ANALYSING SEQUENCES OF TV-FRAMES

H.-H. Nagel
Institut fuer Informatik dcr Universitaet Hamburg
Schlueterstrasse 70, D-2000 Hamburg 13/Cera.any

A system is being implemented to analyse sequences of TV-frames from real world scenes like a street intersection with cars. A monocular black/white TV-camera will record - without changing its position or viewing direction - a scene in realtime on an AKPEX analog disk . The analysis system must not manipulate objects and their environment by mechanical devices. Changes in color or direction of the illumination as well as laser ranging or slit illumination will be excluded.

A (partial) 3-D description of non-stationary (moving objects) as well as stationary scene components (fore- and background) should be derived without reliance upon domain specific knowledge. Continuity in tine and other domain independent knowledge - as about the changing perspective projections of a moving body - will be exploited to attain this goal.

Frame-to-frame changes will have to be explained as being due to one or more solid, opaque moving object. As a first step, non-stationary image components are isolated from a sequence of TV-frames using an algorithm described by Jain, Militzer and Nagel 1977. Relations are then established between non-stationary image components from different frames. Based on such a series, a 3-D description of an object is hypothesized. The trajectory of this hypothesized object and its rotation within its center of gravity system have then to be inferred by comparing the resulting projections into the image plane with the observed non-stationery image components. Emphasis lies more on a symbolic rather than the numeric object description derived by Baumgart 1974 (isolated objects on a turntable), Contrary to the approach of Chien and Jones 1975, no attempt is made to track objects in real-time.

To identify domain independent knowledge "chunks" in experiments with this scene analysis system, it will be endowed with a subsystem to evaluate operator questions about the analysed image sequence and to answer them. An unsatisfactory answer will only be counted as a system deficiency if the operator can prove that the information to answer his question could have been extracted solely from the analysed image sequence - without recourse to other scene specific knowledge unavailable to the system.

The system is structured into a "goal determination" subsystem which interfaces to the "image-analysis" and to the "display/natural language frontend" subsystem. The goal determination subsystem contains the domain independent knowledge in the form of a production system. The image analysis subsystem describes the image sequence at six different levels of abstraction:

- analog and digital TV-frames;
- estimation of stationary and non-stationary image components;
- regions and their boundaries for stationary and non-stationary image components, obtained by segmenting them according to greyvalue characteristics;
- candidates for projection of object surfaces, combining different regions representing the projection of one object candidate into the image plane;
- object and compound object candidates in a 3-D description;
- abstractions based on relations between objects.

Results for a simple situation - object translation parallel to the image plane, with image size, form and orientation remaining constant - have been reported by Nagel 1976. A minimum spanning tree clustering approach has been developed Dreschler 1977, Nagel 1977] for feature descriptors of non-stationary image components. It allows to connect object projections despite drastic overall changes of size, form or orientation - provided such changes build up in small steps from frame to frame. This approach has been applied successfully to three TV-frame sequences of street scenes with 25 to 50 frames each, covering, e.g., a car taking a left turn into a side street and coping with overall size changes in excess of 2:1 within 25 frames.

The availability of extended frame sequences is exploited by gradually building up a reliable description based on evidence from many frames without undue emphasis on the first frame. Rather the initial frames of a sequence are evaluated to form a plan indicating which subsection of the following frames should be investigated in more detail: the plan formation concept of Kelly 1970 is extended into the time dimension! A more detailed discussion of this system may be found in Nagel 1977, including comparisons with other approaches like the one of Chow and Aggarwal 1977.

References:
Baumgart 1974, Geometric Modeling for Computer Vision, STAN-CS-74-463 (October 74), Comp. Science Dep., Stanford Univ.

Chien and Jones 1975, Acquisition of Moving Objects and Hand-Eye Coordination, p. 737, IJCAI4, Tbilisi/ Georgia, USSR, September 75

Chow and Aggarwal 1977, Computer Analysis of Planar Curvilinear Moving Images, IEEE Trans.Comp. C-26, 179 (1977)

Dreschler 1977, Diplomarbeit in Vorbereitung, Institut fur Informatik, Universitat Hamburg

Jain, Militzer and Nagel 1977, Separating non-stationary from stationary scene components in a sequence of real world TV-images, IJCAI-77

Nagel 1976, Formation of an Object Concept by Analysis of Systematic Time Variations in the optically Perceptible Environment, Inst.Informatik, Hamburg Univ. Ifl-HH-B-27/76 (to appear in CG & IP)

Nagel 1977, Analysing Sequences of TV-Frames:System Design Considerations, Inst.Informatik, Hamburg Univ. Ifl-HH-B-33/77 (March 77)