Description Logics for Image Interpretation

Using description logics for
- **knowledge representation (visual phenomena, background knowledge)**
- **inferences (meaning assignment, interpretation)**

- **Important application as web service**
  - automatic annotation of images
  - content-based retrieval
  - multimedia content services
- **Current research topics**
  - logics of multimedia interpretation?
  - standard inference services?
  - part of Semantic Web?

Ucraine´s Andrey Sokolovskiy clears 2,38 in Rome

1

Logics of Image Interpretation

Image interpretation can be formalised as:

- ***Abduction***

  Construct an explanation of real-world evidence from your formal knowledge about the real-world domain.

  An interpretation is an instantiation of formal knowledge which allows to deduce the evidence.

- ***Partial Model Construction***

  Construct a partial mapping of the symbols of your formal knowledge about the world into a real-world domain.

  An interpretation is a partial instantiation of formal knowledge consistent with evidence about the real-world domain.

2

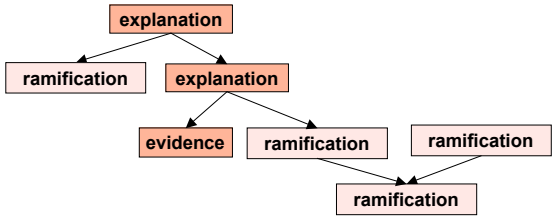
Scene Interpretation by Abduction

Shanahan, M. (2005): Perception as abduction: Turning sensor data into meaningful representation. Cognitive Science 29, 104-134

Compute Δ such that $\Sigma \cup \Delta \mapsto \Gamma$ with

- Σ background knowledge
- Γ evidence
- Δ explanation

Abduction focusses on evidence and does not provide additional ramifications.



3

Abduction in Description Logics (DLs)

- Abduction has only recently been introduced as a "non-standard inference service" in DLs.
- Growing interest in media interpretation for the Semantic Web.

First implementation in the commercial DL system RacerPro:

Solve $\Sigma \cup \Gamma_1 \cup \Delta \mapsto \Gamma_2$

Σ = ABox + TBox
Γ_1 = facts not needing an explanation
Γ_2 = facts needing an explanation
Δ = explanation

4

TBox for Table-Laying Domain

```
(implies plate dish)
(implies saucer dish)
(implies cup dish)
(implies napkin (or paper cloth))
(equivalent  cover
     (and  configuration
       (exactly  1  has-plate   plate)
       (exactly  1  has-saucer  (and saucer (near plate)))
       (exactly  1  has-cup     (and cup (on saucer)))
       (atmost 1 has-napkin   napkin)))
       (same-as   has-saucer ∘ near   has-cup)

(X Y near)   <=   (and  (Z cover)
                        (Z X has-plate)(X plate)
                        (Z Y has-saucer)(Y saucer))
(X Y on)     <=   (and  (Z cover)
                        (Z X has-cup)(X cup)
                        (Z Y has-saucer)(Y saucer))
```

DL-safe rules for representing constraints

5

Providing Rules for Explanations

```
(equivalent  cover
          (and  configuration
               (exactly  1  has-plate    plate)
               (exactly  1  has-saucer saucer)
               (exactly  1  has-cup      cup)
               (atmost 1 has-napkin   napkin)))
```

automatic conversion of all conjuncts of an aggregate definition

```
(X configuration)  <=   (and  (X cover)(X configuration)
(Y plate)          <=   (and  (X cover)
                              (X Y  has-plate)(Y plate)
(Y saucer)         <=   (and  (X cover)
                              (X Y has-saucer)(Y cup)
(Y cup)            <=   (and  (X cover)
                              (X Y has-cup)(Y cup)
```

DL-safe rules to allow abduction by backward-chaining

6

Abduction Example

Calling compute_explanations(Σ, Γ_1, Γ_2) in RacerPro for the table-laying knowledge base:

Γ_2 = {(plate1 plate)(saucer1 saucer)(plate1 saucer1 near)}

Δ = {(cover1 cover)(cover1 plate1 has-plate)(cover1 saucer1 has-saucer)}

7

Scene Interpretation by Partial Model Construction

Given a knowledge base with

- **general domain knowledge,**
- **specific context information,**
- **specific sensory evidence**

construct a mapping of

- **constant symbols into scene elements D,**
- **predicate and relation symbols into predicate and relation functions over D**

such that all predicates and relations are true.

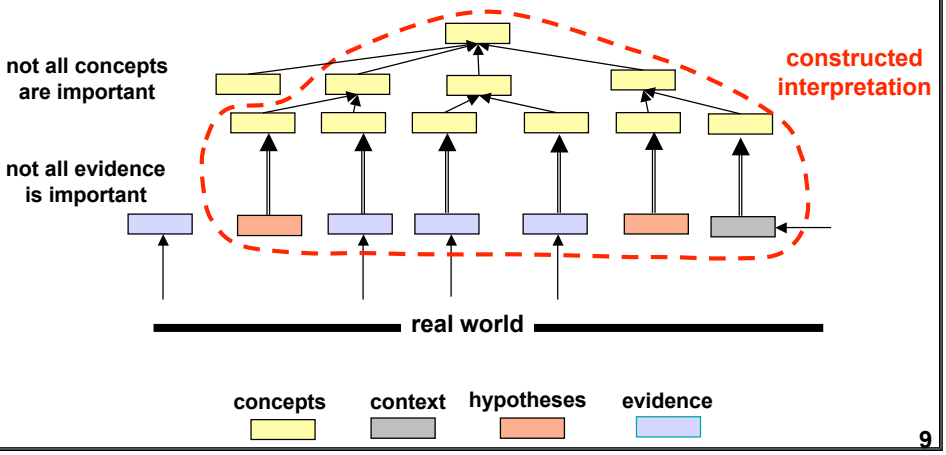
- **Operational semantics of low-level vision provide mapping into primitive constant and predicate/relation symbols.**
- **Hypotheses need no evidence.**

Clowes: "Vision is controlled hallucination"

8

Consistent Interpretations in Compositional Hierarchies

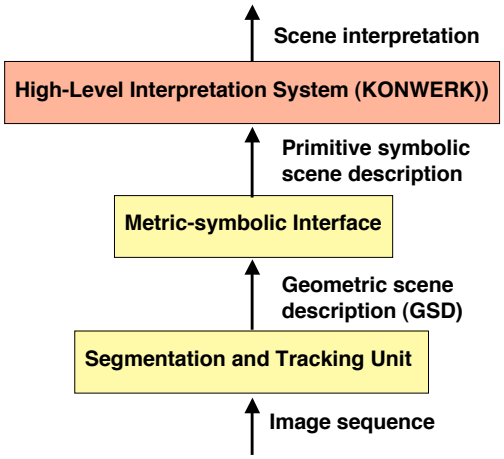
A scene interpretation is a scene description in terms of instantiated aggregate concepts consistent with evidence and context information.



9

Scene Interpretation as Configuration

Model Construction is also the basis of knowledge-based configuration

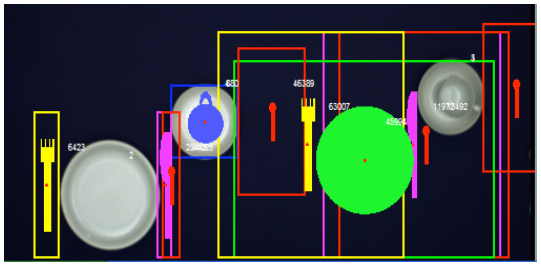


SCENIC uses a configuration system framework for scene interpretation

Hotz & Neumann 2005
Scene Interpretation as a Configuration Task
Künstliche Intelligenz, 3/2005, BöttcherIT Verlag, Bremen, 59-65

10

Experimental Results (1)



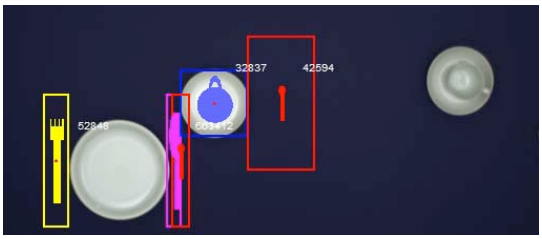
natural views = evidence
coloured shapes = hypotheses
boxes = expected locations

Intermediate state of interpretation after 51 interpretation steps:

- **"lay-dinner-for-2" hypothesis based on partial evidence**
- **predictions about future actions and locations**
- **high-level disambiguation of low-level classification**
- **influence of context**

11

Experimental Results (2)



- **alternative interpretation in terms of "dinner-for-one" and "cluttered-table" (after backtracking)**

12

Interpretation Issues Left Open by Logical Framework

- **Task-dependent scope and abstraction level**
  - **no need for checking all predicates**
    e.g. propositions outside a space and time frame may be uninteresting
  - **no need for maximal specialization**
    e.g. geometrical shape of "thing" suffices for obstacle avoidance
- **Ambiguous choices for interpretation steps**
  - **evidence classfication is naturally ambiguous**
  - **bad choices may cause inconsistency and backtracking**
- **Real-world agents need single "best" scene interpretation**
  - **requires uncertainty rating for evidence and context (propositions)**
  - **requires preference measure for scene interpretations**

Logical model property provides only loose frame for possible scene interpretations.

13