Dickmanns, E.D. 1997: ***Vehicles Capable of Dynamic Vision.*** Proc. 15th International Joint Conference on Artificial Intelligence (IJCAI-97), Vol. 2, Nagoya, Japan, pp 1577-1592

Abstract

A survey is given on two decades of developments in the field, encompassing an increase in computing power by four orders of magnitude. The '4-D approach' integrating expectation-based methods from systems dynamics and control engineering with methods from AI has allowed to create vehicles with unprecedented capabilities in the technical realm: Autonomous road vehicle guidance in public traffic on freeways at speeds beyond 130 km/h, on-board-autonomous landing approaches of aircraft, and landmark navigation for AGV's, for road vehicles including turn-offs onto cross-roads, and for helicopters in low-level flight (real-time, hardware-in-the-loop simulations in the latter case).

1 Introduction

Road vehicle guidance based on video-signal processing has been picked up independently in Japan [Tsugawa et al., 1979], in Europe [Meissner, 1982], and in the USA [Klass, 1985]. While in Japan analog signal processing has been used and (quasi-steady) AI-methods predominated in the US, recursive estimation methods well known from systems engineering have been extended to image sequence processing at the author's institute (UBM); the resulting method had been dubbed '4-D approach', in contrast to the 2-D, 2.5-D, and 3-D methods under discussion then, disregarding time as the fourth independent variable in the problem domain. The numerical efficiency and compactness in state representation of recursive estimation which directly allowed control applications for generating behavioral capabilities, finally, led to its wide-spread acceptance in the vision community. Artificial neural nets (ANN) also found wide acceptance in the USA [Pomerleau, 1992] and around the globe even though image resolution used (about 1000 pixel = 1Kpel), usually, was much less than with recursive estimation (80 Kpel per image, even at a higher image rate).

Both methods allowed road vehicles to run autonomously along highways and other types of roads up to rather high speeds, initially on empty roads only [Dickmanns and Zapp, 1987, Pomerleau, 1989] but finally in normal freeway traffic also [Dickmanns et al., 1994, Pomerleau, 1992]; however, while ANN's stayed confined to either lateral [Pomerleau, 1992; Mecklenburg et al., 1992] or longitudinal control [Fritz, 1996] at a time (the other mode had to be controlled by a human driver), the 4-D approach allowed to detect, track and determine the spatio-temporal state (position and velocity components on a 3-D surface) relative to about a dozen other objects in a range of up to 100 meters in front of and behind the own vehicle pickmanns, 1995a]. The two final demonstrator vehicles in the European project Prometheus: VITA_2 of Daimler-Benz and VaMP of UBM [Ulmer, 1994; Dickmanns et al., 1994], may well be considered as the first two road vehicles of a new species capable of understanding (part of) their environment and of reacting properly to the actual needs on their own (completely autonomous).

Dynamic remote sensing for intelligent motion control in an environment with rapidly changing elements requires the use of valid spatio-temporal models for efficient handling of the large data streams involved. Other objects have to be recognized with their relative motion components, the near ones even with high precision for collision avoidance; this has to be achieved while the own vehicle body carrying the cameras moves in an intended way and is, simultaneously, subject to perturbations hardly predictable.

For this complex scenario, inertial sensing in addition to vision is of great help; negative angular rate feedback to a viewing direction control device allows to stabilize the appearance of stationary objects in the image sequence. Measured accelerations and velocities will, via signal integration, yield predictions for.translational and rotational positions affecting the perspective mapping process. These predictions are good in the short run, but may drift slowly in the long run, especially when inexpensive inertial sensors are used. These drifts, however, can easily be compensated by visual interpretation of static scene elements.