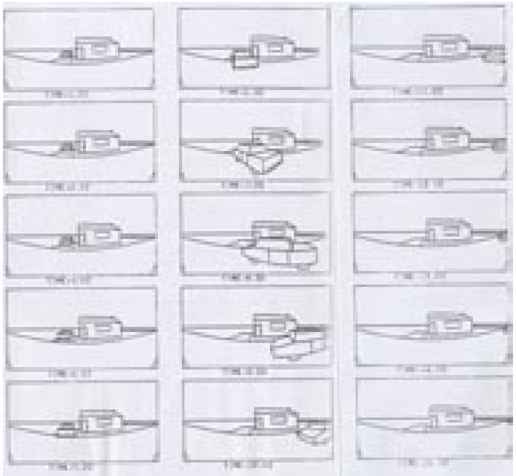
Historical Examples for High-level Vision

Early Traffic Scene Analysis (Badler 75)



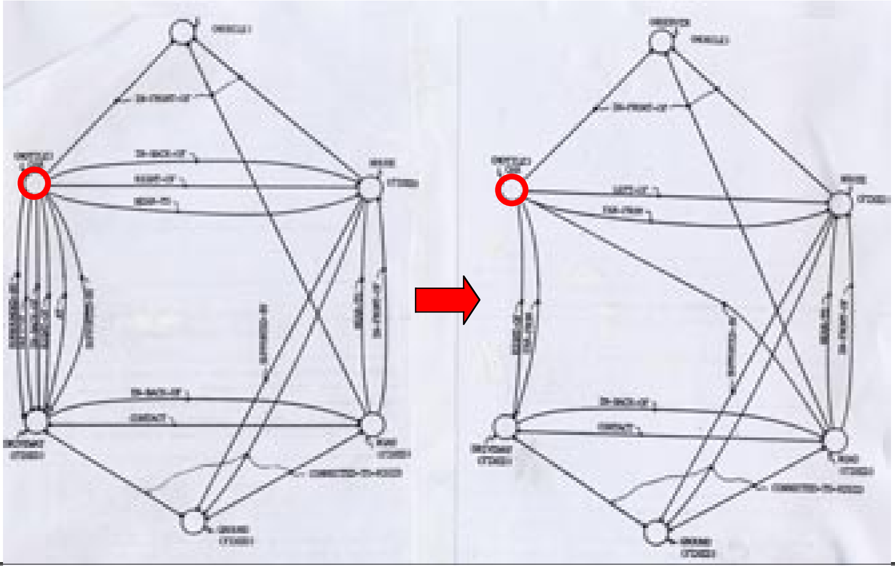
Task:
Describe motion in terms of changing spatial relations

15 "snapshots" of a car leaving the driveway of a house

Directional Adverbials for Motion Description (Badler 75)

ACROSS
AFTER
AGAINST
AHEAD-OF
ALONG
APART
AROUND
AWAY
AWAY-FROM
BACK
BACK-AND-FORTH
BACKWARD
BEHIND
BY
CLOCKWISE
COUNTERCLOCKWISE
DOWN
FORWARD
FROM
IN
IN-THE-DIRECTION-OF
INTO
INWARD
OFF
OFF-OF
ON
ONTO
ONWARD
OUT
OUT-OF
OUTWARD
OVER
SIDEWAYS
THROUGH
TO
TO-AND-FRO
TOGETHER
TOWARD
UNDER
UP
UP-AND-DOWN
UPWARD
WITH

Changing Scene Graph for Car Scene (Badler 75)



Demon Representation of "ACROSS" Motion (Badler 75)

A NEAR-TO relation with one side of an object is broken and replaced by a similar relation with the other side. There is an implicit sense of passage ABOVE the object.

Precondition 1
NEAR-TO(X S1).
SUB-PART(Y S1) for some object Y and SUB-PART [chain] to object S1.
FRONT or BACK or LEFT-SIDE or RIGHT-SIDE(Y S1).
ACROSS remains active as long as NEAR-TO(X Y) and A□BOVE(X Y) hold.

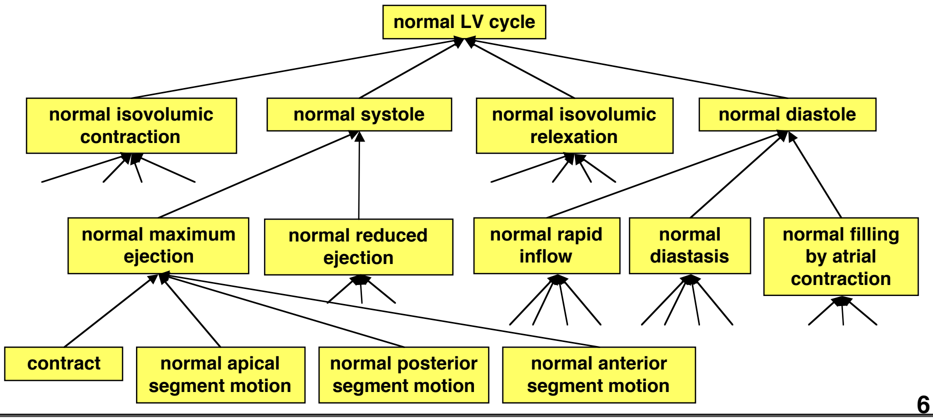
Precondition 2
NEAR-TO(X S2).
SUB-PART(Y S2) for a SUB-PART [chain] to object S2.
FRONT or BACK or LEFT-SIDE or RIGHT-SIDE(Y S2) where S1 ≠ S2 and at least one of the ORIENTATION relations to S1 (from Precondition 1) no longer holds.

Postcondition
SUBJECT X
DIRECTION PCONS((ACROSS Y), DIRECTION)

Left-ventricular Motion PART-OF Hierarchy (Tsotsos 79)

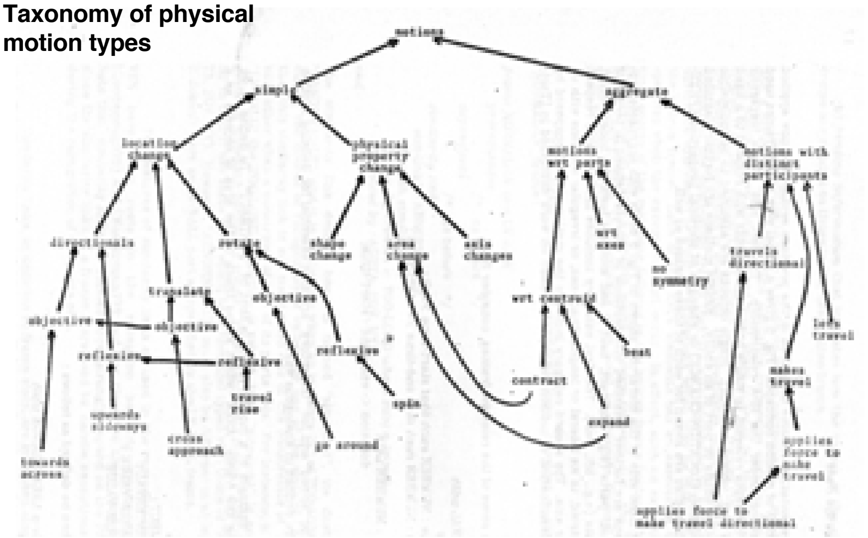
Task:
Recognize normal and abnormal heart conditions from ultrasound heart images

PART-OF structure supports part-whole reasoning in recognition processes



Motion IS-A Hierarchy (Tsotsos 79)

Taxonomy of physical motion types

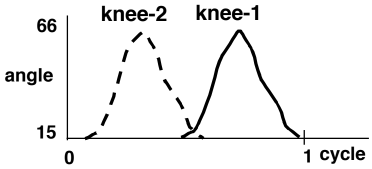


Model-based Prediction for Tracking a Jointed Moving Object (Hogg 84)

Task:
Describe highly coordinated motion of parts. Use quantitative measures along time axis.

Posture curves + constraints represent coordinated motion of joints of walker.

Example:

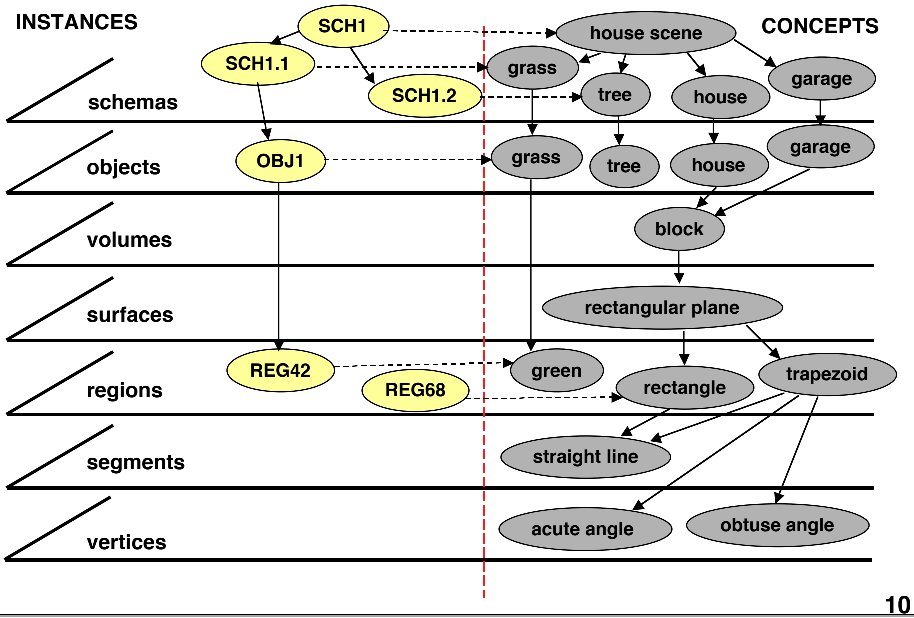


The VISIONS Image Interpretation System (Hanson & Riseman 78)

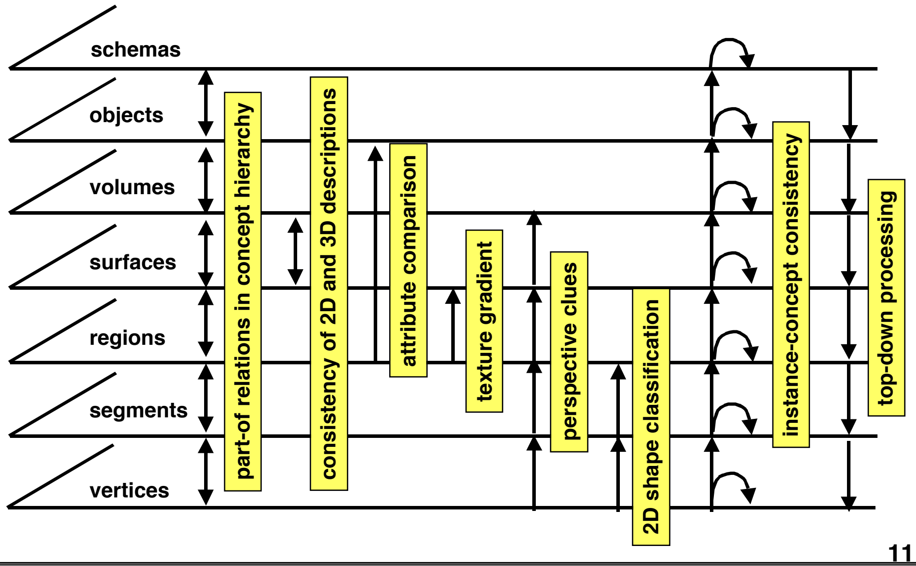
Long-term research about the interpretation of land-house scenes

(original in colour)

Representation Hierarchy in VISIONS



Knowledge Sources for VISIONS



schemas

objects

volumes

surfaces

regions

segments

vertices

part-of relations in concept hierarchy

consistency of 2D and 3D descriptions

attribute comparison

texture gradient

perspective clues

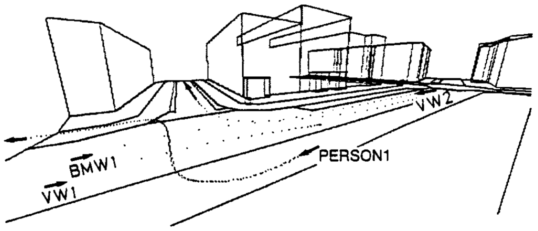
2D shape classification

instance-concept consistency

top-down processing

NAOS - Natural Language Description of Object Motions in Traffic Scenes

(Neumann & Novak 1986)



English paraphrase of automatically generated description:

The scene contains four moving objects: three cars and a pedestrian.

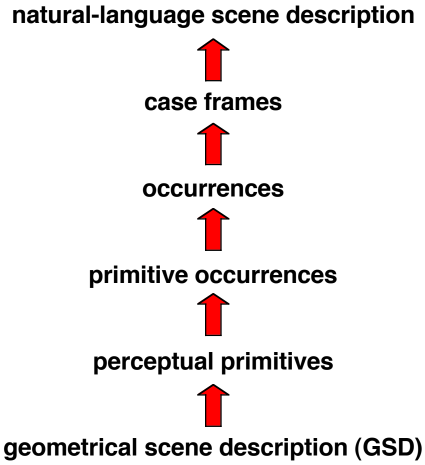
A VW drives from the Alte-Post to the front of the FBI. It stops.

Another VW drives towards Dammtor. It turns off Schlueterstrasse. It drives on Bieberstrasse towards Grindelhof.

A BMW drives towards Hallerplatz. While doing so, it overtakes the VW which has stopped, before Bieberstrasse. The BMW stops in front of the traffic lights.

The pedestrian walks towards Dammtor. While doing so, he crosses Schlueterstrasse in front of the FBI.

From Scene Data to a Natural-language Scene Description (NAOS)



natural-language scene description

case frames

occurrences

primitive occurrences

perceptual primitives

geometrical scene description (GSD)

Geometrical Scene Description (GSD) in NAOS

Quantitative description of all objects in a time-varying scene:

- name of all objects (class or identity)
- position of all objects at all times (location and orientation)
- illumination (if required for high-level description)

Example of a synthesized GSD in NAOS:

```
(LAGE VW2 (779. 170. 0.) (-1.0 0.0 0.0) 0)
(LAGE VW2 (753. 170. 0.) (-1.0 0.0 0.0) 1)
(LAGE VW2 (727. 170. 0.) (-1.0 0.0 0.0) 2)
(LAGE VW2 (701. 170. 0.) (-1.0 0.0 0.0) 3)
(LAGE VW2 (675. 170. 0.) (-1.0 0.0 0.0) 4)
(LAGE VW2 (649. 170. 0.) (-1.0 0.0 0.0) 5)
(LAGE VW2 (623. 170. 0.) (-0.999 0.037 0.0) 6)
(LAGE VW2 (596. 171. 0.) (-1.0 0.0 0.0) 7)
(LAGE VW2 (570. 171. 0.) (-1.0 0.0 0.0) 8)
(LAGE VW2 (544. 171. 0.) (-1.0 0.0 0.0) 9)
.
.
.
```

location orientation time

Occurrence Model for "OVERTAKE" (NAOS)

(OVERTAKE OBJ1 OBJ2 T1 T2) <=>
(MOVE OBJ1 T1 T2)
(MOVE OBJ2 T1 T2)
(BEHIND OBJ1 OBJ2 T1 T3)
(BESIDE OBJ1 OBJ2 T3 T4)
(BEFORE OBJ1 OBJ2 T4 T2)
(APPROACH OBJ1 OBJ2 T1 T3)
(DIS-APPROACH OBJ1 OBJ2 T4 T2)

➡ temporal constraint satisfaction for occurrence recognition

➡ principled definition of primitive occurrences

Temporal Relations in NAOS

- Observations provide begin and end time points of occurrences
- Models express qualitative constraints on time points

Unary temporal constraints: $t_{min} \le t \le t_{max}$

Binary temporal constraints: $t_1 \ge t_2 + c_{12}$

Convex interval relations may be expressed by inequalities:

I_1 during I_2 => $I_2.tb \le I_1.tb$
$I_1.te \le I_2.te$

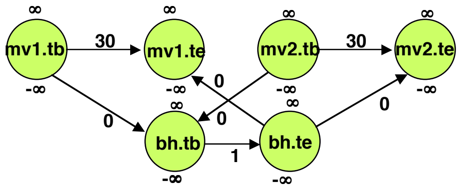
NAOS temporal constraint propagation was later identified as a convex time point algebra [Vila 94].

Constraint Propagation for Occurrence Verification (1)

Example:

Verify occurrence "two moving objects, one behind the other"

1. **Initialize constraint net of occurrence model**



2. **Compute primitive events for scene**

ID: move1
instance: move
parts: mv-ob = obj1
mv-tr = trj1
times: mv-tb = 13
mv-te = 47

ID: behind1
instance: behind
parts: bh-ob1 = obj1
bh-obj2 = obj2
times: bh-tb = 20
bh-te = 33

(and many more)

Constraint Propagation for Occurrence Verification (2)

3. **Instantiate parts in occurrence model**

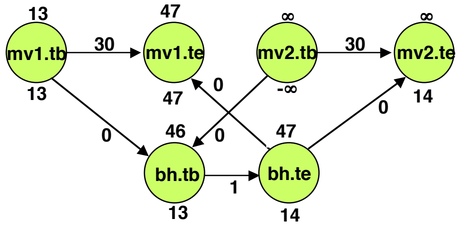
propagate minima and maxima of time points through constraint net:

- minima in edge direction $t_{2min}´ = \max\{t_{2min}, t_{1min} + c_{12}\}$

- maxima against edge direction $t_{1max}´ = \min\{t_{1max}, t_{2max} - c_{12}\}$

Example: **move1 in scene instantiates mv1 of model**

ID: move1
instance: move
parts: mv-ob = obj1
mv-tr = trj1
times: mv-tb = 13
mv-te = 47



Constraint Propagation for Occurrence Verification (3)

4. **Consistency and completeness test**

   A (partially) instantiated model is inconsistent, if for any node T one has: Tmin > Tmax

   => search for alternative instantiations or terminate with failure

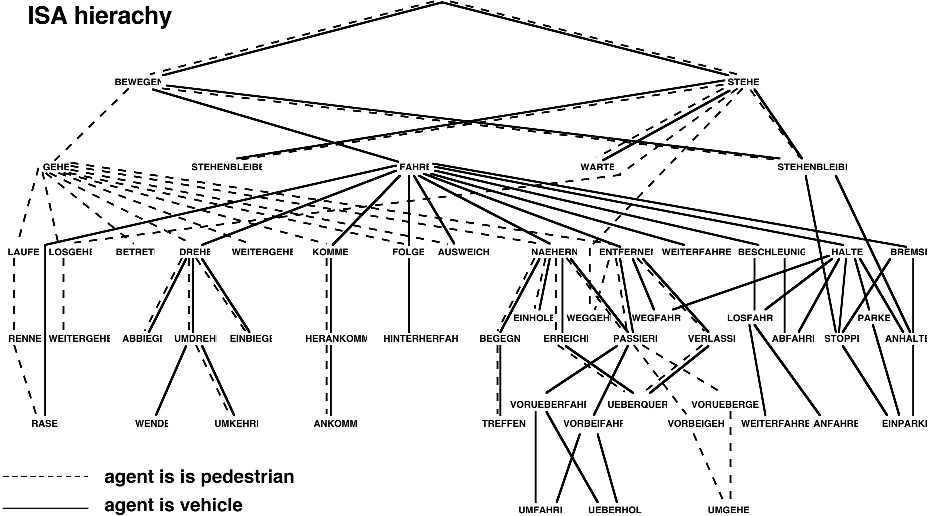
   An occurrence has been recognized if the occurrence model is instantiated with sufficient completeness and the instantiation is consistent.

Note:

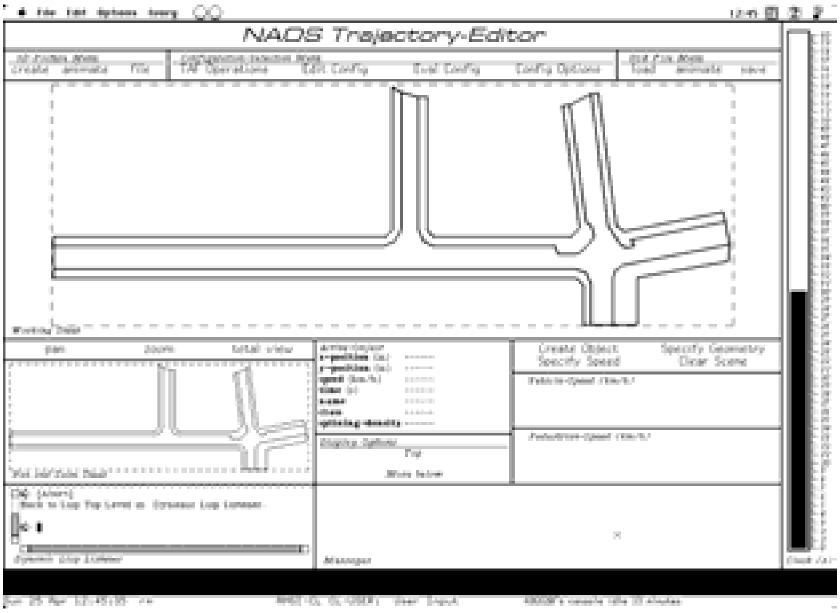
- **Incremental occurrence recognition follows an evolving scene**
- **A-posteriori occurrence recognition is carried out after observing a scene (choice of order!)**
- **Partially instantiated models may be used for scene prediction**

Hierachy for Object Motions in Street Traffic (NAOS)

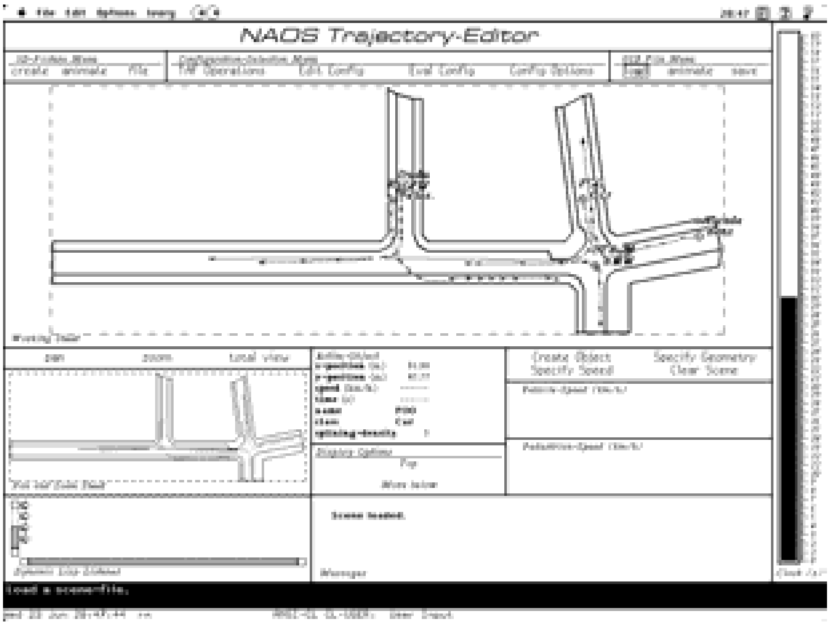
mix of part_of and ISA hierachy



NAOS System: Editor



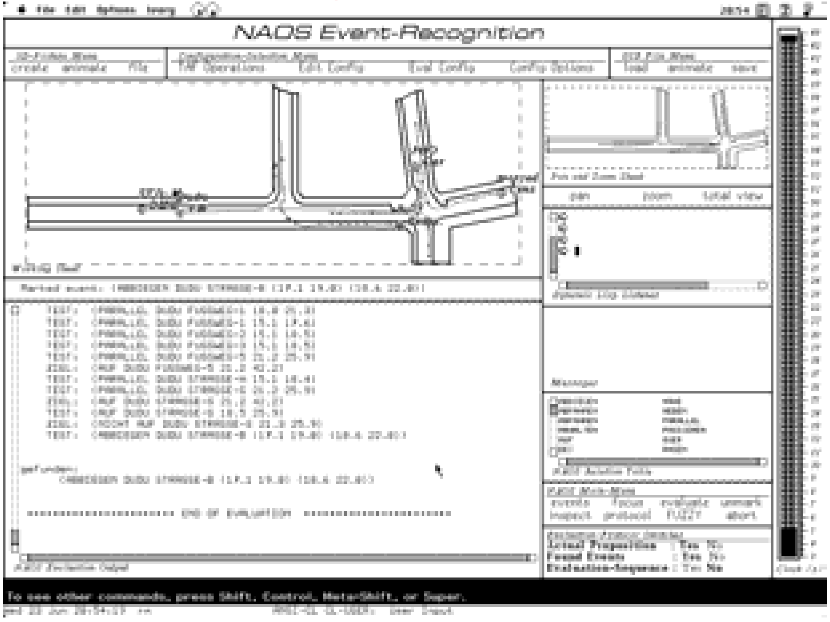
NAOS System: Synthetic Traffic Scene



Scene loaded.

Load a scene-file.

NAOS System: Recognizing a Turning-off Occurrence



Generating a Natural-language Description

Principle:

sentence
↑
case frame
↑
occurrence

} language-oriented AI techniques

Problems:

- Which occurrences should be selected for verbalization?
- Which deep cases should be filled?
- Which addtitional time or location information is required?
- In which order should the information be presented?

Solution:

Speech planning based on hearer simulation

informing a hearer <=> enabling a hearer to imagine the scene

Standard Plan for Generating Natural-language Scene Descriptions in NAOS

- rules which assure that the hearer will be able to imagine the scene
- summary + descriptions of all object trajectories, each in chronological order
- no explicit hearer simulation

Description of an object trajectory

1. Each time interval is described by the most special occurrence
2. The first occurrence begins at the beginning of the scene
3. The next occurrence follows in temporal order
4. Location information is given by prepositional expressions as required
5. Temporal information is given by prepositional expressions or references to other occurrences as required

Lessons from Historical Examples for High-level Scene Interpretation

- High-level scene interpretation requires representation and recognition of object motions.
- Representations may involve taxonomies and partonomies.
- Representations may be in quantitative and/or qualitative terms.
- Representations may involve temporal and spatial constraints on objects.
- Recognition may be incremental or a-posteriori.
- A natural-language description is one possible form of a high-level scene interpretation.