Knowledge-based Scene Interpretation

Bernd Neumann
Hamburg University, Germany
neumann@informatik.uni-hamburg.de

in cooperation with

Francois Brémont and **Monique Thonnat**
INRIA, Sophia Antipolis, France
{Francois.Bremont, Monique.Thonnat}@sophia.inria.fr

Intended Audience

- The slides are intended for a graduate course of roughly 20 hours (14 lectures of 90 min each).
- Students are expected to possess basic knowledge in Computer Vision and Artificial Intelligence.

2

Website

The website for this course can be reached via

http://kogs-www.informatik.uni-hamburg.de/~neumann/KBSI-WS-2006/

You will find PDF copies of the slides and possibly other useful information related to the course.

The website will be updated each week on Monday.

Contents (1)

Contents (2)

5

What is Computer Vision?

Computer Vision is the academic discipline dealing with task-oriented reconstruction and interpretation of a scene by means of images.

scene: **section of the real world**
stationary (3D) or moving (4D)

image: **view of a scene**
projection, density image (2D)
depth image (2 1/2D)
image sequence (3D)

reconstruction and interpretation: **computer-internal scene description**
quantitative + qualitative + symbolic

task-oriented: **for a purpose, to fulfill a particular task**
context-dependent, supporting actions of an agent

6

What Is Scene Interpretation?

Scene Interpretation is the task of "understanding" or interpreting a scene beyond single-object recognition. Typical examples are traffic scene interpretation for driver assistance, inferring user intentions in smart-room scenarios, recognizing team behavior in robocup games, discovering criminal acts in monitoring tasks.

Characteristics:

- **Interpretations involve several objects and occurrences.**
- **Interpretations depend on temporal and spatial relations between parts of a scene**
- **Interpretations describe the scene in qualitative terms, omitting geometric details.**
- **Interpretations include inferred facts, unobservable in the scene.**
- **Interpretations are based on conceptual knowledge and experience about the world.**

"Scene interpretation" means roughly the same as "high-level vision".

Examples for Scene Interpretation (1)

scene interpretation means understanding every-day occurrences

Garbage collection in Hamburg (1 frame of a sequence)

We want to recognize parts, activities, intentions, spatial & temporal relations

Examples for Scene Interpretation (2)

Scene interpretation is silent movie understanding

Buster Keaton in "The Navigator"

We want to recognize episodes, the "story", emotions, funnyness

Some Application Scenarios for Scene Interpretation

- Street traffic observations (long history)
- Cameras monitoring parking lots, railway platforms, supermarkets, nuclear power plants, ...
- Video archiving and retrieval
- Automatic soccer commentator
- Smart room cameras
- Autonomous robot applications
  (e.g. robot watchmen, playmate for children, assistance for elderly)

Technological Challenges of Scene Interpretation Tasks

- Problem area combines Computer Vision (CV) and Artificial Intelligence (AI), not well attended by CV and AI research
- Interpretations may build on common sense knowledge, common-sense knowledge representation is an unsolved issue
- Application scenarios may be large and highly diverse, knowledge engineering is a challenge
- Visual learning and adaptation may be required
- Reliability and complexity management may become important issues
- Economical application development requires generic approach

Cognitive Computer Vision

Scene interpretation is strongly related to "cognitive vision", a term created for vision comparable to human vision:

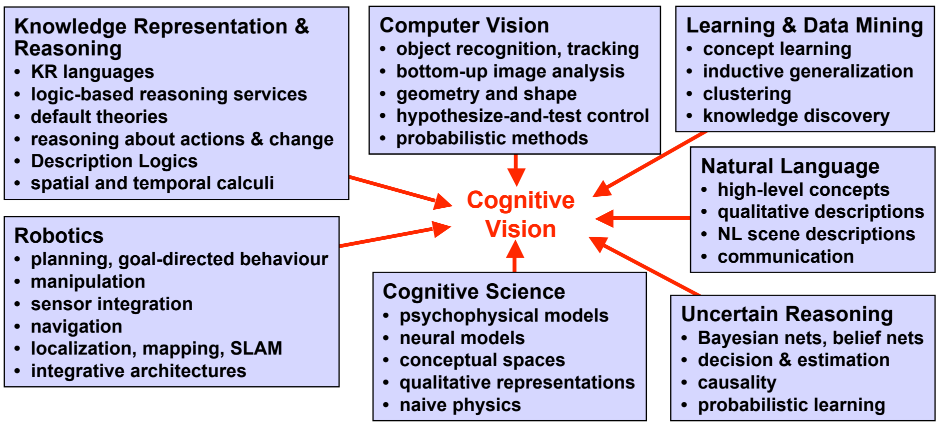
Cognitive computer vision is concerned with integration and control of vision systems using explicit but not necessarily symbolic models of context, situation and goal-directed behaviour. Cognitive vision implies functionalities for knowledge representation, learning, reasoning about events & structures, recognition and categorization, and goal specification, all of which are concerned with the semantics of the relationship between the visual agent and its environment.

Topics of cognitive vision:

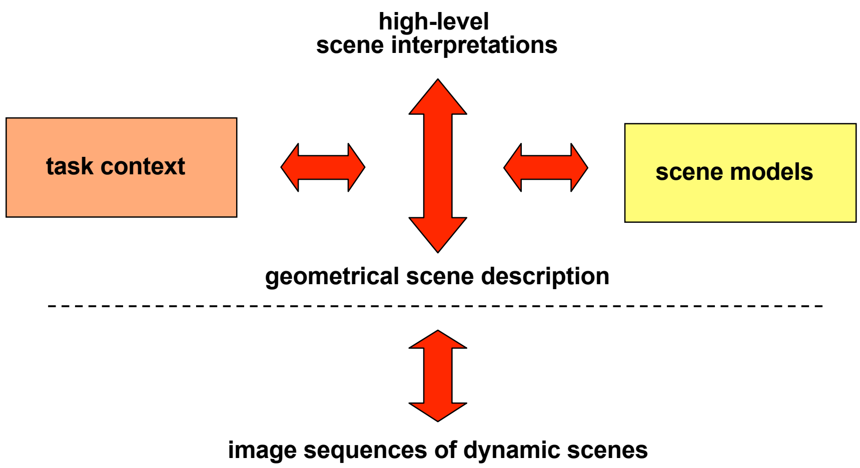
- integration and control
- explicit models
- not necessarily symbolic
- context
- situation
- goal-directed behaviour
- knowledge representation
- learning
- reasoning
- recognition
- categorization
- goal specification
- visual agent

Multidisciplinary Contributions to Cognitive Vision

Cognitive Vision research requires multidisciplinary efforts and escape from traditional research community boundaries.



Basic Structure of Knowledge-based Scene Interpretation



Representation Levels for High-level Scene Interpretation

symbolic scene descriptions — symbolic domain

pictorial scene representations

metric feature representations

(metric domain: pictorial scene representations, metric feature representations)

map-type feature representations — signal domain

raw image sequence

Context and Task Dependence

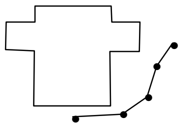
Interpretations may depend on

- domain context
- spatial context
- temporal context
- intentional context
- task context
- communicative context
- focus of attention
- a priori probabilities

Constructing an interpretation is <u>not</u> a mapping from image data into interpretation space.

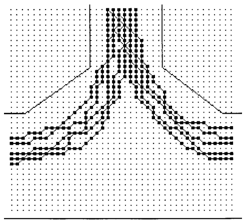
Signal-symbol Problems (1)

Mapping from quantitative into qualitative representations

Example:  ➡ *"along"*

"Citytour"
(Wahlster 87)

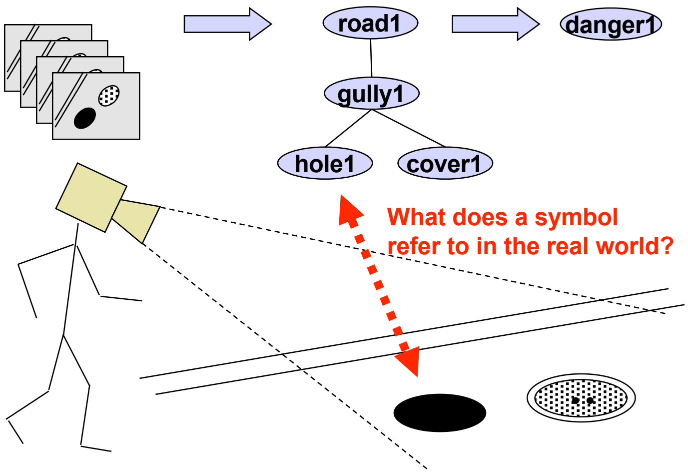
Mapping from qualitative into quantitative representations

Example: *"abbiegen"* ➡ 

"typicality fields"
(Mohnhaupt & Neumann 91, Herzog 94)

Signal-symbol Problems (2)

Symbol grounding



Common-Sense Problems

Common-sense reasoning

Deductions from symbolic knowledge about a scene should not only be correct w.r.t. to domain-related definitions but also w.r.t. to common sense.

Examples:

(implies (and house (some near lake)) mosquito-house)
(instance house1 house)
(instance lake1 lake)
(related house1 lake1 near)
(instance house1 (not (mosquito-house)))
=> inconsistent by domain-related definitions

(instance house1 house)
(instance cup1 cup)
(related house1 cup1 inside)
=> inconsistent by common sense

Uncertainty Problems (1)

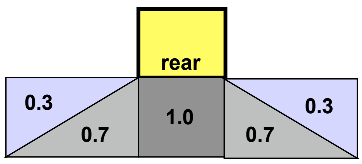
Fuzzyness of concepts

Many high-level concepts have unsharp boundaries.

"behind" "overtake" "meet"

=> mapping into logical propositions may be problematic

- Fuzzy set theory offers "degree of applicability"
- Probability theory offers statistical measures for language use

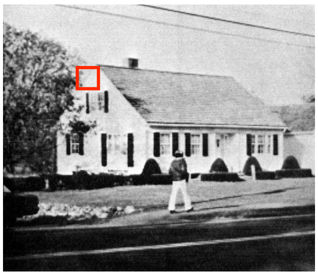


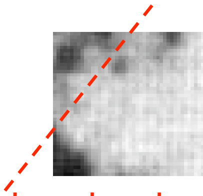
Fuzzy definition of behind

Uncertainty Problems (2)

Uncertainty of data

<u>Example</u>: Object boundaries





house boundary is not discernable

Strict bottom-up image interpretation is fundamentally ill-defined

Uncertainty Problems (3)

Exploring multiple hypotheses

Answers from several disciplines:

- **graph matching**
- **heuristic search**
- **optimization theory**
- **logic theories**
- **probability & utility theory**
- **case-based reasoning**
- **neural networks**
- **particle physics**

(and others)

Mixed bottom-up and top-down interpretation strategies have been rarely explored

Uncertainty Problems (4)

Cultural clash betweeen logical and probabilistic reasoning

Probabilistic methods are nor yet seamlessly integrated with logical calculi

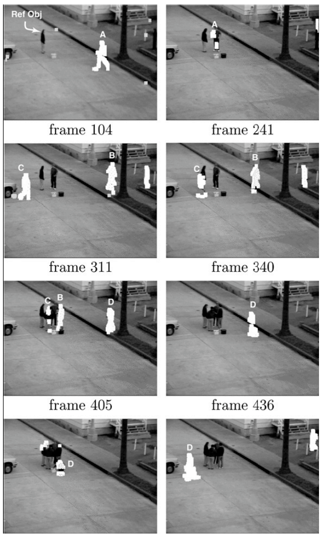
<u>Interesting recent developments</u>:

- **First-order probabilistic inference (Poole 03)**
- **Probabilistic relational models (http://dags.stanford.edu/PRMs/)**

Example for reasoning in image interpretation:
**(from Kanade´s invited lecture at IJCAI-03:
"Computer Vision: AI or Non-AI Problem?")**

**car on left side of street
(uncertain orientation of car)** }
japanese signs => left-hand traffic } **orientation of car resolved**

State-of-the-art Example of Scene Interpretation



S. Hongeng, R. Nevatia and F. Bremond.
Video-Based Event Recognition: Activity Representation and Probabilistic Recognition Methods.
Computer Vision and Image Understanding, Vol. 96 (2004), 129 - 162.

Recognising "Stealing by Blocking":

"A" approaches a reference object (a person standing in the middle with his belongings on the ground). "B" and "C" then approach and block the view of "A" and the reference person from their belongings. In the mean time, "D" comes and takes the belongings.