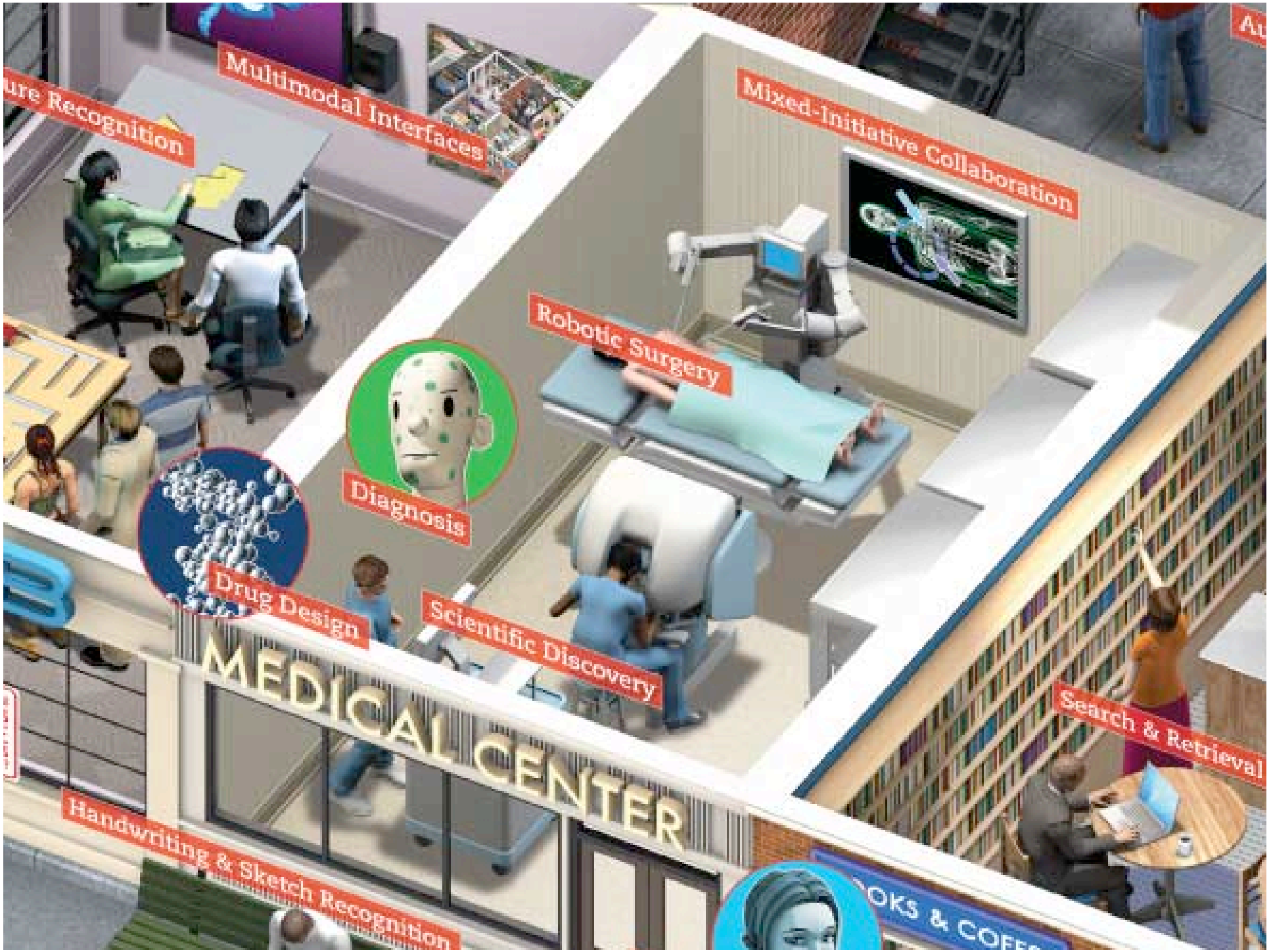
Image Processing and Scene Interpretation - Artificial Intelligence in Operation

Prof. Bernd Neumann

Cognitive Systems Laboratory / HITeC
Hamburg University
Germany

The AI Landscape


ure Recognition
Multimodal Interfaces
Mixed-Initiative Collaboration
Robotic Surgery
Diagnosis
Drug Design
Scientific Discovery
MEDICAL CENTER
Search & Retrieval
Handwriting & Sketch Recognition


Planning
AI & Creative Expression
Plan Recognition
PV=nRT
Intelligent Tutoring
Ubiquitous Computing
Social Computing
AI


Whitehead
Lovelace
Russell
Turing
Security & Privacy
User Modeling
Vehicle Navigation
Communications Triage
Machine Translation
CAN AI BRING PEOPLE TOGETHER?
HOW CAN AI SYSTEMS HARNESS COMMON SENSE?

Agenda

- Basic ingredients for Scene Interpretation
- Object Recognition with SIFT Features
- Ontologies with OWL
- Probabilistic Scene Interpretation

Scene Interpretation (1)

Scene interpretation means understanding every-day occurrences ...

Scene Interpretation (2)

... or recognizing
rare events

Some Application Scenarios for Scene Interpretation

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

Significant progress in the last 35 years

Activity Recognition at Blagnac Airport

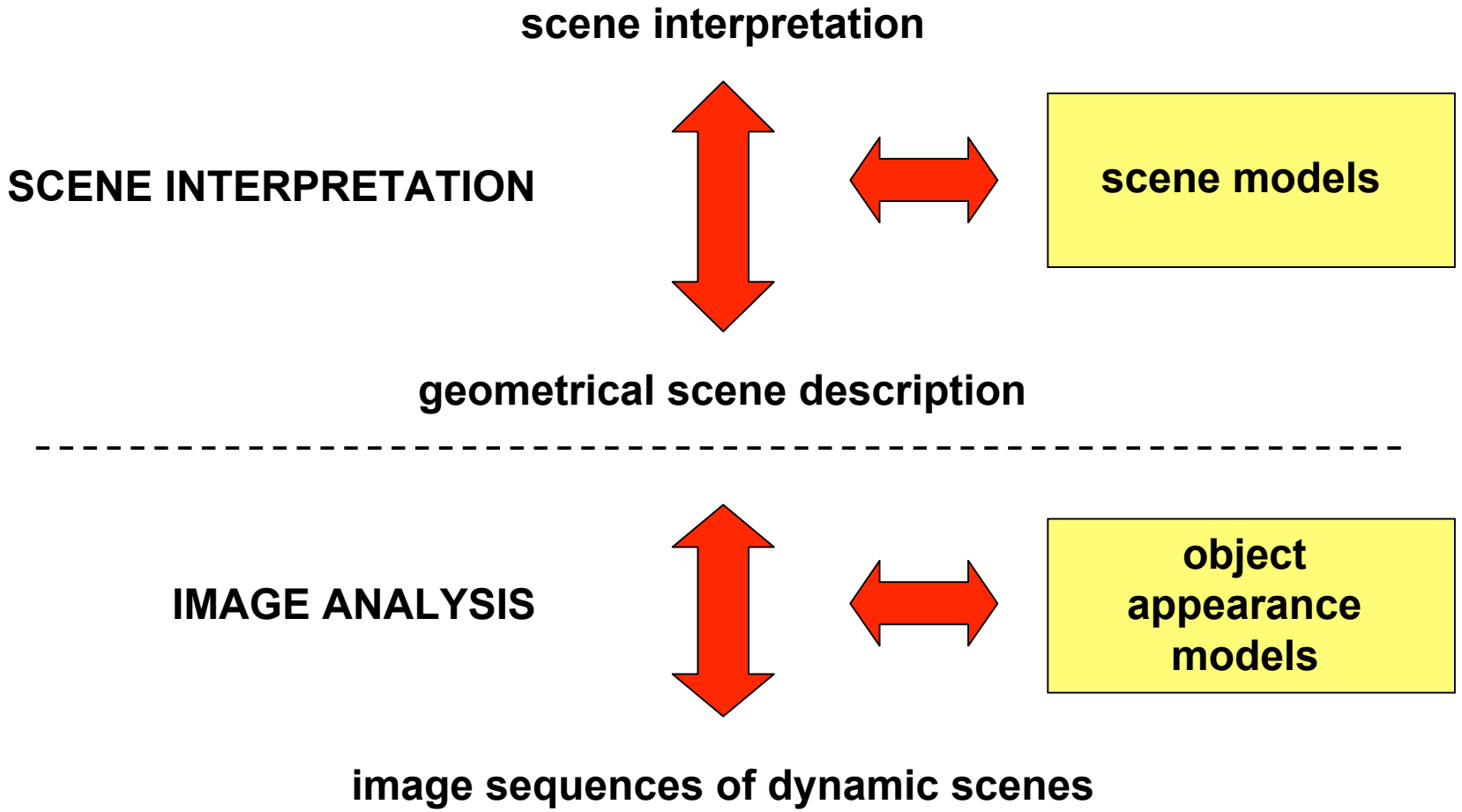
Application scenario

- Aircraft servicing operations at Toulouse-Blagnac Airport are observed by eight cameras
- Moving objects are tracked by a low-level vision system
- Activities such as refueling or baggage unloading are recognised by a high-level vision system

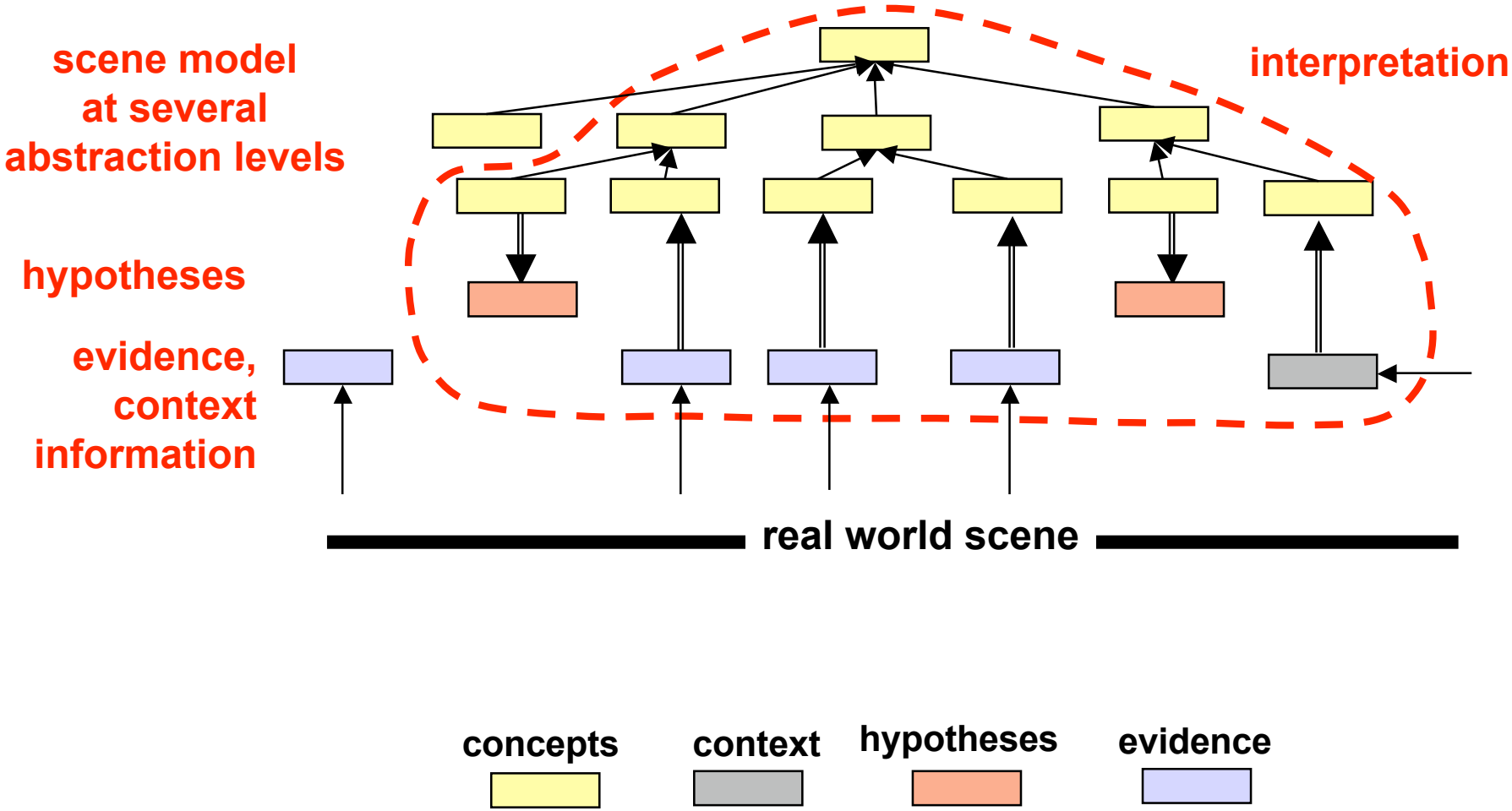
Project goals

- Reliable on-line interpretation of extended multi-camera video sequences
- Learning new activities from examples
- Robust recognition performance based on a rich domain ontology

Basic Structure of a Knowledge-based Vision System



Generic Scene Interpretation Process



Technological Challenges of Scene Interpretation Tasks

- Problem area combines Computer Vision (CV) and Artificial Intelligence (AI), not well attended by CV and AI research
- Reliable object recognition has progressed but is still a bottle-neck
- Interpretations may require large knowledge bases and common-sense reasoning
- Visual learning and adaptation may be necessary to build up and maintain knowledge bases
- Robust interpretation processes must be devised to cope with uncertain and incomplete visual information
- Economical application development requires a generic approach

But: High-level context may support low-level image analysis!

Advances in Image Analysis: SIFT Features

- Create object models in terms of sets of scale and rotation invariant SIFT features
- For recognition, use SIFT features of unknown object as index into model base
- Verify hit by least-squares fit

Detect "interest points" in image at multiple scales

Compute rich local description of image intensities at keypoints

Determine one or more main orientations

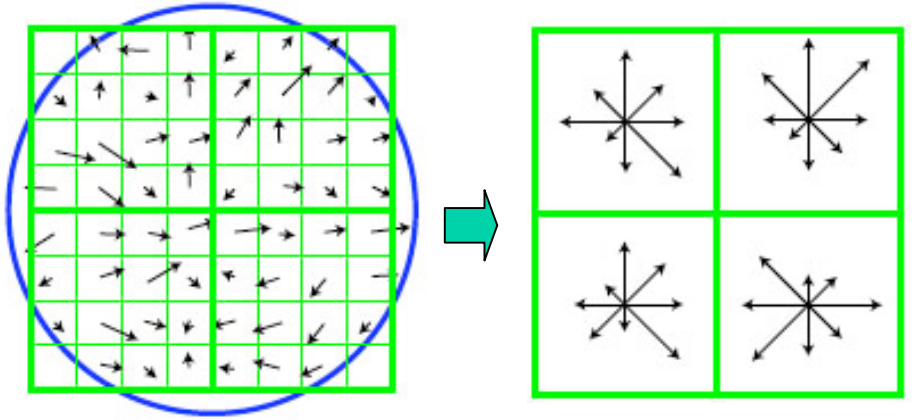
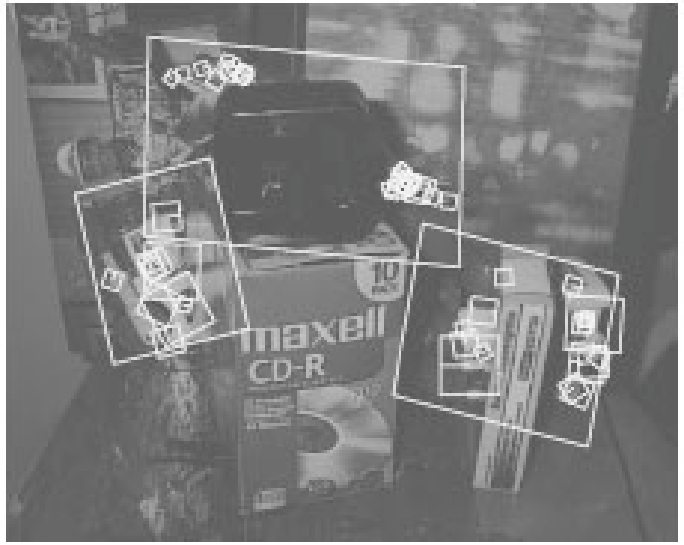


Image gradients **Keypoint descriptors**

SIFT-based Object Recognition

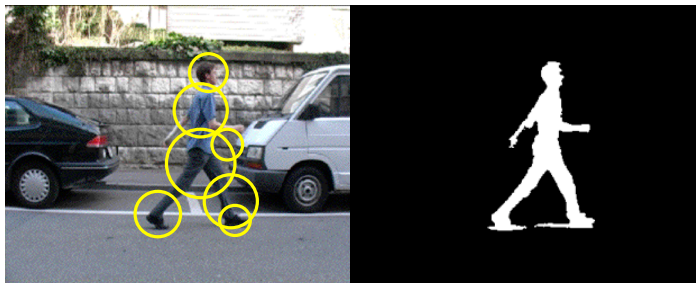
training images

cluttered image

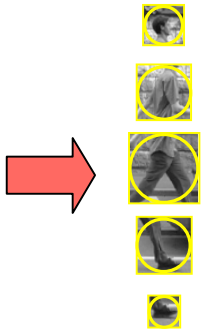


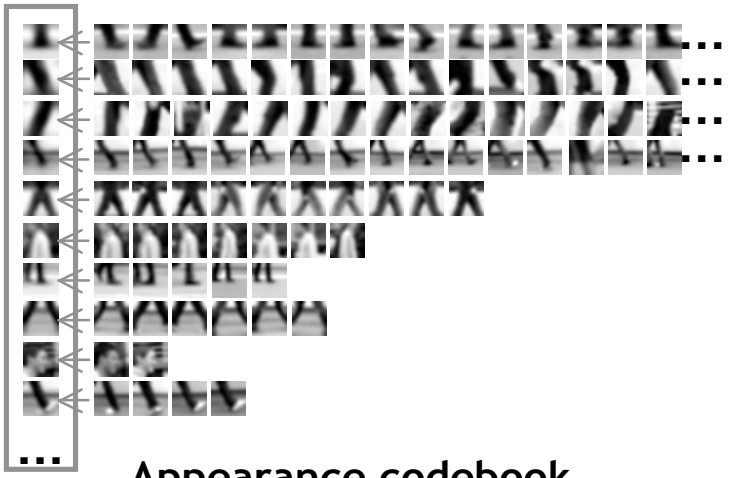
recognized objects

Learning Appearance Models



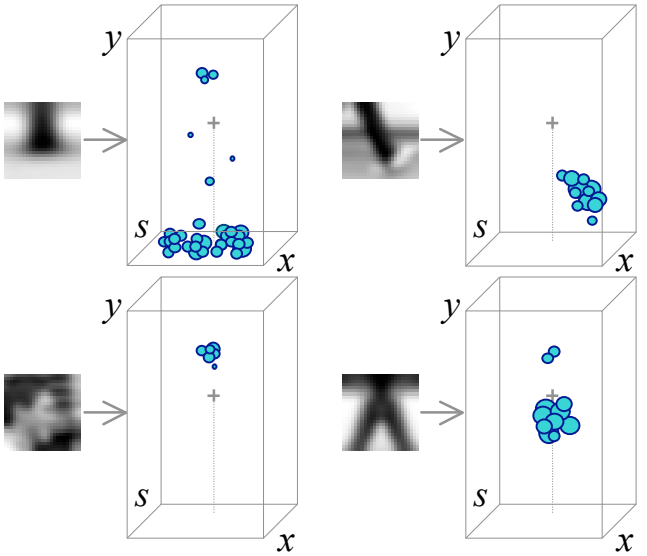
105 training images
(+ motion segmentation)





Appearance codebook

- Learn appearance codebook
  - Extract patches at DoG interest points
  - Agglomerative clustering ⇒ codebook
- Learn spatial distributions
  - Match codebook to training images
  - Record matching positions on object



Spatial occurrence distributions

Car Recognition by Appearance Models

- Recognizes different kinds of cars
- Robust to clutter, occlusion, noise, low contrast

Modelling High-Level Knowledge

- **Model-based scene interpretation requires extensive knowledge bases**

  => formal representation framework and reasoning support needed

- **Standardized knowledge representation with the Semantic Web ontology language OWL and RDF**

  - definition of classes with properties (e.g. "Person", "Vehicle")

  - definition of relationships between classes (e.g. subClassOf, disjointWith)

  - definition of individuals (e.g. "GPU-Access-Area", "Front-Loading-Area")

  - definition of class memberships

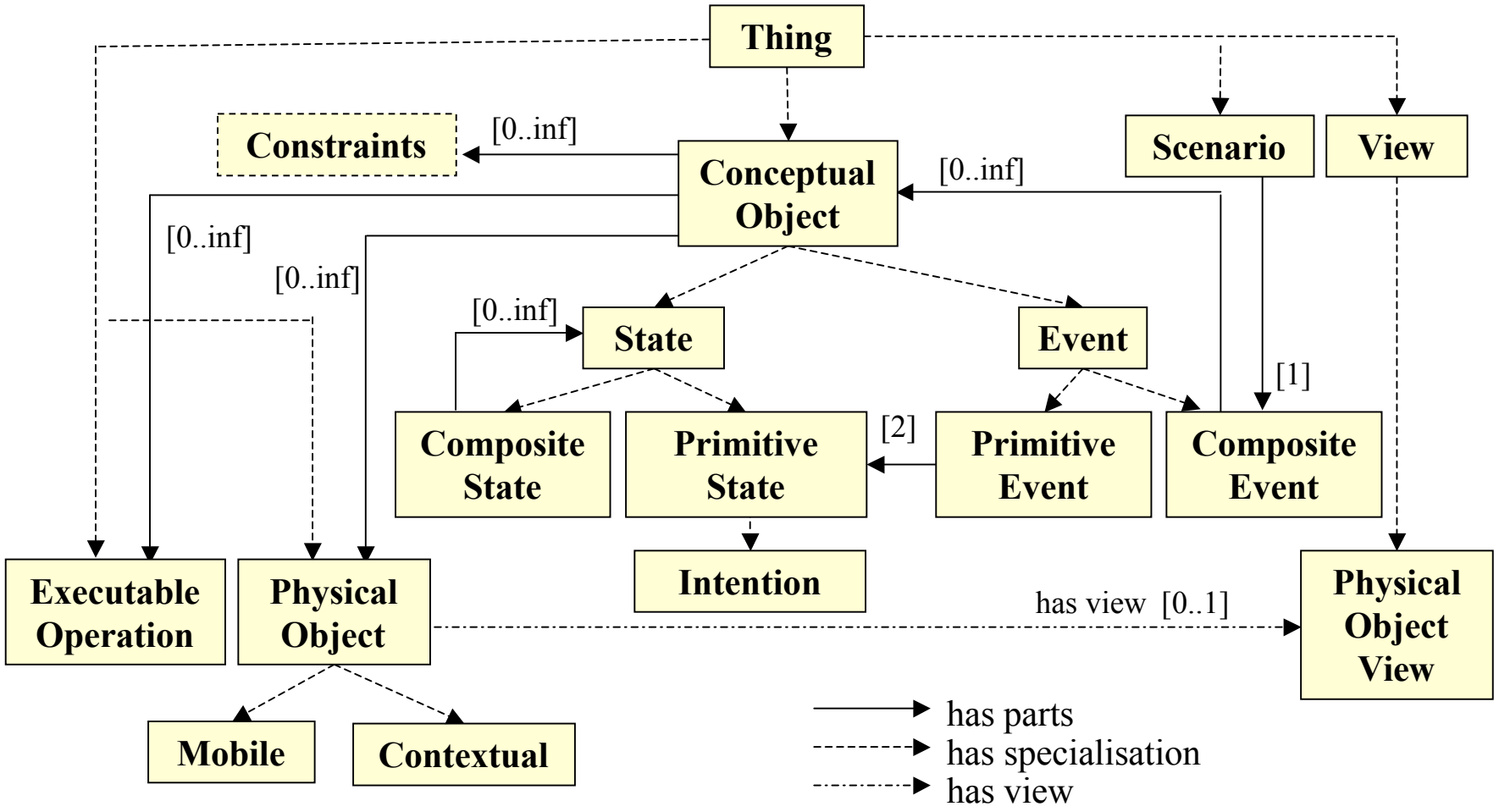
  - consistency checking

- **Commercially available reasoning systems (e.g. Pellet, Racer)**

Interplay of OWL knowledge base with scene interpretation is current research topic

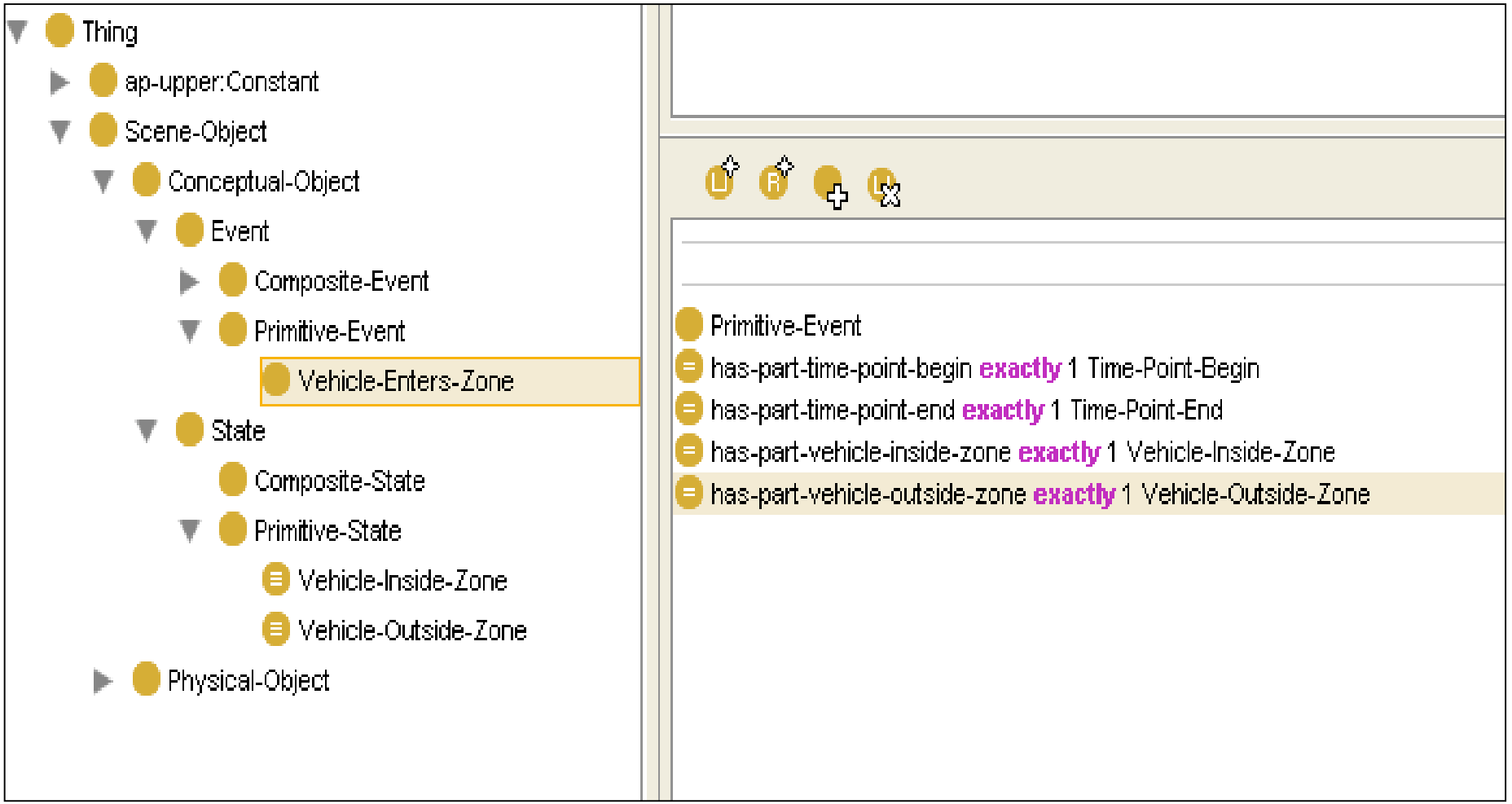
Example: Upper Model for Activity Recognition

Definition of ontological relationships between essential concepts



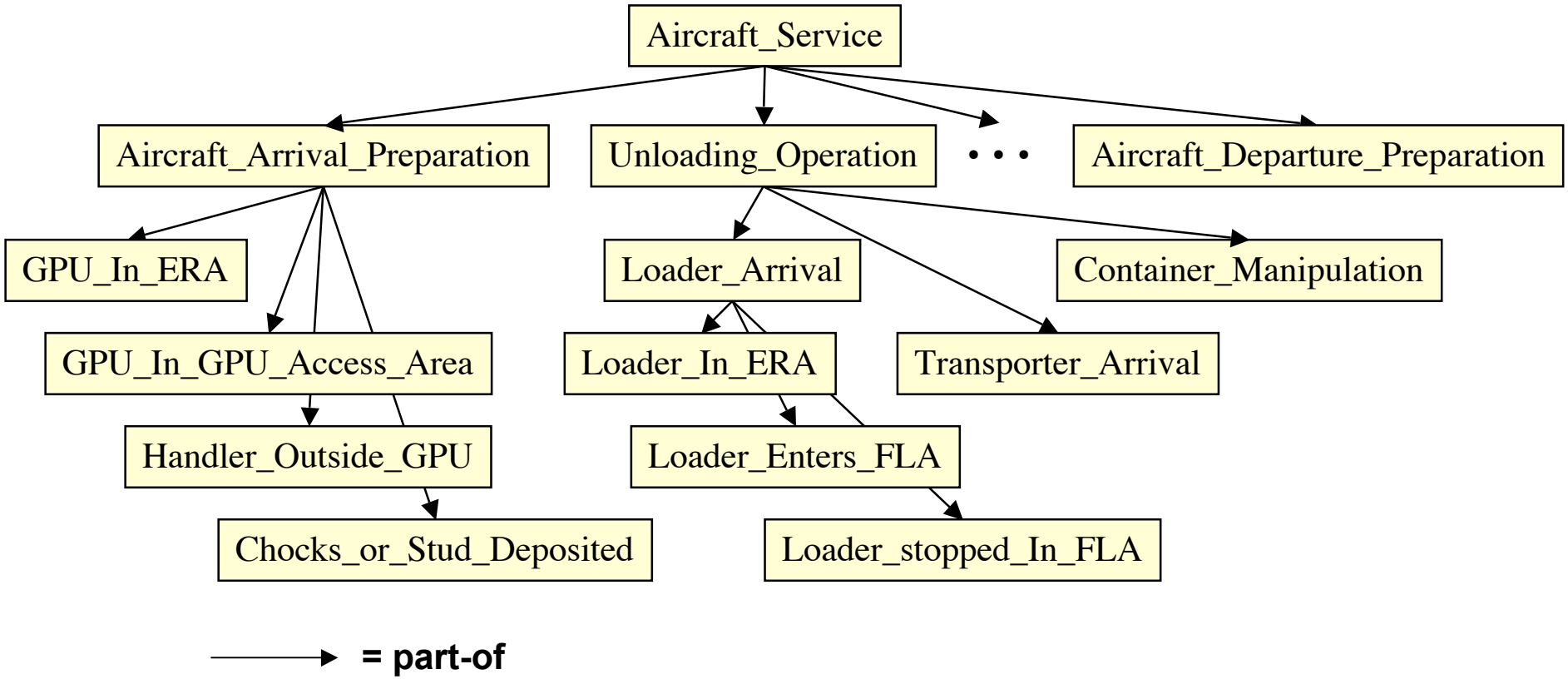
Protégé Ontology Editor for OWL

Definition of "Vehicle-Enters-Zone" for aircraft activity rcognition

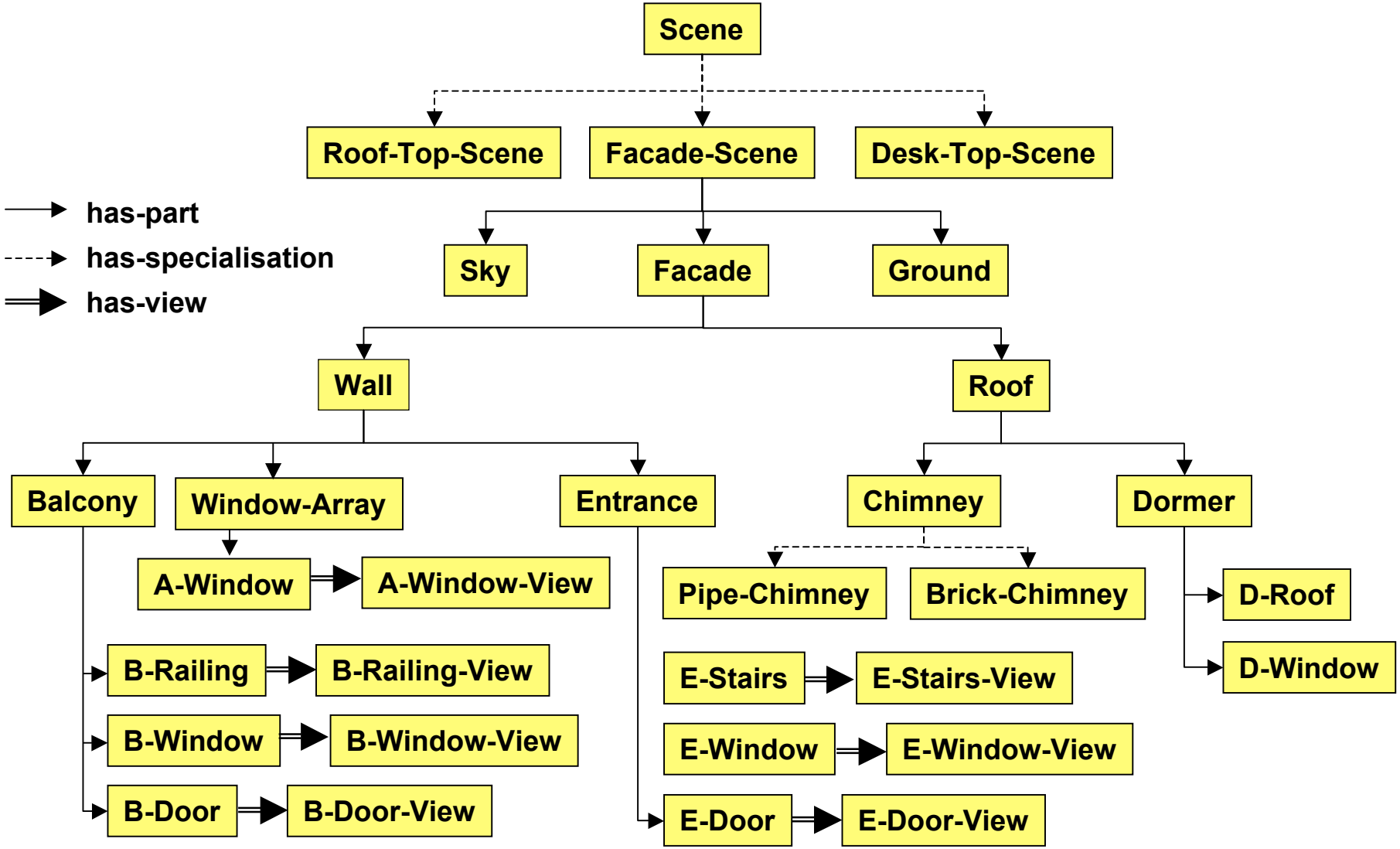


Compositional Hierarchy for Aircraft Services

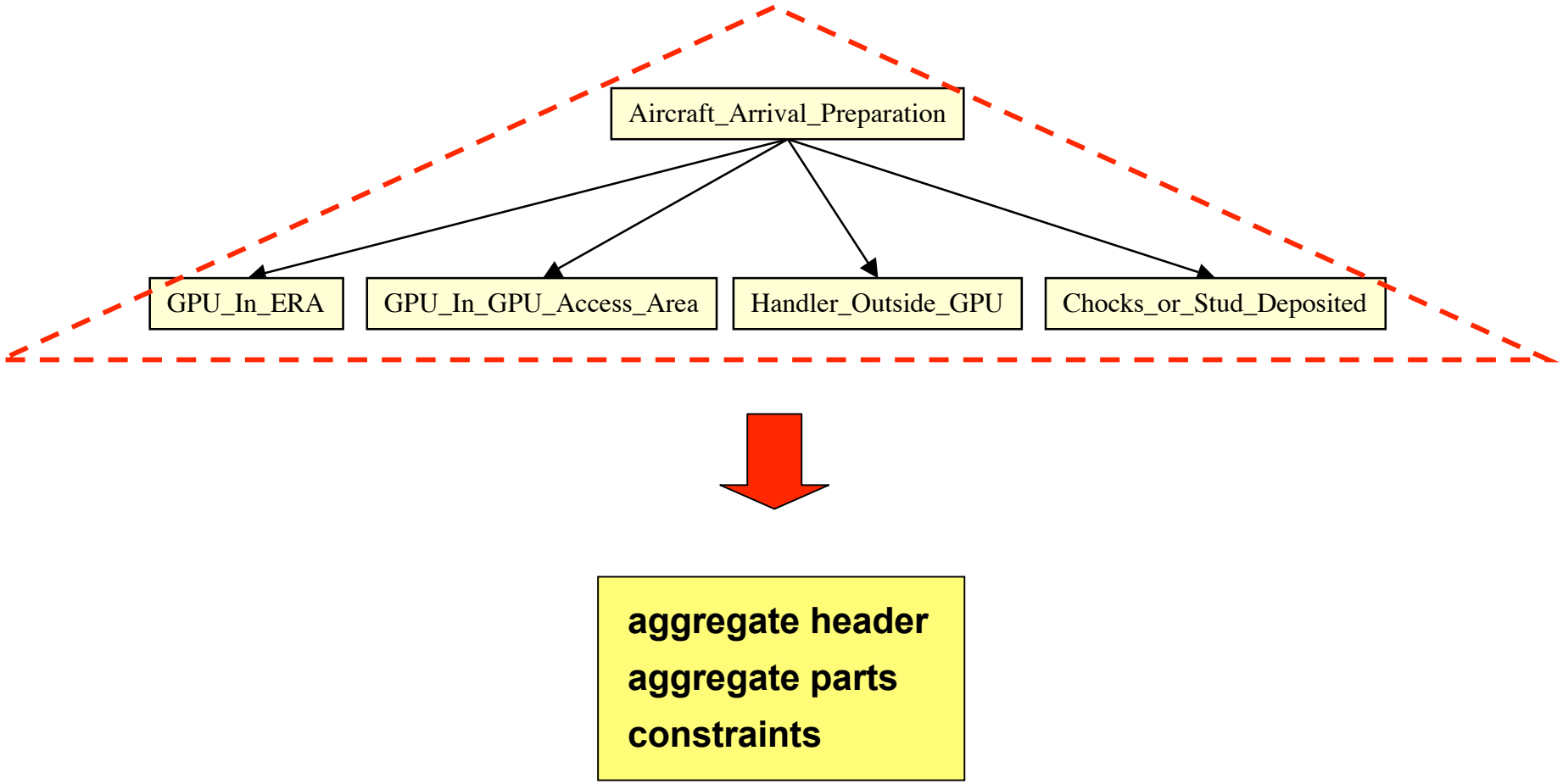
Compositional (or part-of) hierarchies are the backbone for scene interpretation



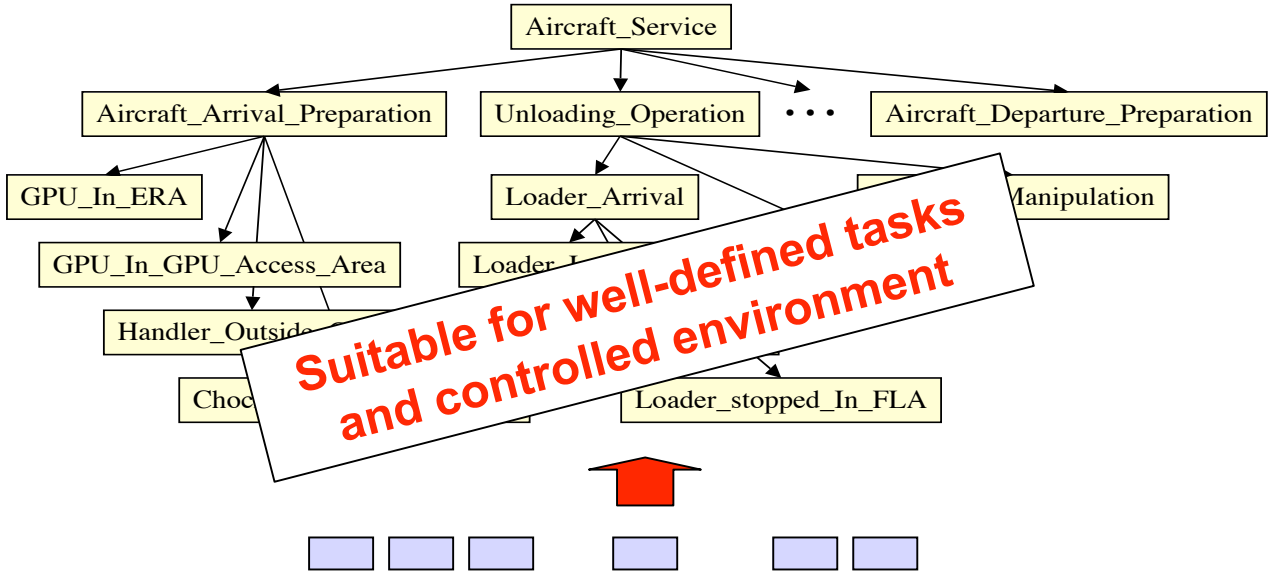
Compositional Hierarchy for Facade Scenes



Generic Object-oriented Aggregate Definitions



Simple Bottom-up Scene Interpretation



- Bottom-up image analysis provides primitive pieces of evidence
- Evidence instantiates leaves of compositional hierarchy
- Aggregates are instantiated, when all parts are instantiated and constraints are satisfies
- Interpretation is complete, when root of hierarchy is instantiated

Uncertainty Management

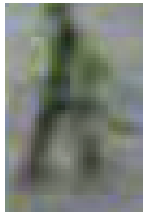
Evidence is often ambiguous or misleading:

- **Noise, occlusion, image analysis deficiencies**

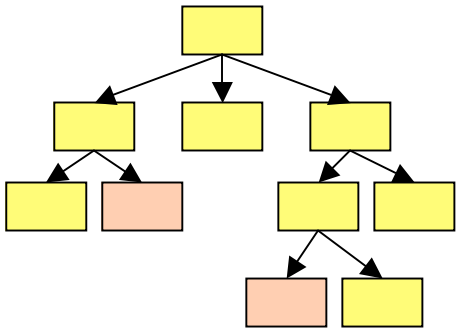
?

Transporter
GPU
Loader
Service_Vehicle
...

- **Multiple roles in scene model**



?



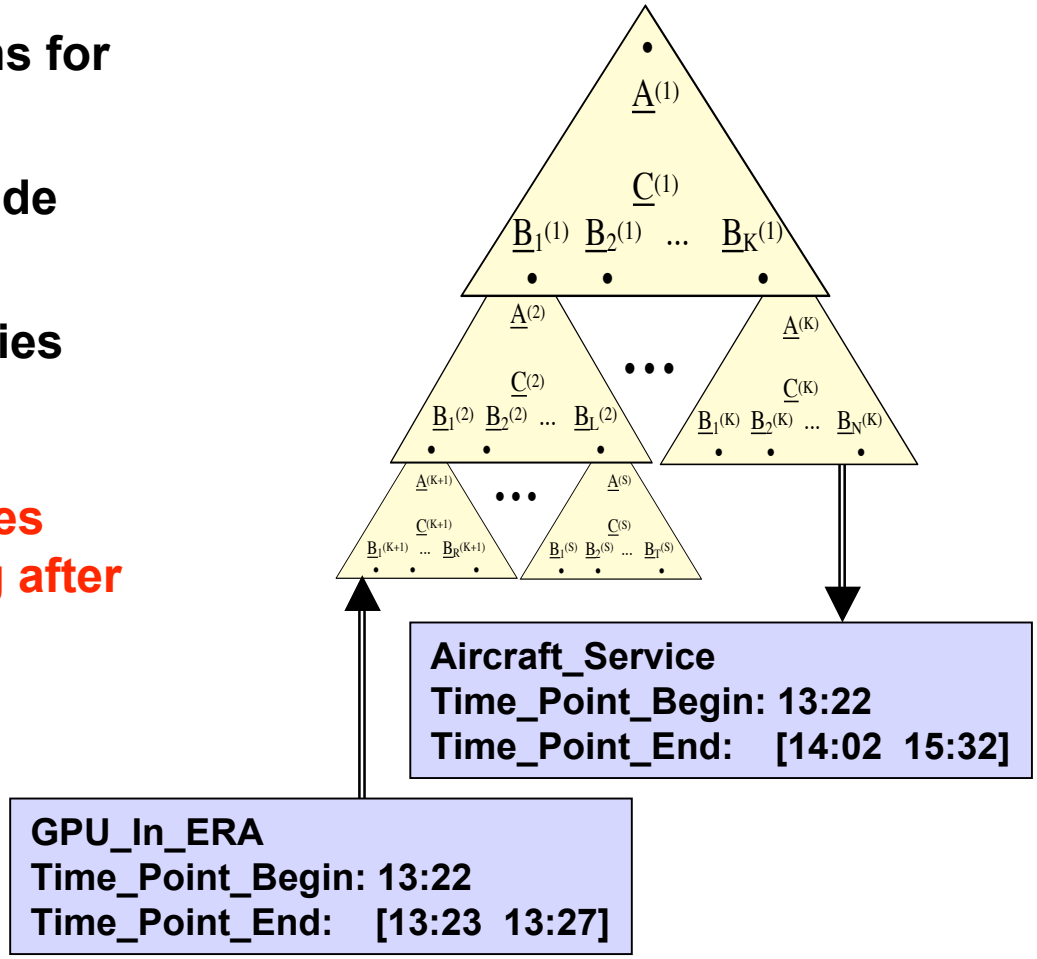
Walking_Person

- **False positives and negatives**

Probabilistic Guidance

- Provide probability distributions for aggregates in scene model
- Compute dynamic priors to guide evidence assignment
- Obtain estimates for all quantities based on current evidence

Bayesian Compositional Hierarchies allow efficient probability updating after evidence assigment steps



Learning

- **Learning methods have greatly improved in the last decade**
  - **- SVMs (Support Vector Machines) for pattern classification**
  - **- Version Space Learning for logic-based representations**
  - **- Probabilistic learning for Bayesian Networks**
- **Learning may be the only way to populate large knowledge bases**
- **Learning visual appearances may require months of continuous presentation of real-life phenomena**
- **Structural learning for high-level knowledge bases is not yet well developed**

Learning of Facade Structures

Summary

- AI contributes to virtually all technology areas
- Computer Vision and AI have to cooperate for scene interpretation
- Object Recognition has been improved but remains a bottleneck
- The structure of high-level scene models is well understood, but building extensive knowledge bases including common sense knowledge remains a problem
- Semantic Web technologies provide a welcome standard for ontology representation and reasoning


Thank you for your attention !