Ernst D. Dickmanns, Prof. Dr.-Ing.
Universitaet der Bundeswehr Munich
Department of Aero-Space Technology
Institut fuer Systemdynamik und Flugmechanik
D-85577 Neubiberg, Germany

ARO-Workshop December 2000, Canberra, Australia; talk on:

The advantages of EMS-Vision for mission performance of autonomous ground vehicles

Extended Abstract: Expectation-based, Multi-focal, Saccadic (EMS-) Vision has been designed to cope with many different aspects of mission performance. A wide field of view (f.o.v., >~100°) nearby, realized by two wide-angle cameras with divergent optical axes, allows to detect and avoid obstacles at low speed and to negotiate tight curves. Trinocular stereo vision in a small central f.o.v. yields depth estimations in the near range from just one single well recognizable feature. The additional tele-camera covering the vertical center of the area of overlap between the two wide-angle cameras is a high-resolution 3-chip-color camera; its focal length is 3 to 4 times the one of the two wide-angle cameras. Active gaze control allows shifting this central f.o.v. to where it is needed, and to inertially stabilize the viewing direction for eliminating motion blur; a fourth camera with a strong tele-lens may be added for increased resolution. Search and smooth pursuit of objects with the tele-cameras are modes of operation.
This provides advantages for detection and recognition of objects or landmarks far away; a factor of ten in focal length as compared to the wide-angle lenses has been realized. This combination of cameras mounted fix relative to each other on a single pan & tilt platform acts like a vertebrate eye and requires quick gaze control for achieving good resolution in areas which show interesting features in the wide f.o.v. This concept trades two orders of magnitude in video data rate for a short time delay (a few tenths of a second) until high resolution of this part of the scene is available (like in the vertebrate eye). According to the major constituents: **M**ulti-focal camera set, **a**ctive/**r**eactive gaze control relative to the **V**ehicle carrying it, it has been dubbed '**MarVEye**'.

This concept requires distributed processors for handling the huge video data stream (up to 55 MB/s); three commercial-off-the-shelf (COTS-) dual processor industrial PC systems with multiple frame-grabbing capability have been selected for image sequence processing at video rate (25 Hz or 40 ms cycle time). The system architecture will be discussed in the talk. A fourth dual processor system takes care of mission performance, situation assessment, and decision making for control actuation on the upper system level. Measurement data input and actual control output to the vehicle is done by a separate transputer system remaining from the previous generation of processors.
Scene representation is done in a dynamic scene tree exploiting homogeneous coordinate transformations like in computer graphics; however, in computer vision, the variables entering these transformations and those describing the shape of the objects seen, are the unknowns of the problem. Up to about a dozen objects may be handled in parallel, presently. Parameter, shape and state adaptation is done by prediction-error-feedback exploiting information in the Jacobian matrices as linear approximations of the relationship between

image features and 4-D models (Extended Kalman Filter for perspective projection). This core process also takes care of intelligent image feature extraction by proper specification of the parameters in the algorithms; orders of magnitude in computing efficiency may be gained by doing this carefully.

The **D**ynamic **O**bject data**B**ase (**DOB**) - containing the scene tree representation among other knowledge components about the vehicle status - is the central (vertical) layer separating the 'systems-engineering' lower part of the overall cognitive system from the more 'Artificial Intelligence'-oriented upper part. On the higher levels, the situation is assessed, and behavioral decisions are taken in the mission context. Here, knowledge about the effects of maneuvers and of the application of feedback control laws is available. Actual maneuver performance and control computations are done on the lower levels with dedicated processors in the distributed overall system.

Experimental results with the test vehicle 'VaMoRs' (a 5-ton van) will be shown. Driving on unmarked dirt roads, detecting a crossroad and turning-off onto it will be demonstrated including multi-focal and saccadic vision. A second application example is driving on a grass surface, detecting negative obstacles like ditches, and stopping autonomously in front of it. This has been done with an US-partner on our test site recently, using their real-time stereo processing hard- and software in connection with our EMS-vision system and test vehicle. Ditches as small as 60 cm in width have been detected and verified sufficiently early to stop the vehicle from a speed of up to 10 mph autonomously.

An outlook on further developments will be given.

Ernst D. Dickmanns