remain aligned with the local trihedron defined by the object's trajectory. Based on this observation they used the Frenet-Serret motion model, and showed that knowing how the Frenet-Serret frame is changing relative to the observer gives essential information for understanding the object's motion. Our present work is a continuation of this work in a more realistic and complicated scenario, in which the camera is also moving.

5.2 Independent Motion Detection

Much work has been done on detection of independently moving objects by a moving observer. However, the work has been related to detection, classification, and tracking of the motion, and has not paid much attention to motion estimation. Clarke and Zisserman in [11] addressed the problem of independently moving rigid object detection, assuming that all motions (including the camera motion) are pure translations. The idea is to track a set of feature points using correlation and movement assumptions derived from the previous frames, and, based on the feature point correspondences, determine the epipole (FOE) for the background points as the



(a) (b)

Figure 7: Identification of distant image points in frames from two sequences.

intersection of the features' image plane trajectories. The assumption is that a majority of the feature points are background points. Moving objects are found by fitting an epipole to those feature matches that are not consistent with the background epipole. The image plane extent of the moving object is defined by the smallest rectangle enclosing these features. The instability of the camera introduces strong rotational components into the relative motion; these are not dealt with in [11].

Torr and Murray in [30] used statistical methods to detect a non-rigid motion. A five-dimensional space of image pixels and spatio-temporal intensity gradients is fit to an affine transformation model in a least squares sense, while identifying the outliers to the fit using statistical diagnostics. The outliers are spatially clustered to form sets of pixels representing independently moving objects. The assumption again is that the majority of the pixels come from background points and their movement is well approximated by a linear or affine vector field, i.e. the distances to the background points are large compared to the variations in these distances.

Nelson in [23] suggested two qualitative methods for motion detection. The first uses knowledge about observer motion to detect independently moving objects by looking for points whose projected velocity behaviors are not consistent with the constraints imposed by the observer's motion. The second method is used to detect so-called "animate motion", which can be found by looking for violations of the motion field smoothness over time. This is valid in cases where the observer motion changes slowly, while the apparent motion of the moving object changes rapidly.

Irani et al. in [20] used temporal integration to construct a dynamic internal representation image of the tracked object. It is assumed that the motion can be approximated by 2D parametric transformations in the image plane. Given a pair of images, a dominant motion is first computed and the corresponding object is excluded from the region of analysis. This process is then repeated and other objects are found. To track the objects in a long image sequence, integrated images registered with respect to the tracked motion are used.

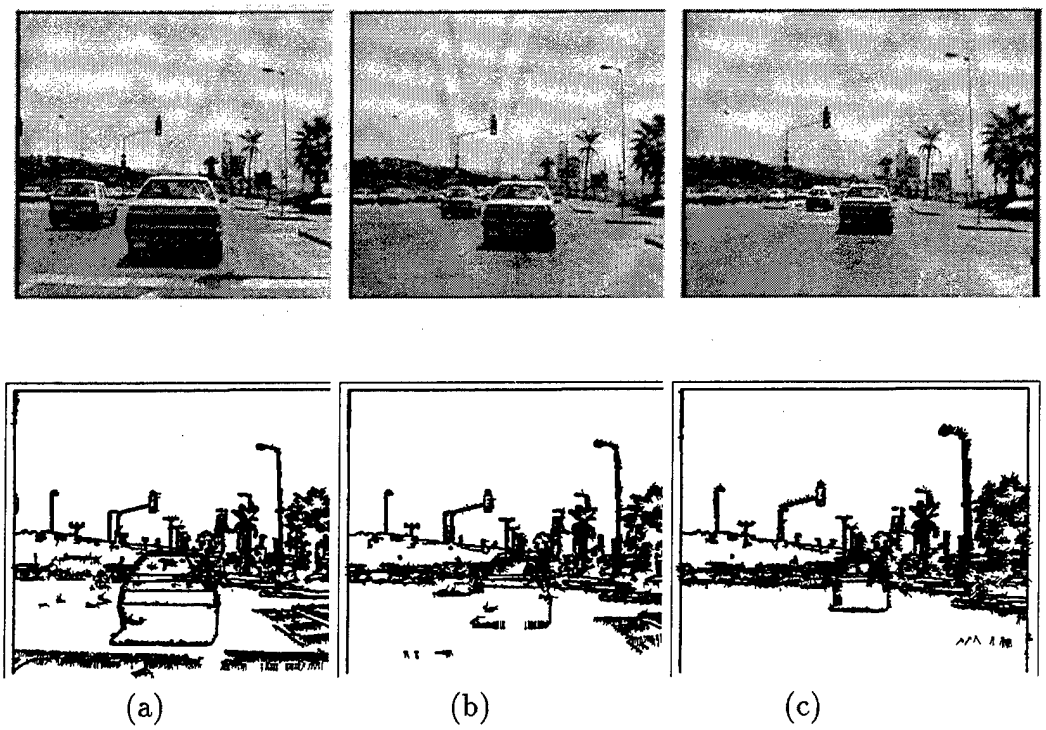


Figure 8: Frames 0, 30, and 60 of a sequence showing two vehicles accelerating. The normal flow results are shown below the corresponding image frames.

Sharma and Aloimonos in [28] demonstrated a method for detecting independent motions of compact objects by an active observer whose motion can be constrained. Fejes and Davis in [18] used constraints on low-dimensional projected components of flow fields to detect independent motion. They implemented a recursive filter to extract directional motion parameters. Independent motion detection was done by a combination of repeated-median-based line-fitting and one-dimensional search. The method was demonstrated on scenes with large amounts of clutter.

5.3 Vehicle Detection and Tracking

In previous work on vehicle detection and tracking by a moving observer, the detection made use of model-based object recognition in single frames. Gil et al. [19] combined multiple motion estimators for vehicle tracking. Vehicle detection was performed using two features: the bounding rectangle of the moving vehicle, where the convex hull of the vehicle is computed for every frame and then translated according to the predicted motion parameters, and an updated 2-D pattern (gray-level mask) based on optimization of the correlation between the pattern and

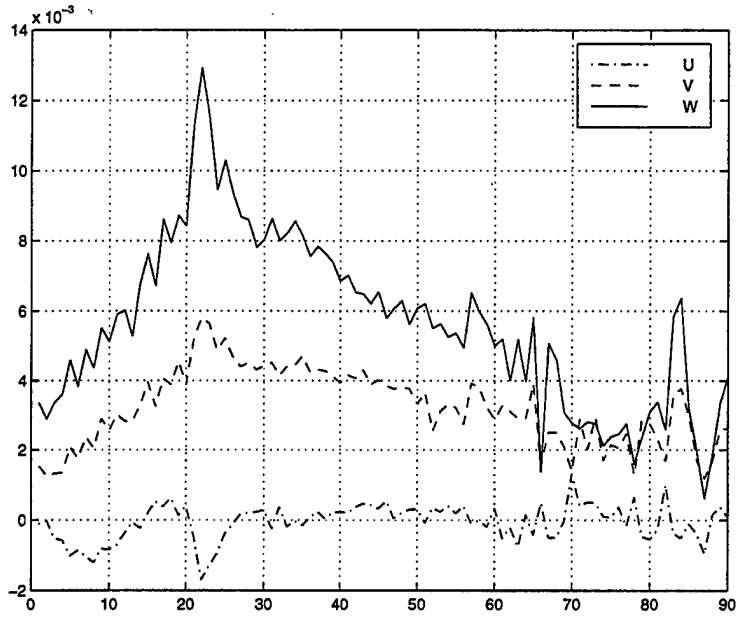


Figure 9: Motion analysis results for the acceleration sequence. U, V, W are the scaled (by an unknown distance Z_c^{-1}) components of the relative translational velocity .

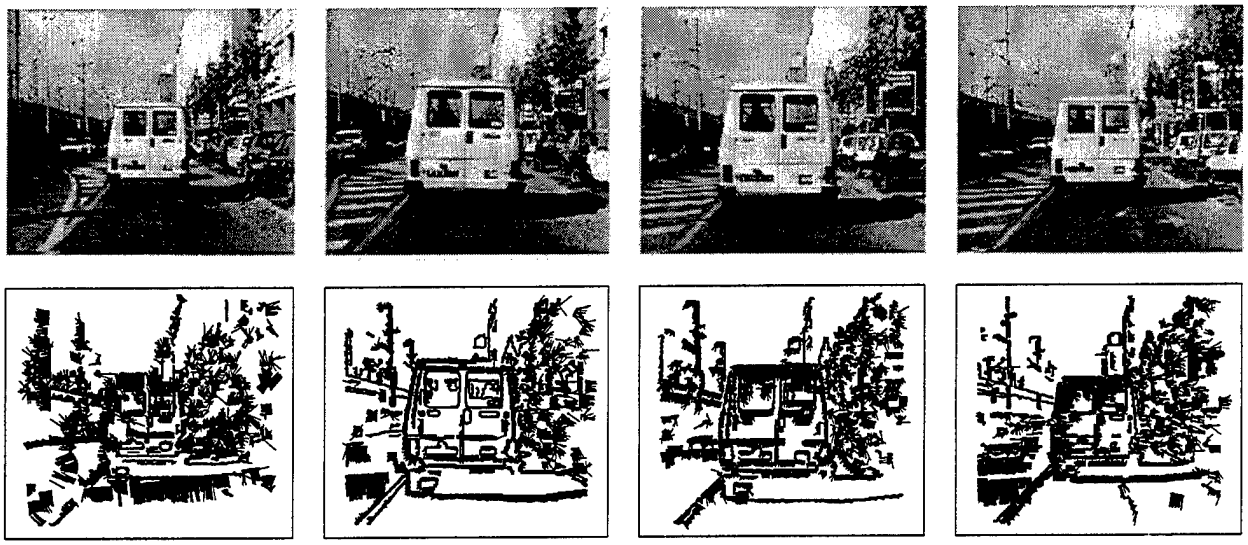
the image using the motion parameters. These results were obtained using a stationary camera mounted above a highway under different road and illumination conditions.

Betke et al. [3] developed a real-time system for detection and tracking of multiple vehicles in a frame sequence taken on a highway from a moving vehicle. The system distinguishes between distant and passing vehicle detection. In case of a passing vehicle the recognition is performed by detecting large brightness differences over small numbers of frames. 2-D car models are used to create a gray-scale template of the detected vehicle for future tracking. Distant vehicles are detected by analyzing prominent horizontal and vertical edges. A square region bounded by such edges, which is strongly enough correlated with a vehicle template, is recognized as a vehicle. For each newly recognized vehicle a separate tracking process is allocated by a real-time operating system, which tracks the vehicle until it disappears and makes sure that no other process tracks the same vehicle. When one vehicle occludes another, one of the tracking processes terminates and the other tracks the occlusion region as a single moving object.

5.4 Road Detection and Shape Analysis

Our work also involved detection of the road markings; in doing this we made no significant use of motion models. There has been considerable work on road boundary detection in images obtained by a vehicle driving on the road. A typical detection procedure involves two steps. First the road is detected in a single frame (or in a small number of frames), and then the road boundaries are tracked in a long sequence of frames.

Schneiderman and Nashman in [27] left the first step to a human operator and gave a solution



(a) (b) (c) (d)

Figure 10: Frames 5, 15, 25, and 35 of the van sequence. The normal flow results are shown below the corresponding image frames.

for the second step alone. Their algorithm is based on road markings. In each frame, edge points are extracted and grouped into clusters representing individual lane markers. The lane marker models built in the first step are updated using the information obtained from successive frames. The markers are modeled by second-order polynomials in the image plane. Tracking is exploited in the sense that edge points are associated with markers based on a road model formed from analysis of the previous frames. Spatial proximity and gradient direction are used as a clues for clustering. The marker models are updated to satisfy a least squares criterion of optimality.

Broggi in [7]-[9] presented a parallel real-time algorithm for road boundary detection. First the image undergoes an inverse perspective transformation and then road markings are detected using simple morphological filters that work in parallel on different image areas. The inverse perspective procedure is based on a priori knowledge of the imaging process (camera position, orientation, optics, etc.).

Richter and Wetzel in [25] used region segmentation for road surface recognition. A road model is adjusted to the segmentation results collected from several frames. The adjusted model is then used for model-based segmentation and tracking. A special lookup mechanism is used to overcome the problem of obstacles and shadows that may cause the segmentation algorithm to fail to detect the road regions.

A sophisticated road model that takes into account both horizontal and vertical curvature was suggested by Dickmanns in [12]. A differential-geometric road representation and a spatio-temporal driving model are exploited to find the parameters of the road and the vehicle.

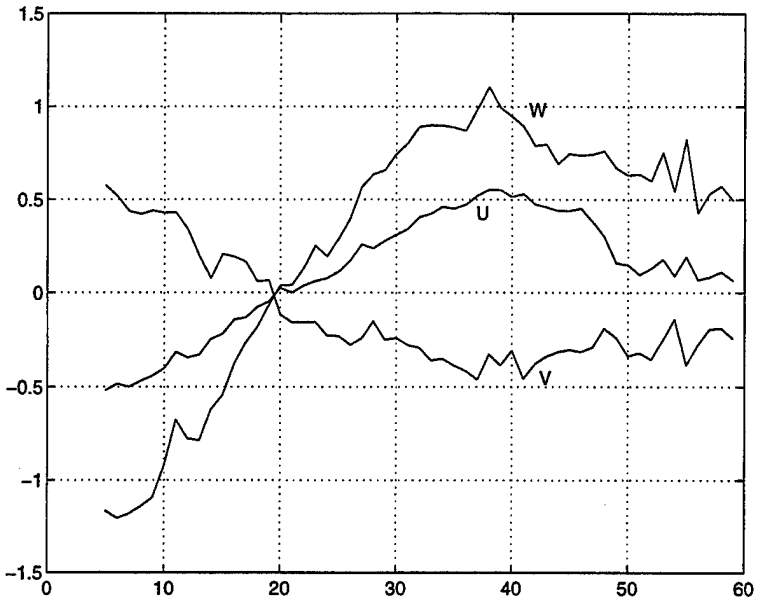


Figure 11: Motion analysis results for the van sequence. U, V, W are the scaled (by an unknown distance Z_c^{-1}) components of the relative translational velocity .

6 Conclusions and Plans for Future Work

Understanding the motions of vehicles from images taken by a moving observer requires a mathematical formulation of the relationships between the observer's motion and the image motion field, as well as a model for the other vehicles' trajectories and their contributions to the image motion field. In this paper a constant relationship between each vehicle's frame and the observer's frame is assumed. The use of the Darboux frame provides a vocabulary appropriate for describing long motion sequences.

We have derived equations for understanding the relative motions of vehicles in traffic scenes from a sequence of images taken by a moving vehicle. We use the Darboux motion model for both the observing vehicle and the nearby moving vehicles. Using a full perspective imaging model we stabilize the observer sequence so that our model for the observed vehicles' motions can be applied. Using the weak perspective approximation we analyze the nearby vehicles' motions and apply this analysis to long image sequences. Expanding our analysis to various classes of traffic events [21], and to articulated vehicles, are directions for future research.

References

[1] American Association of State Highway Officials. *A Policy on Geometric Design of Rural Highways* (9th edition). AASHO, 1977.

[2] M. G. Bekker. *Introduction to Terrain-Vehicle Systems.* University of Michigan Press, Ann Arbor, 1969.

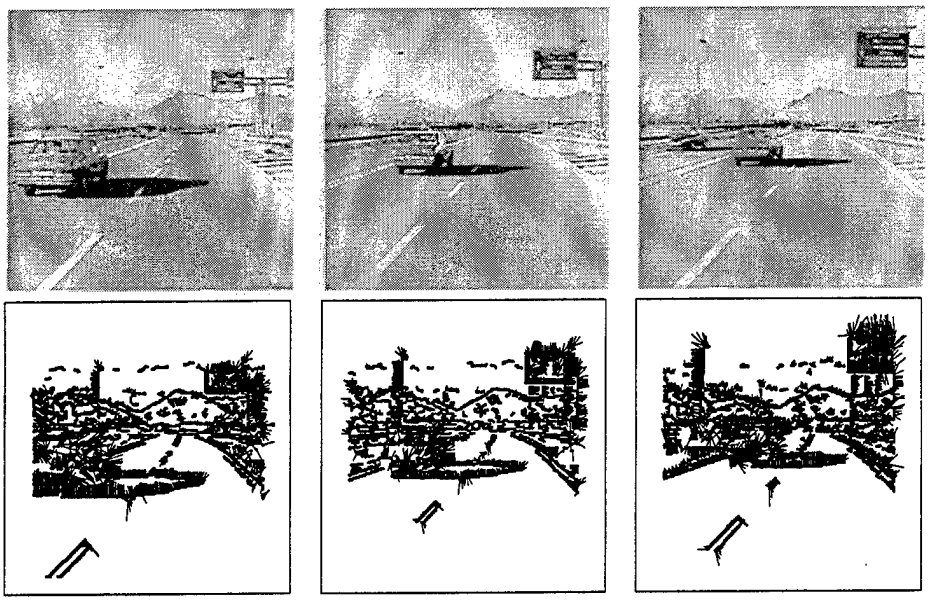


Figure 12: Frames 1, 14 and 26 of the Italian sequence. The normal flow results are shown below the corresponding image frames.

[3] M. Betke, E. Haritaoglu, and L. S. Davis. *Multiple Vehicle Detection and Tracking in Hard Real Time.* Technical Report CS-TR-3667, Computer Vision Laboratory, Center for Automation Research, University of Maryland, College Park, 1996.

[4] M. L. Boas. *Mathematical Methods in the Physical Sciences* (2nd edition). Wiley, New York, 1983.

[5] O. Bottema and B. Roth. *Theoretical Kinematics.* North–Holland, Amsterdam, 1979.

[6] T. J. Broida and R. Chellappa. Estimation of object motion parameters from noisy images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8:90–99, 1986.

[7] A. Broggi. An image reorganization procedure for automotive road following systems. In *Proc. International Conference on Image Processing*, Vol. 3, pages 532-535, 1995.

[8] A. Broggi. Robust real-time lane and road detection in critical shadow conditions. In *Proc. International Symposium on Computer Vision*, pages 353-358, 1995.

[9] A. Broggi. Parallel and local feature extraction: A real-time approach to road boundary detection. *IEEE Transactions on Image Processing*, 4:217-223, 1995.

[10] P. Burlina and R. Chellappa. Time-to-X: Analysis of motion through temporal parameters, In *Proc. Conference on Computer Vision and Pattern Recognition*, pages 461-468, 1994.

[11] J.C. Clarke and A. Zisserman. Detection and tracking of independent motion. *Image and Vision Computing*, 14:565–572, 1996.

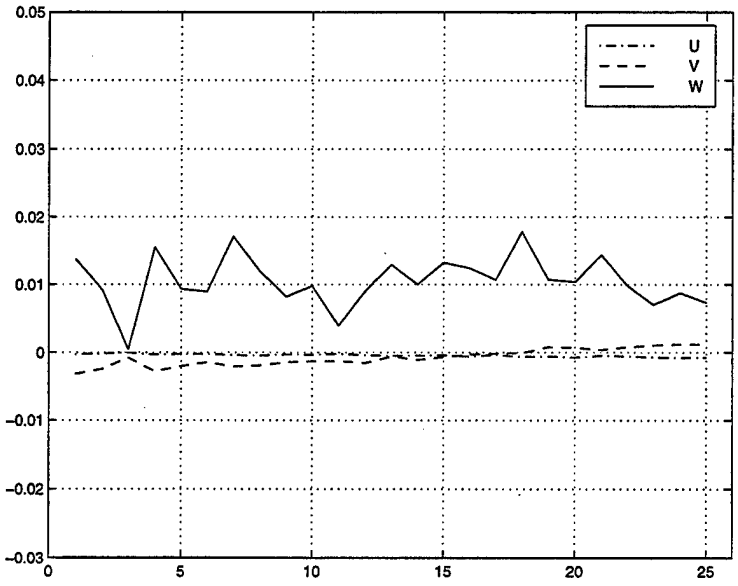


Figure 13: Motion analysis results for the Italian sequence. U, V, W are the scaled (by an unknown distance Z_c^{-1}) components of the relative translational velocity .

[12] E. Dickmanns and B. Mysliwetz. Recursive 3-D road and relative ego-state recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13:149-213, 1992.

[13] Z. Duric, A. Rosenfeld, and J. Duncan. The applicability of Green's theorem to computation of rate of approach. In *Proc. ARPA Image Understanding Workshop*, pages 1209–1217, 1994.

[14] Z. Duric, A. Rosenfeld, and L. S. Davis. Egomotion analysis based on the Frenet-Serret motion model. *International Journal of Computer Vision*, 15:105–122, 1995.

[15] Z. Duric, and A. Rosenfeld. *Stabilization of Image Sequences.* Technical Report CAR–TR–778, Computer Vision Laboratory, Center for Automation Research, University of Maryland, College Park, 1995.

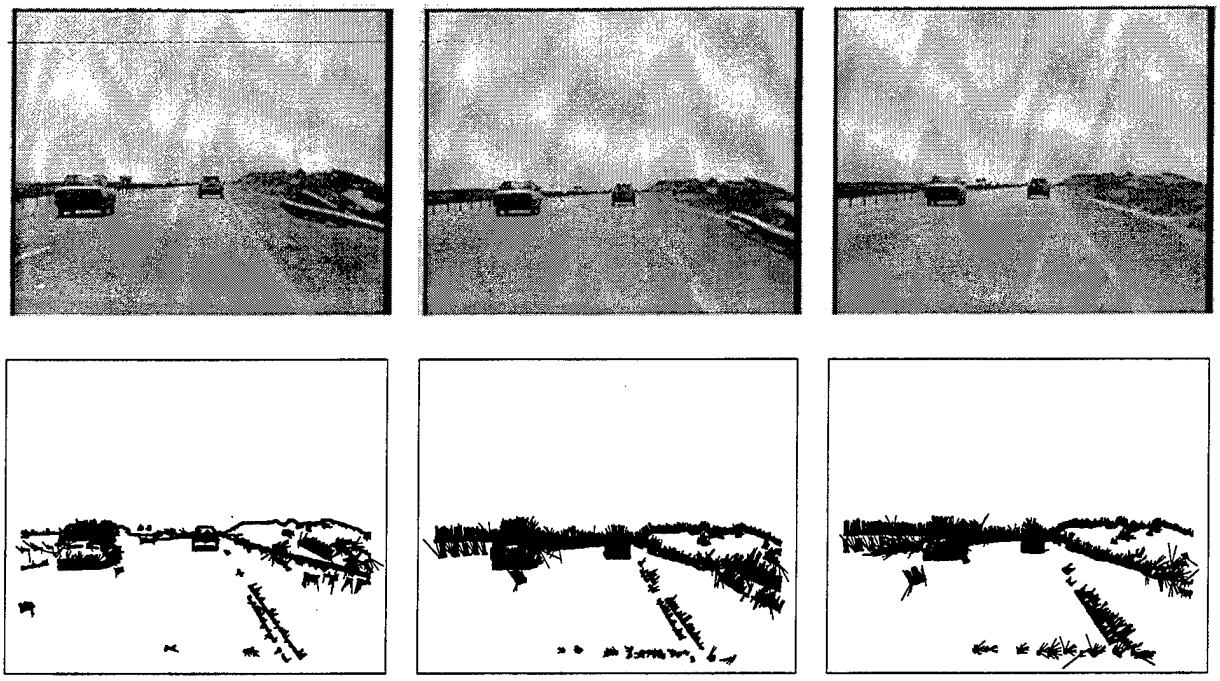
[16] Z. Duric, E. Rivlin and A. Rosenfeld. Understanding Object Motion. In *Proc. International Conference on Computer Vision*, pages 925-932, 1998.

[17] J. R. Ellis. *Vehicle Handling Dynamics.* Mechanical Engineering Publications Limited, London, 1994.

[18] S. Fejes and L.S. Davis. Detection of independent motion using directional motion estimation. In *Proc. International Conference of Computer Vision*, pages 979-986, 1998.

[19] S. Gil, R. Milanese and T. Pun. Combining multiple motion estimates for vehicle tracking, In *Proc. European Conference on Computer Vision*, Vol. 2, pages 307-320, 1996.

[20] M. Irani, B. Rousso, and S. Peleg. Detecting and tracking multiple moving objects using temporal integration. In *Proc. European Conference on Computer Vision*, pages 282–287, 1992.



(a) (b) (c)

Figure 14: Frames 1, 25 and 48 of the Haifa sequence. The normal flow results are shown below the corresponding image frames.

[21] D. Koller, H. Heinze, and H.H. Nagel. Algorithmic characterization of vehicle trajectories from image sequences by motion verbs. In *Proc. Conference on Computer Vision and Pattern Recognition*, pages 90–95, 1991.

[22] E. Kreyszig. *Differential Geometry.* University of Toronto Press, Toronto, Canada, 1959.

[23] R. Nelson. Qualitative detection of motion by a moving observer *International Journal of Computer Vision*, 7:33–46, 1991.

[24] R. Oldenburger. *Mathematical Engineering Analysis.* Macmillan, New York, 1950.

[25] S. Richter and D. Wetzel. A robust and stable road model. In *Proc. International Conference on Pattern Recognition*, Vol. 1, pages 777-780, 1994.

[26] A. Ruhe and A. Wedin. Algorithms for separable nonlinear least squares problems. *SIAM Review*, 22:318–337, 1980.

[27] H. Schneiderman, and M. Nashman. A discriminating feature tracker for vision based autonomous driving. *IEEE Transactions on Robotics and Automation*, 10:769-775, 1994.

[28] R. Sharma and Y. Aloimonos. Early detection of independent motion from active control of normal image flow patterns. *IEEE Transactions on Systems, Man, and Cybernetics B*, 26:42-52, 1996.

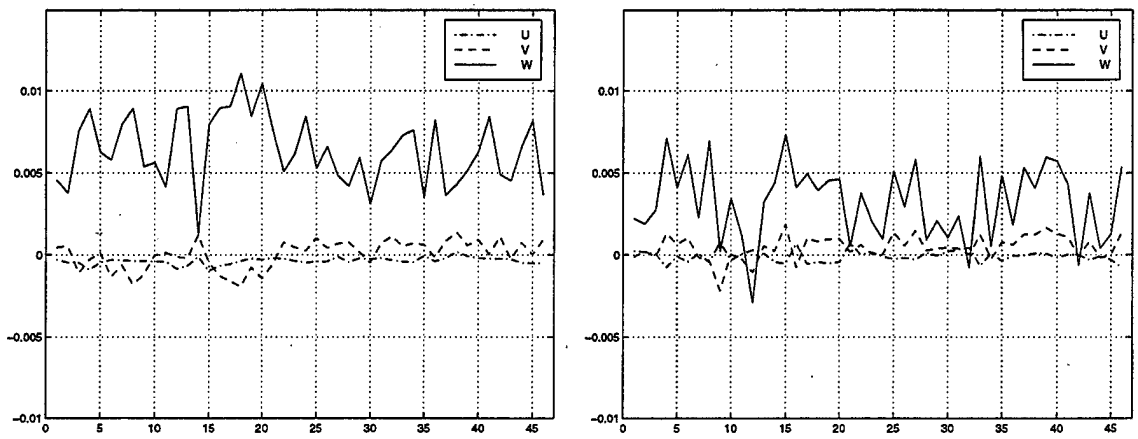


Figure 15: Motion analysis results for the Haifa sequence. U, V, W are the scaled (by an unknown distance Z_c^{-1}) components of the relative translational velocity .

[29] G. W. Stewart. *Introduction to Matrix Computations.* Academic Press, New York, 1973.

[30] P. Torr and D. Murray. Statistical detection of independent movement from a moving camera. *Image and Vision Computing*, 11:180–187, 1993.

[31] A. Verri and T. Poggio. Against quantitative optical flow. In *Proc. International Conference on Computer Vision*, pages 171–180, 1987.

[32] J. Weng, T. Huang, and N. Ahuja. 3-D motion estimation, understanding, and prediction from noisy image sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9:370–389, 1987.

[33] J. H. Williams, Jr. *Fundamentals of Applied Dynamics.* Wiley, New York, 1996.

[34] J. Y. Wong. *Theory of Ground Vehicles* (2nd edition). Wiley, New York, 1993.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	March 1998	Technical Report

4. TITLE AND SUBTITLE

Estimating Relative Vehicle Motions in Traffic Scenes

5. FUNDING NUMBERS

N00014-95-0521

6. AUTHOR(S)

Zoran Duric, Roman Goldenberg, Ehud Rivlin and Azriel Rosenfeld

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Center for Automation Research
University of Maryland
College Park, MD 20742-3275

8. PERFORMING ORGANIZATION REPORT NUMBER

CAR-TR-881
CS-TR-3882

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Office of Naval Research
800 North Quincy Street, Arlington, VA 22217-5660

Advanced Research Projects Agency
3701 North Fairfax Drive, Arlington, VA 22203-1714

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release.
Distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

Autonomous operation of a vehicle on a road calls for understanding of various events involving the motions of the vehicles in its vicinity. In this paper we show how a moving vehicle which is carrying a camera can estimate the relative motions of nearby vehicles. We present a model for the motion of the observing vehicle, and show how to "stabilize" it, i.e. to correct the image sequence so that transient motions resulting from bumps, etc. are removed and the sequence corresponds more closely to the sequence that would have been collected if the motion had been smooth. We also model the motions of nearby vehicles and show how to detect their motions relative to the observing vehicle. We present results for several road image sequences which demonstrate the effectiveness of our approach.

14. SUBJECT TERMS

15. NUMBER OF PAGES

30

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL