Visual Surveillance for Moving Vehicles

JAMES M. FERRYMAN, STEPHEN J. MAYBANK AND ANTHONY D. WORRALL
Computational Vision Group, Department of Computer Science,
The University of Reading, Berkshire RG6 6AY, England, UK
j.m.ferryman@reading.ac.uk

Abstract. An overview is given of a vision system for locating, recognising and tracking multiple vehicles, using an image sequence taken by a single camera mounted on a moving vehicle. The camera motion is estimated by matching features on the ground plane from one image to the next. Vehicle detection and hypothesis generation are performed using template correlation and a 3D wire frame model of the vehicle is fitted to the image. Once detected and identified, vehicles are tracked using dynamic filtering. A separate batch mode filter obtains the 3D trajectories of nearby vehicles over an extended time. Results are shown for a motorway image sequence.

Keywords: model-based vision, surveillance, traffic scene analysis, vehicle tracking, filtering

1. Introduction

In today’s crowded traffic conditions the cost of an accident can be heavy both in human terms and in its effect on a road network working at or near to full capacity. Any reduction in the number or in the severity of accidents would have large social and economic benefits. Recent work in computer vision may provide the means for such a reduction. With the rapidly falling costs of cameras and computers it may soon be feasible to equip a car with an autonomous artificial vision system for monitoring nearby vehicles, keeping track of their motions and warning the driver whenever there is danger of a collision.

This paper describes a laboratory based vision system for locating, recognising and tracking multiple vehicles, using an image sequence taken by a single camera mounted on a moving vehicle.

1.1. Previous Work

The Intelligent Vehicle Symposia, 1994–1996, describe a number of vision systems for detecting and tracking road lane markings, vehicles and obstacles using monocular, binocular or trinocular vision. Existing methods for motion estimation and tracking typically employ optical flow or other 2-D image based techniques e.g. Bertozzi and Broggi (1996), Betke et al. (1996), Giachetti et al. (1994), Matthews et al. (1995) and Smith (1995). These approaches are adequate for tasks such as the estimation of time to collision or object detection but (i) they cope poorly with overlapping vehicles, shadows, rain etc; and (ii) in the absence of a model they are not accurate enough to generate an understanding of the 3D forms and positions of vehicles.

The first vision systems for monitoring road traffic relied on static cameras (Beymer and Malik, 1996; Koller et al., 1994; Sullivan, 1992, 1994). The system described in Sullivan (1992, 1994) can locate and track vehicles in 3D as they move across the ground plane, classify their trajectories and take account of occlusions of vehicles by stationary parts of the scene or occlusions between vehicles.

More recent work has included the application of model-based methods to imagery collected from moving cameras (Dickmanns and Myśliwetz, 1992; Dickmanns, 1997; Thomae et al., 1992; Weber, 1995; Zhao and Thorpe, 1998). In the CMU NavLab project, Dellal et al. (1998) track a vehicle using a...
single vehicle mounted camera. They model the target vehicle by a box, and have a one large filter to estimate the parameters they need, including the position and dimensions of the target vehicle, the curvature and width of the lane, the motion of the test vehicle and part of the camera calibration. Luong et al. (1995) and Bertozzi and Broggi (1996) and Weber et al. (1995) describe vehicle mounted stereo systems for tracking road lanes and detecting obstacles. In Luong et al. (1995) the distance to the vehicle in front is monitored as well. A review of Dickmanns work can be found in Dickmanns (1997).

1.2. This Paper

New results are reported from a system for the model based tracking of road vehicles. The use of vehicle models has many advantages. Model-based algorithms are robust to variations in illumination and colour, and to changes in the position of the object relative to the camera. They deal properly with occlusions and make it possible to impose object-centered dynamic constraints. The algorithms can be easily extended to deal with camera movement and to utilise the information from several cameras. The model gives the tracker a level of stability and accuracy that is impossible to achieve in 2D methods of tracking.

Section 2 describes methods for estimating the camera motion (egomotion) from point matches and for detecting far-off target vehicles. Section 3 describes the model-based techniques for inferring 3D trajectories of vehicles from an image sequence and Section 4 describes a batch mode filter for estimating vehicle trajectories and predicting their future motions of vehicles.

1.3. System Overview

The main sources of information for the system are a monocular sequence of video images, camera calibration, knowledge of the road markings and wire frame models of vehicle geometry. It is assumed throughout that vehicles are subject to the Ground Plane Constraint (GPC), i.e. vehicles are confined to the ground plane and move with three degrees of freedom, namely position \((x, y)\) on the ground plane, and orientation \(\theta\) about an axis normal to the ground plane.

A world coordinate frame is defined, with the origin in the ground plane, \(x\) axis transverse to the road lanes, \(y\) axis along the road lanes and the \(z\) axis vertically upward. The \(x\) axis is directed from the outer edge of the road (hard shoulder) inwards, giving a right handed coordinate frame in the UK where vehicles are driven on the left.

The structure of the system is shown in Fig. 1. The Egomotion Estimation module estimates the position and translational velocity of the host vehicle. These estimates are smoothed and supplied to the Model-based Instantiation Module and the Model-based Tracker. At instantiation a wire frame vehicle model is chosen. In the Model-based Tracker the chosen wire frame model is fitted to the image and used to track the vehicle through the image sequence. The tracker uses the Dynamic Filter to predict the position of the vehicle in the next, and thus reduce the cost of matching the vehicle model to the next image. The Collision Alert module inputs from the Model-based Tracker the measured positions and orientations of a nearby vehicle and uses a batch filter to fit a path to the measurements over an interval of several seconds. Batch filtering is essential for obtaining the good estimates of vehicle trajectories needed to warn the driver of impending collisions.

2. Egomotion Estimation

The estimation of the position and velocity of the camera is simplified by a prior off-line calibration to determine the internal orientation of the camera and its position relative to the ground plane. The camera calibration is obtained interactively, using a rectangular grid which is adjusted in 3D until the projections of the grid lines coincide with the motorway lane markings.