# High-level Vision

Bernd Neumann
Slides for the course in WS 2004/05

Faculty of Informatics
Hamburg University
Germany

neumann@informatik.uni-hamburg.de
http://kogs-www.informatik.uni-hamburg.de

## Contents (1)

# Contents (2)



3

# Website

**The website for this course can be reached via**

http://kogs-www.informatik.uni-hamburg.de/~neumann/HBD-WS-2004/

**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 Friday.**

4

# Introduction: What is High-level Vision?

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**<br>stationary (3D) or moving (4D) |
| **image:** | **view of a scene**<br>projection, density image (2D)<br>depth image (2 1/2D)<br>image sequence (3D) |
| **reconstruction and interpretation:** | **computer-internal scene description**<br>quantitative + qualitative + symbolic |
| **task-oriented:** | **for a purpose, to fulfill a particular task**<br>context-dependent, supporting actions of an agent |

6

# What Is High-level Vision?

***High-level vision is the task of "understanding" a scene beyond single-object recognition. Typical examples are traffic scene understanding 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 understanding" means roughly the same as "high-level vision".**

# Examples for High-level Vision (1)

**high-level vision 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 High-level Vision (2)

**High-level vision is silent movie understanding**

**Buster Keaton in "The Navigator"**

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

# Some application scenarios for high-level vision

- **street traffic observations (long history)**
- **cameras monitoring parking lots, railway platforms, supermarkets, nuclear power plants, ...**
- **video archiving and retrieval**
- **soccer commentator**
- **smart room cameras**
- **autonomous robot applications (e.g. robot watchmen, playmate for children )**

# Characteristics of High-level Scene Interpretation Tasks

- interpretations typically involve several interrelated objects
- spatial and temporal relations are important
- interpretations may build on common sense knowledge
- application scenarios are highly diverse
- domains may be very large
- learning and adaptation may be required
- reliability and complexity management may become important issues
- economical application development requires generic approach

# Cognitive Computer Vision

High-level vision is strongly related to "cognitive vision", a term created for vision comparable to human vision:

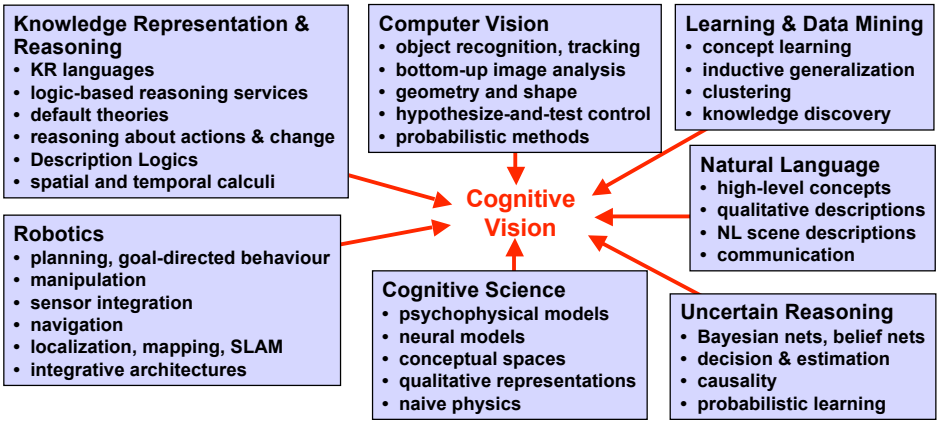
***Cognitive computer vision is concerned with integration and control of vision systems using explicit but not necessarily symbolic models of context, situation and goal-directed behaviour. Cognitive vision implies functionalities for knowledge representation, learning, reasoning about events & structures, 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:**

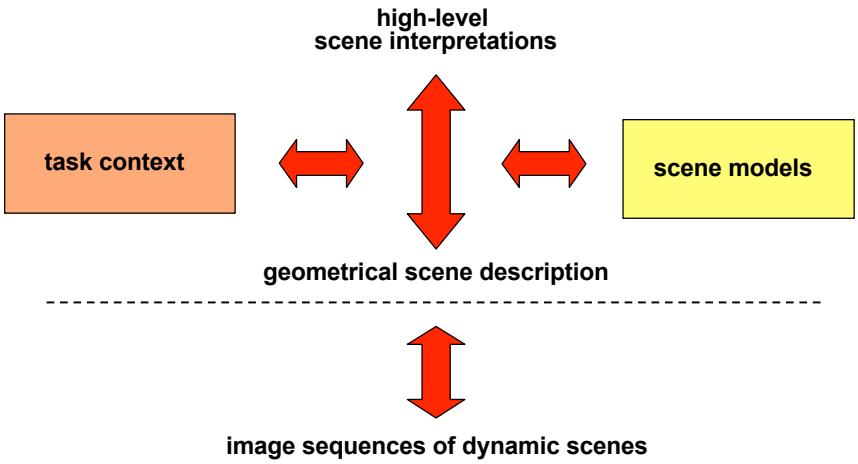
- integration and control
- explicit models
- not necessarily symbolic
- context
- situation
- goal-directed behaviour
- knowledge representation
- learning
- reasoning
- recognition
- categorization
- goal specification
- visual agent

# Multidisciplinary Contributions to Cognitive Vision

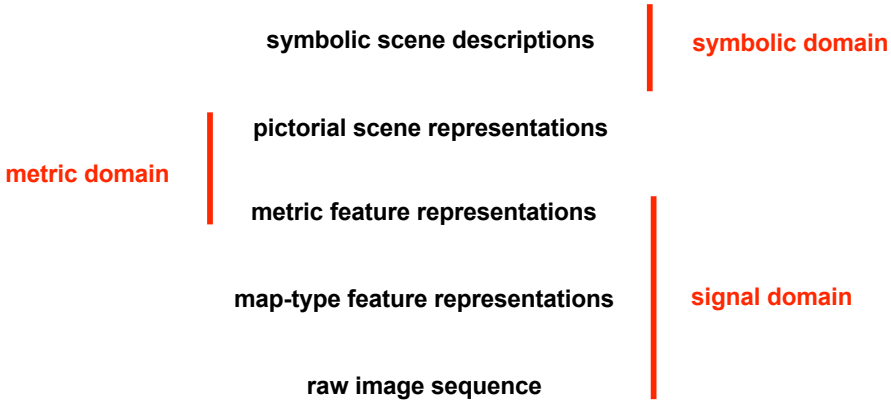
**Cognitive Vision research requires multidisciplinary efforts and escape from traditional research community boundaries.**




# Basic Structure of High-level Scene Interpretation



# Representation Levels for High-level Scene Interpretation




# Context and Task Dependence

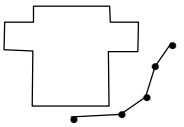
Interpretations may depend on

- 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.**

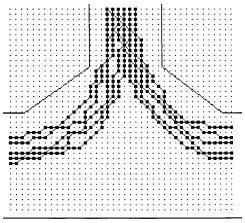
# Signal-symbol Problems (1)

**Mapping from quantitative into qualitative representations**

**Example:**  *"along"*

"Citytour"
(Wahlster 87)

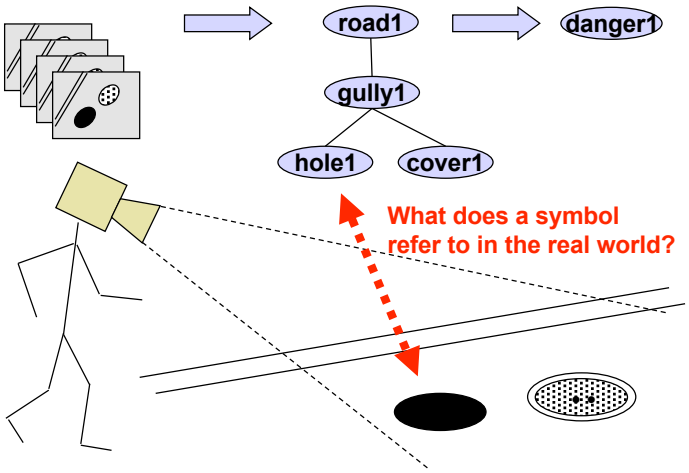
**Mapping from qualitative into quantitative representations**

**Example:** *"abbiegen"* 

"typicality fields"
(Mohnhaupt &
Neumann 91)

# Signal-symbol Problems (2)

**Symbol grounding**



# Signal-symbol Problems (3)

**Grounded symbolic reasoning**

Deductions from symbolic knowledge about a scene should not only be correct w.r.t. to the underlying logic 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!

(instance house1 house)
(instance cup1 cup)
(related house1 cup1 inside)
=> inconsistent???

# Uncertainty Problems (1)

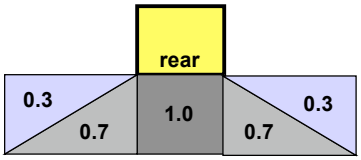
**Fuzzyness of concepts**

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

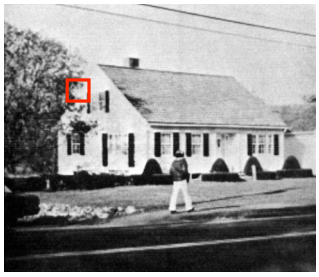


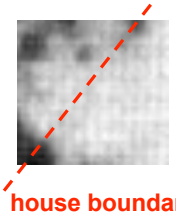

Fuzzy definition of behind

# Uncertainty Problems (2)

**Uncertainty of data**

**Example: Object boundaries**





**house boundary is not discernable**

**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**

# A State-of-the-art Example of Scene Interpretation


**recognizing assaults**




**recognizing thefts at a phonebooth**