Visual Object Recognition

Lecture 1: Introduction

Per-Erik Forssén, docent
Computer Vision Laboratory
Department of Electrical Engineering
Linköping University

Lecture 1: Introduction

- Visual Object Recognition
  what is the problem?

- Terminology and Taxonomy
  Classification, Categorisation, Recognition, Detection, Pose estimation, Articulation, Expression, Feature extraction, Learning.

- About this course
  lecture format, projects, exam

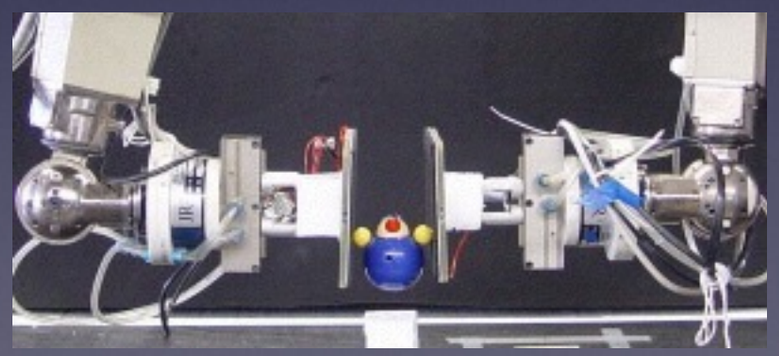
Recognition

- In humans, recognition is a holistic process that involves all senses/modalities.
- **Vision**, sound, touch, taste, smell, body sense.

Recognition

- In humans, recognition is a holistic process that involves all senses/modalities.
- **Vision**, sound, touch, taste, smell, body sense.

- Recognition in computers typically involves just one modality (or a few).

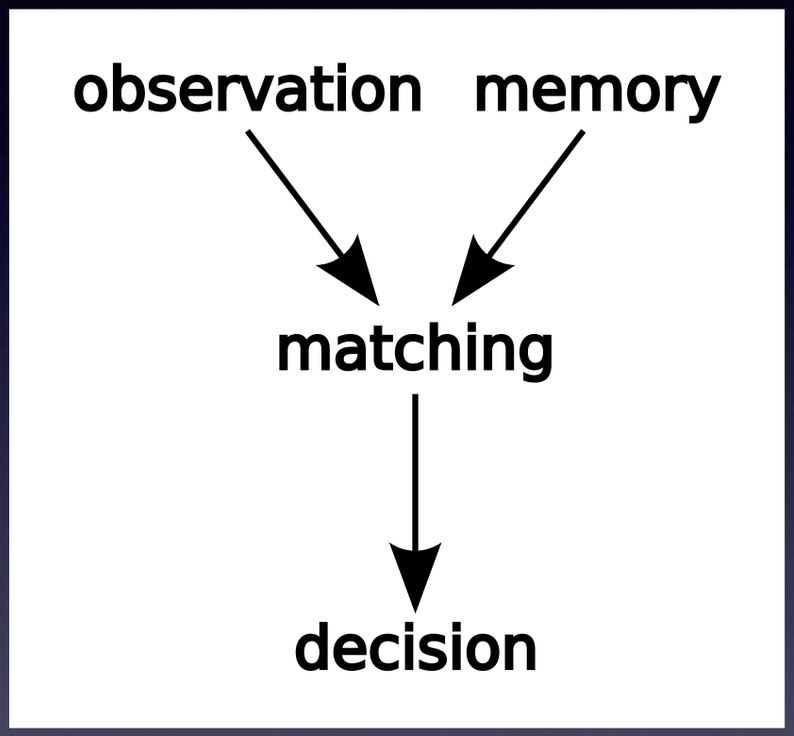


tactile recognition
Neuroinformatics at Uni. Bielefeld

sound recognition
Shazam app

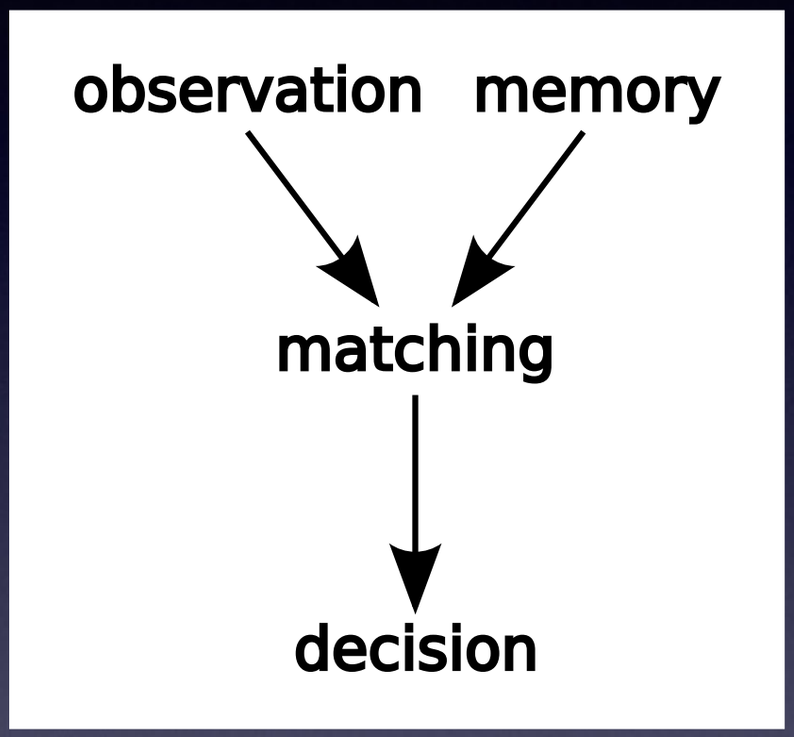
visual recognition
Google goggles app

Recognition



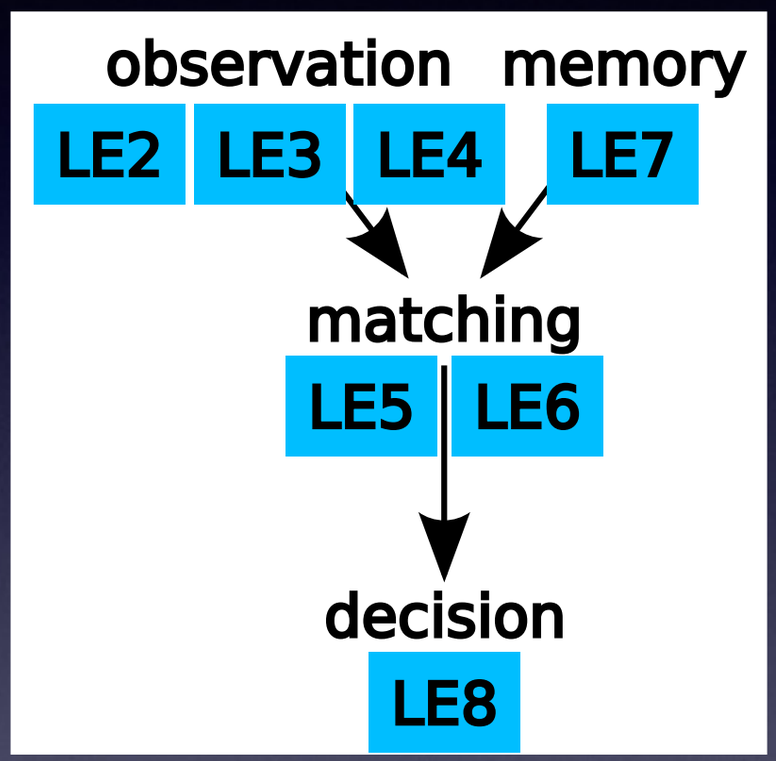
- **Recognition** is a comparison of an observation with what is stored in memory.

Recognition



- **Recognition** is a comparison of an observation with what is stored in memory.
- **Feature extraction** constructs an observation
- **Learning** constructs the memory

Recognition



- **Recognition** is a comparison of an observation with what is stored in memory.
- **Feature extraction** constructs an observation
- **Learning** constructs the memory

Visual Object Recognition

- OR happens very quickly in the human visual system. Bottom up process takes less than 150ms (S. Thorpe et al. 1996).

Visual Object Recognition

- OR happens very quickly in the human visual system. Bottom up process takes less than 150ms (S. Thorpe et al. 1996).
- This has evolutionary reasons...

Visual Object Recognition

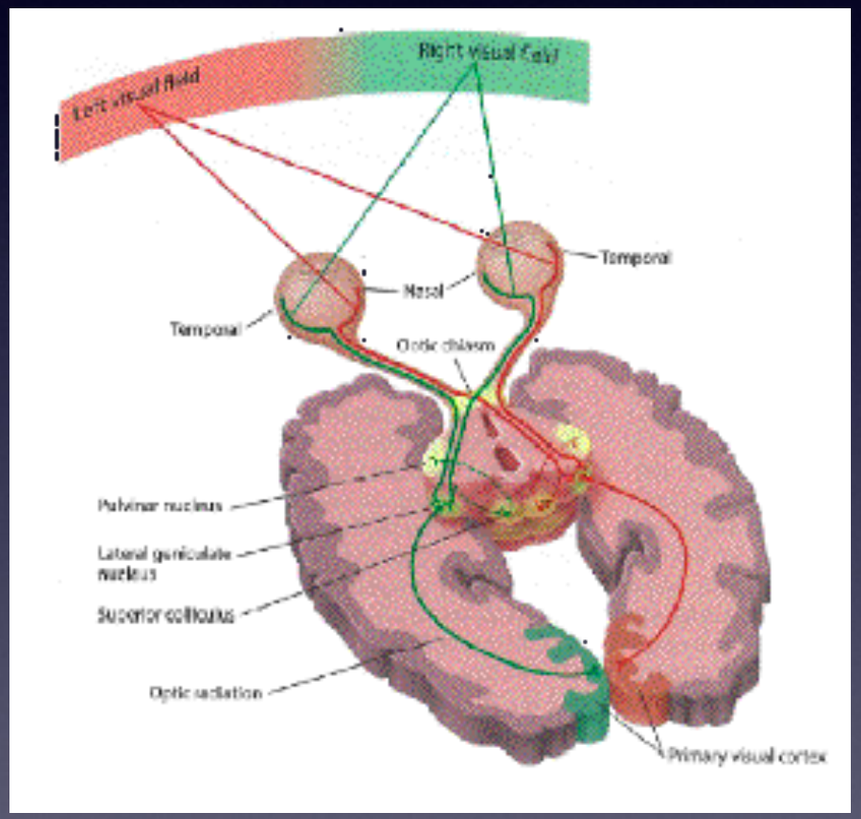
- OR happens very quickly in the human visual system. Bottom up process takes less than 150ms (S. Thorpe et al. 1996).
- This has evolutionary reasons...
- Since it is a pre-conscious process in our brains, we do not intuitively think of object recognition as being difficult.

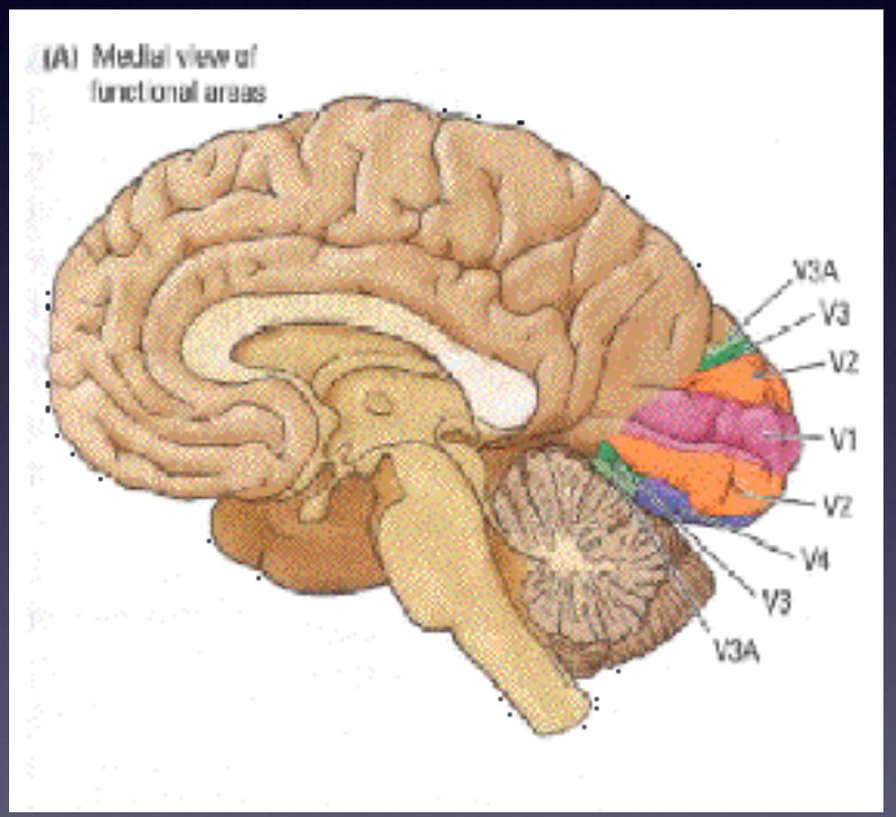
Moravec's Paradox

- Initially in AI computer vision was assumed to be simple, and logical inference hard.
- "Just detect the objects in an image and generate the appropriate symbols"
- Only symbolic reasoning and and inference was seen as proper AI problems.
- Now we know that computer vision is much more complex than logical inference.

Visual Object Recognition

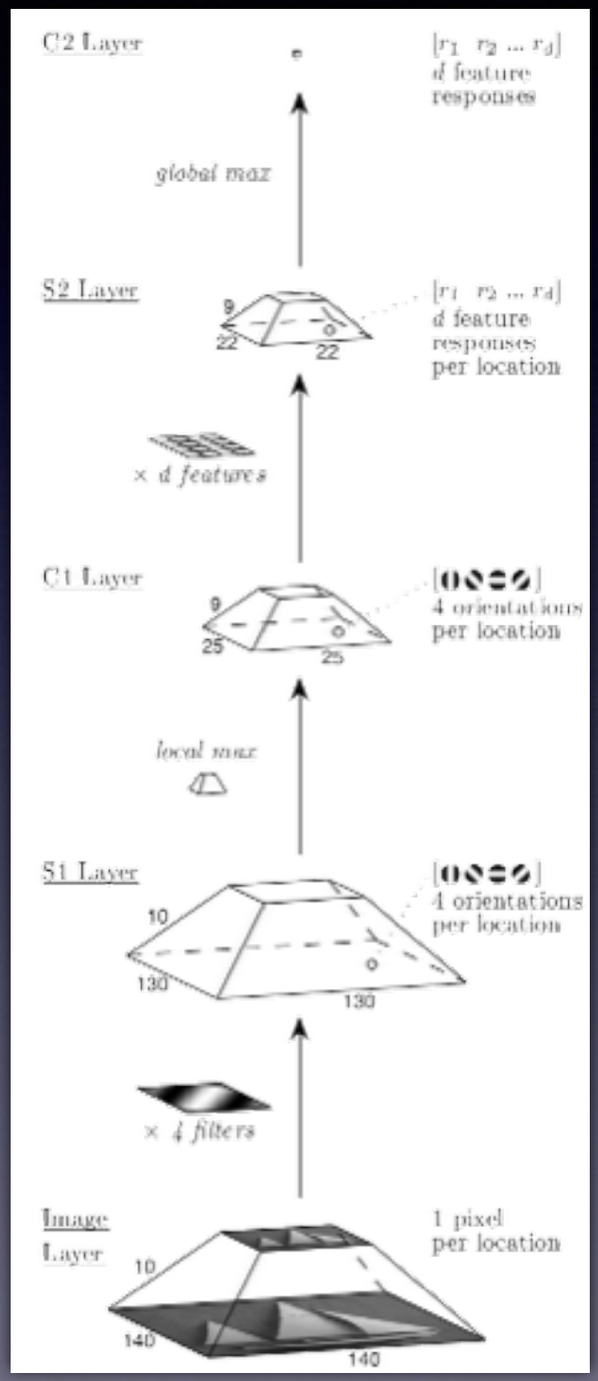
- von Neumann vs. Biological architectures





- Visual cortex, eyes, muscles, cerebellum, mid-brain.

Visual Object Recognition



Mutch&Lowe CVPR'06

- The "Standard Model", Riesenhuber&Poggio, Nature Neuroscience vol.2 no.11, 1999
- Alternating template matching and local max operations.
- Decreasing spatial resolution, increasing number of feature types
- Perception only, no motor functions (head&eye movements)

What do we mean by Visual Recognition?

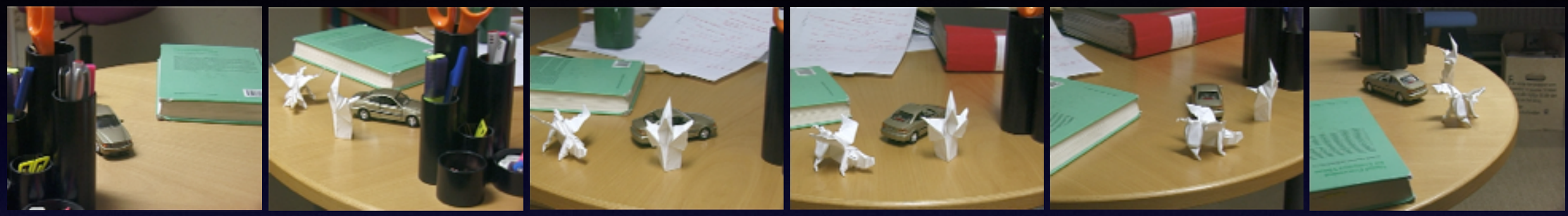
- The same object instance?
- The same class? category?
- The same pose?
- The same articulation? expression?

Johansson&Moe CRV05

Lowe PAMI91

www.cwu.edu/~warren/

Object instance recognition



Johansson&Moe CRV05

- Recognition of the same object
- Different view/pose (also pose estimation)
- Different illumination
- The same articulation/expression

Object instance recognition

- Application: Pose estimation for bin picking
    - Several identical instances of the object need to be distinguished

Random Bin Picking

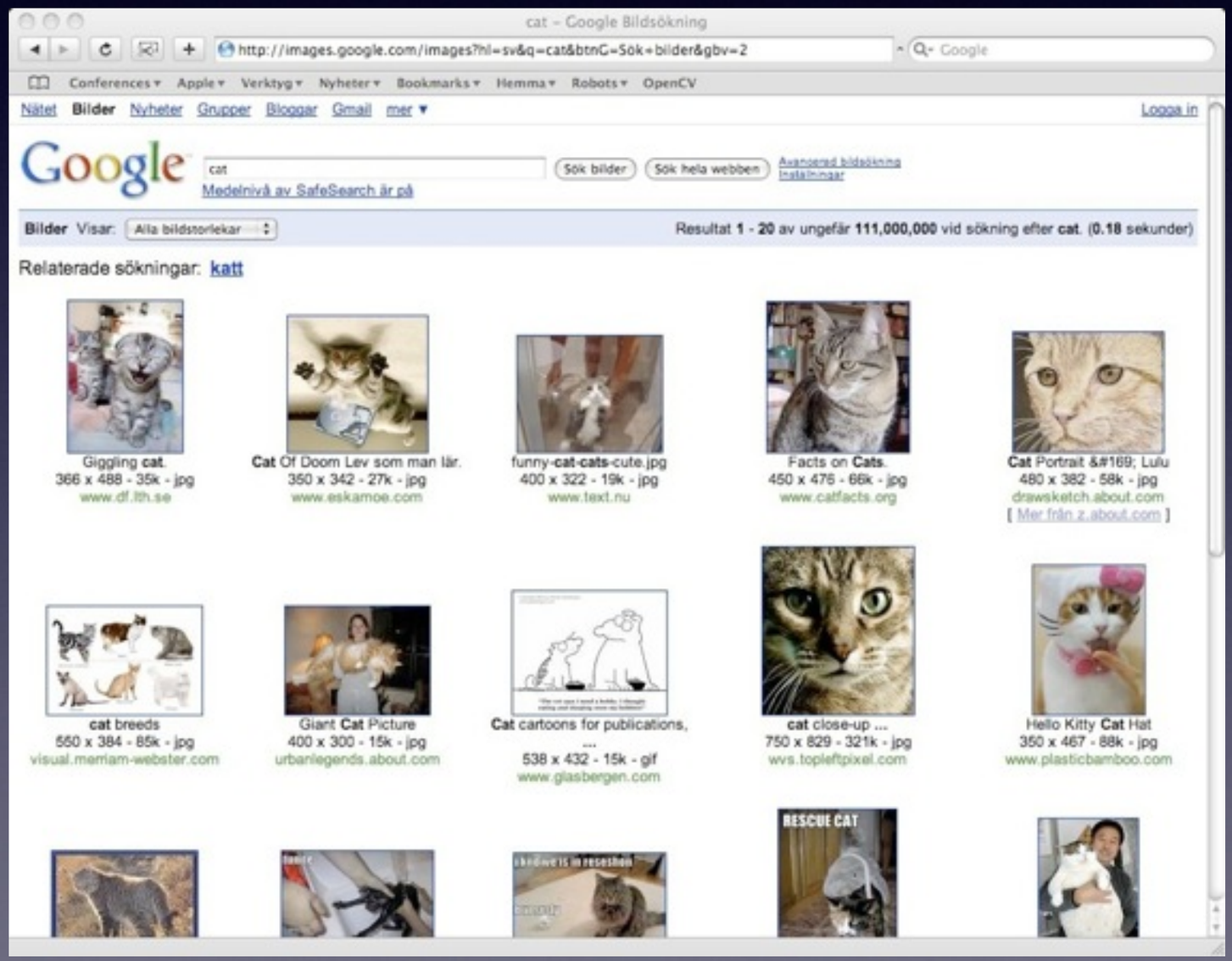
http://www.youtube.com/watch?v=09LzufOnbX0

Object class recognition

- Recognition of an object class/category
  - Different instance
  - Different view/pose
  - Different illumination
  - Different articulation/expression

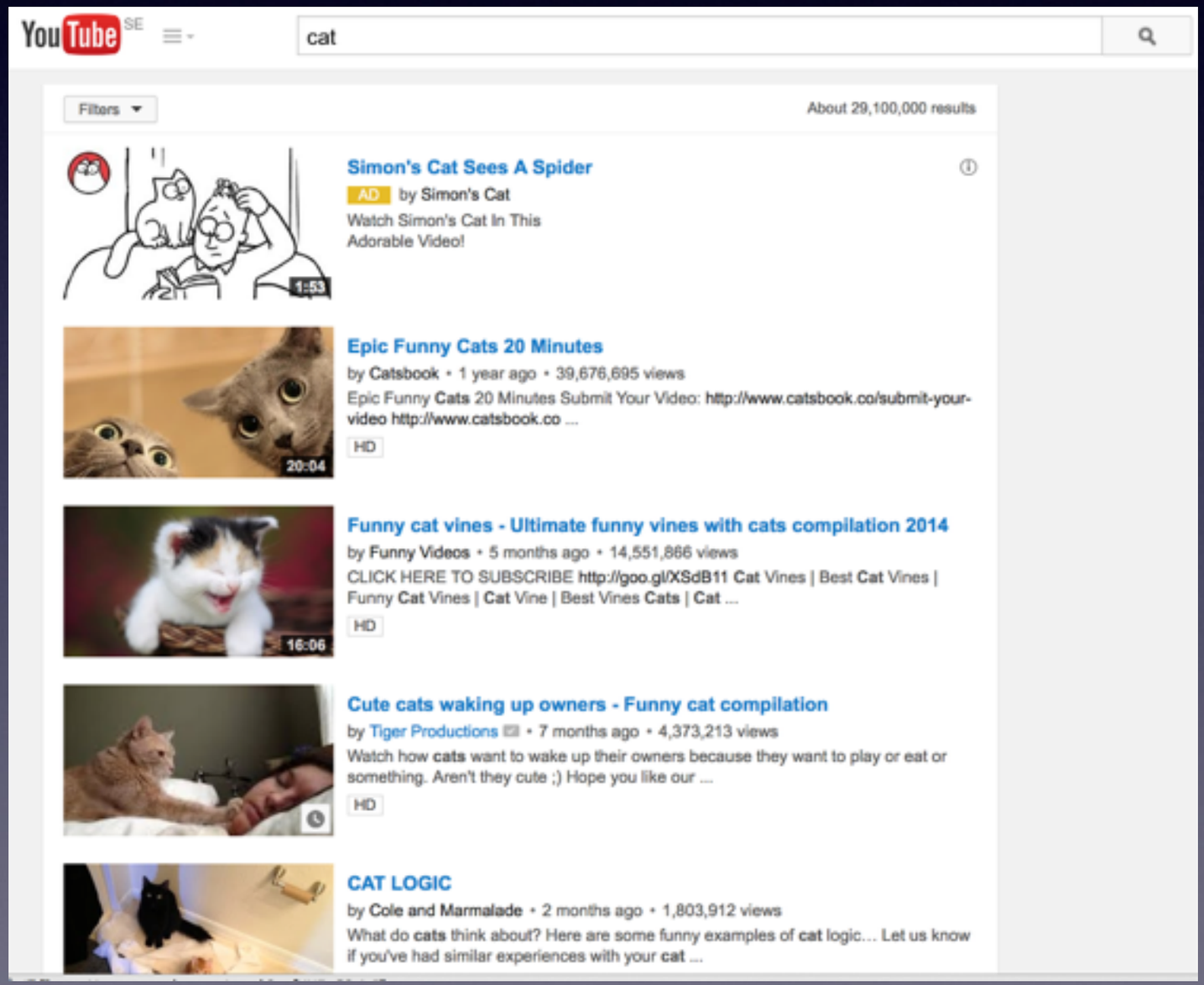
Object class recognition

- Main application is image database search:



Object class recognition

- Now also video database search:

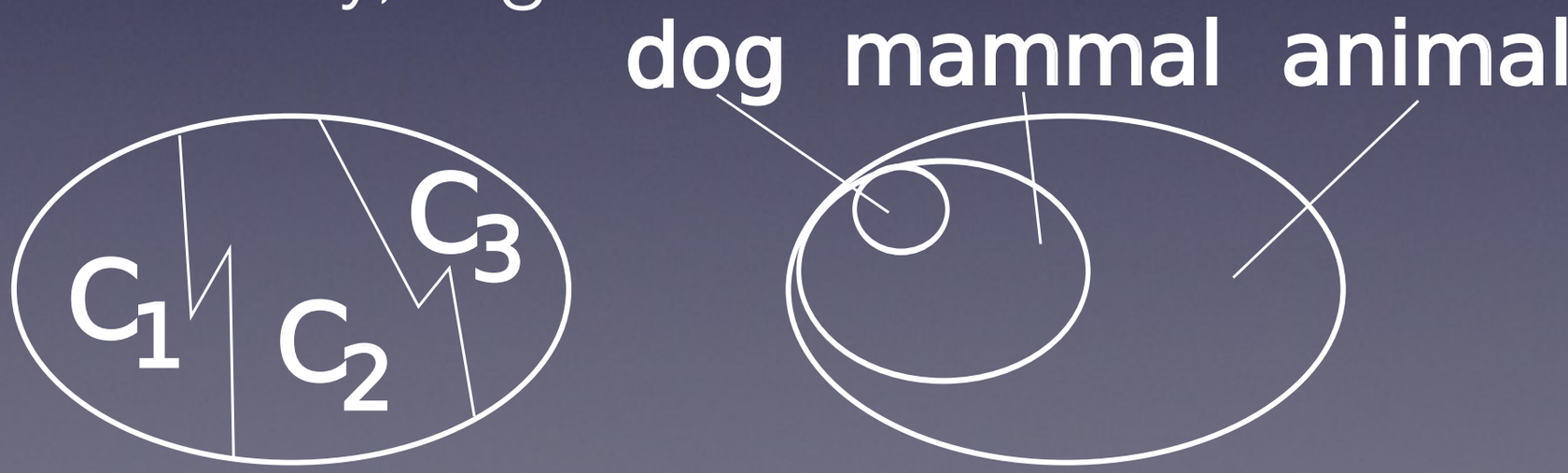


Object class recognition

http://www.semantic-robot-vision-challenge.org/

Classes and Categories

- Object class is a computer science construct
- Implicit assumption: It is possible to partition a dataset into disjoint classes
- This fits poorly to the structure of natural language where categories are often nested hierarchically, e.g.



Classes and Categories

- Natural categories tend not to be defined by appearance alone.
- Applicable actions also matter

e.g. a “chair” is something you sit on. Number of legs, colour etc. does not matter.

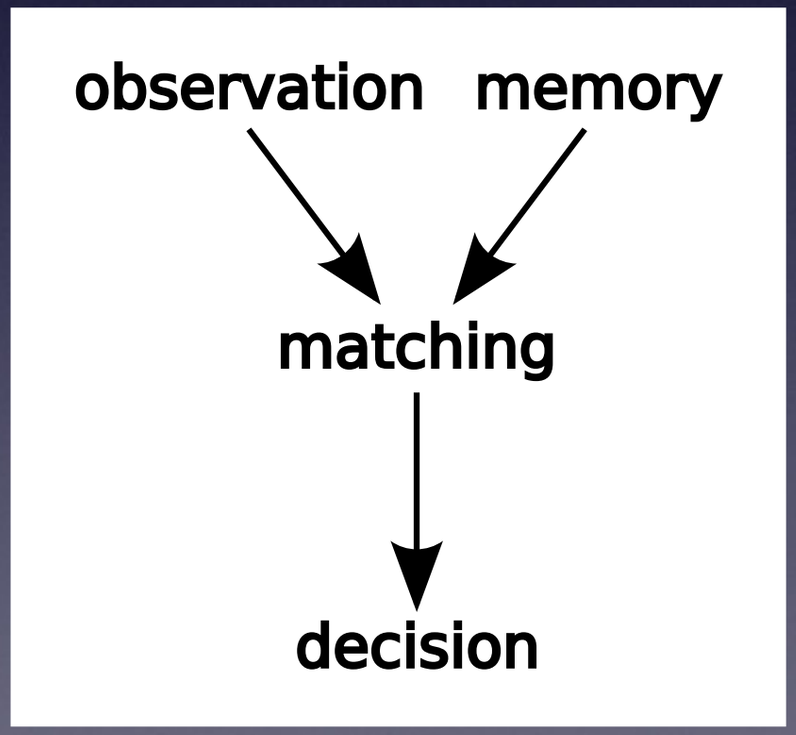
Classes and Categories

- A category member is instead recognized as being similar to one of possibly several prototypical category members.
- Category membership is not a binary decision.
- Lakoff, George, “Women, Fire, and Dangerous Things - what categories reveal about the mind”. University of Chicago Press. 1987

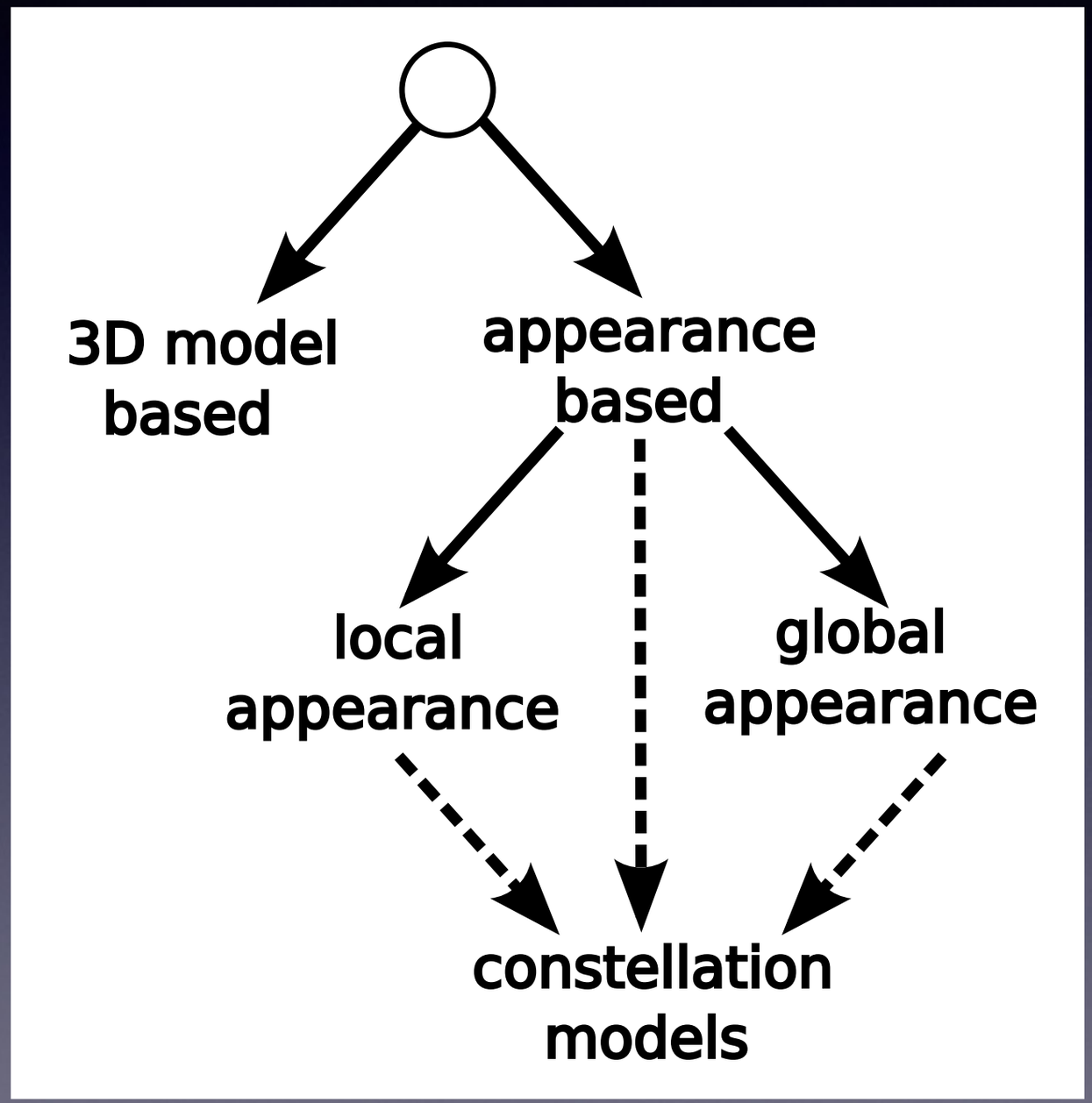
Embodied Recognition Systems

CVL research project EVOR

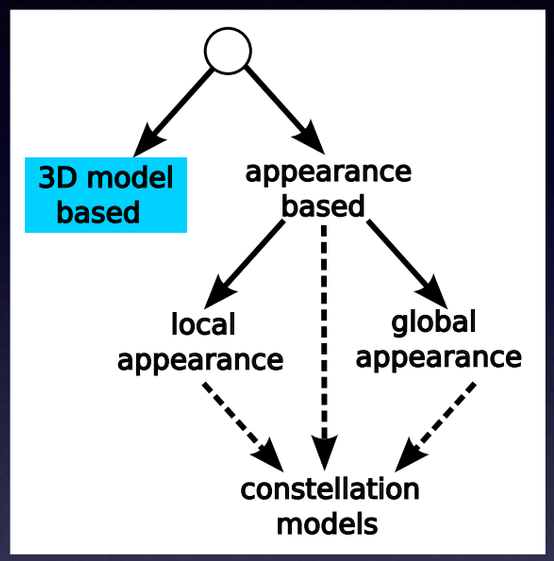
- An embodied system arranges its own **memory** and collects its own **observations**



Taxonomy of Recognition Approaches

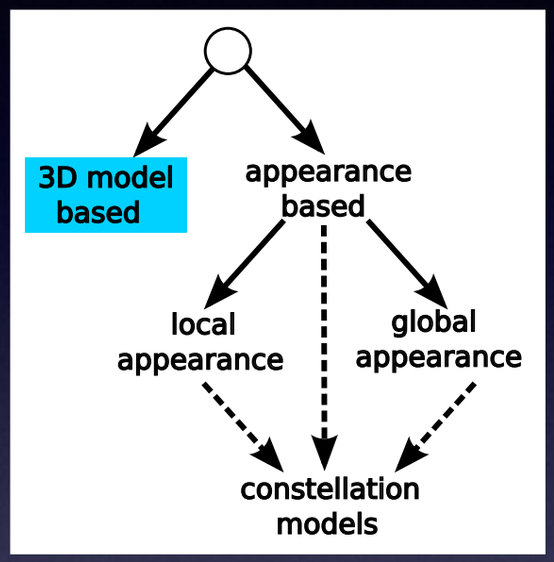


3D model based



- Memory: A 3D model generated from multiple views for each object
- Observation:
  A new 3D model

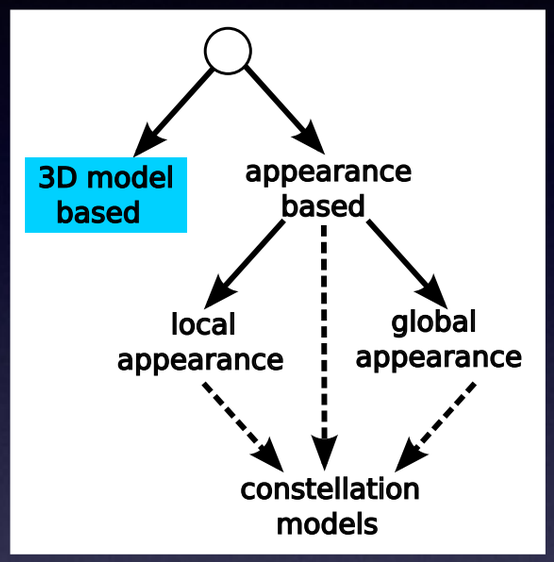
3D model based



- Memory: A 3D model generated from multiple views for each object
- Observation: A new 3D model

A.S. Mian et al. TPAMI 2006

3D model based



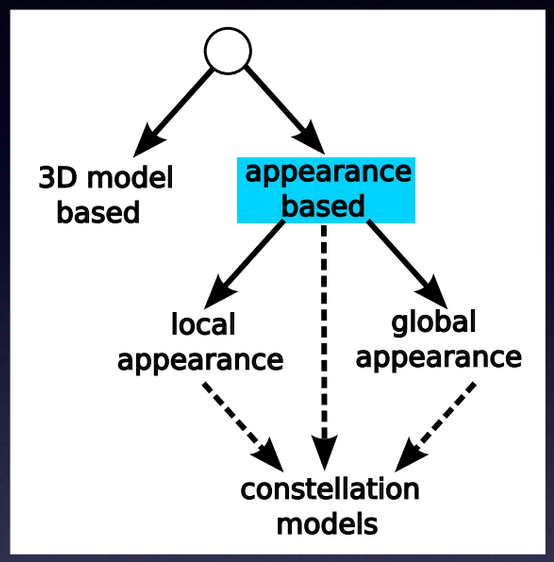
- Memory: A 3D model generated from multiple views for each object
- Observation:
  A new 3D model

+ Illumination invariant

- No abstraction, exact matches only

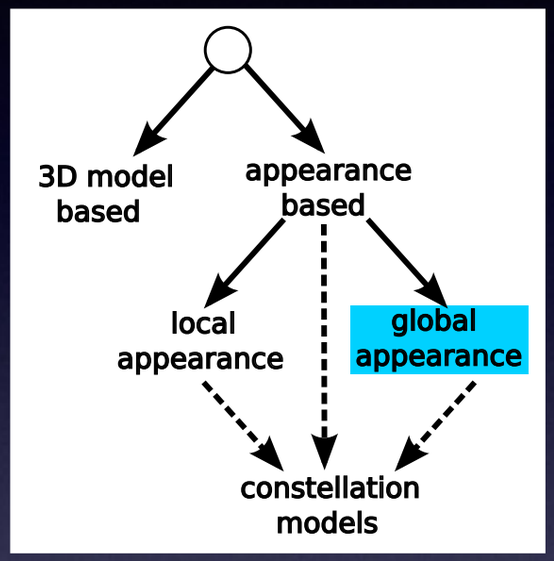
Used in some machine vision applications, e.g. bin-picking

Appearance based

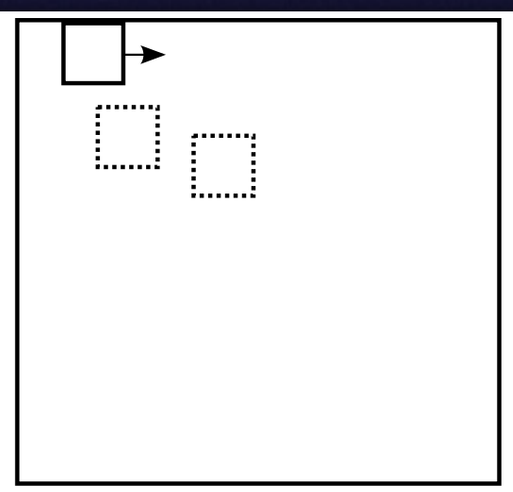


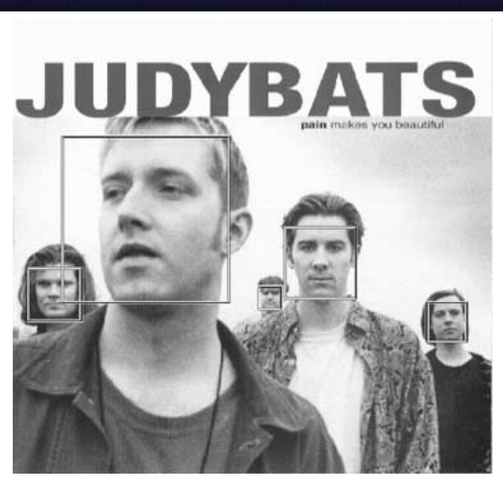
- Used by HVS
- **Global appearance**: recognition of a visual pattern
- **Local appearance**: recognition of many small patterns
- **Constellation models**: recognition of many small patterns and their arrangement

Global appearance

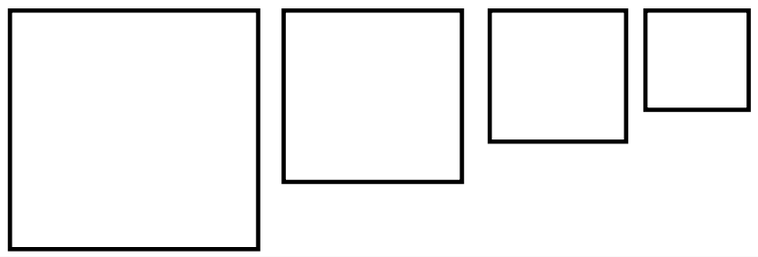


- E.g. Run a fast pattern recognition algorithm as a sliding window detector.

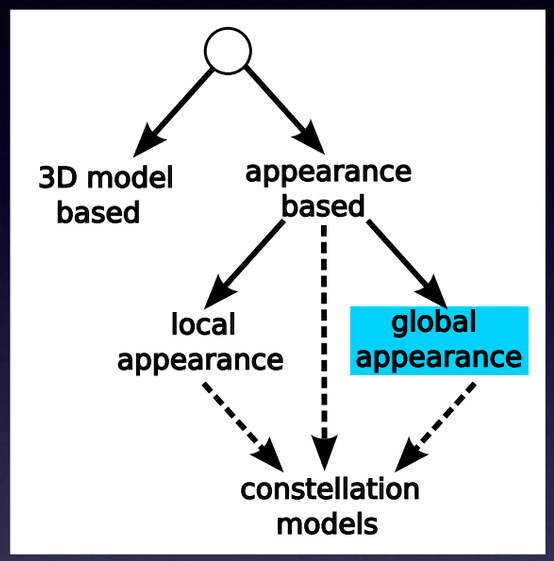




Scale pyramid for scale invariance

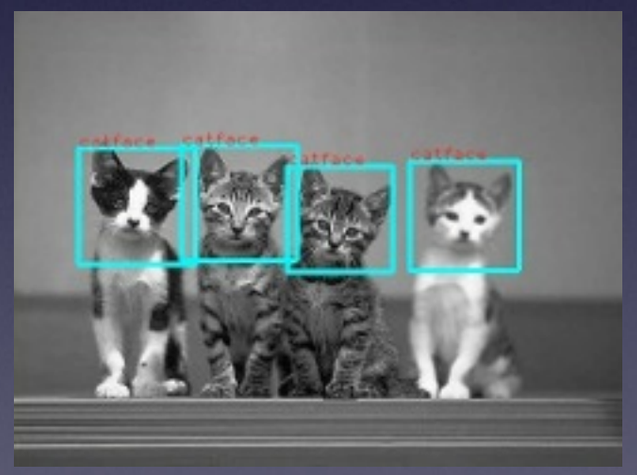


Global appearance



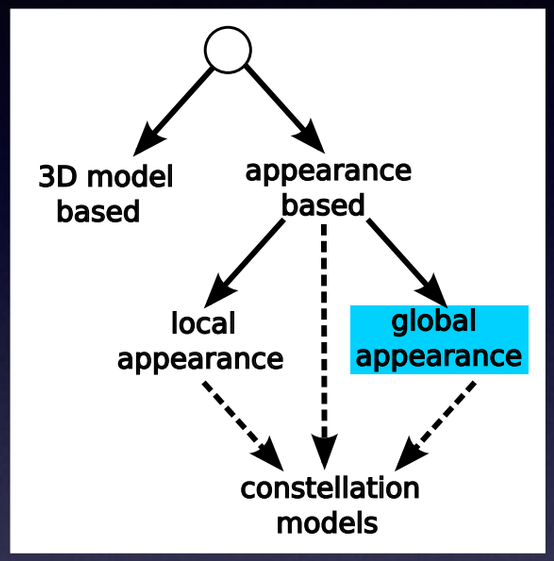
- E.g. Run a fast pattern recognition algorithm as a sliding window detector.
- Cascaded face **detection**

Viola&Jones IJCV'04



Ivan Laptev

Global appearance

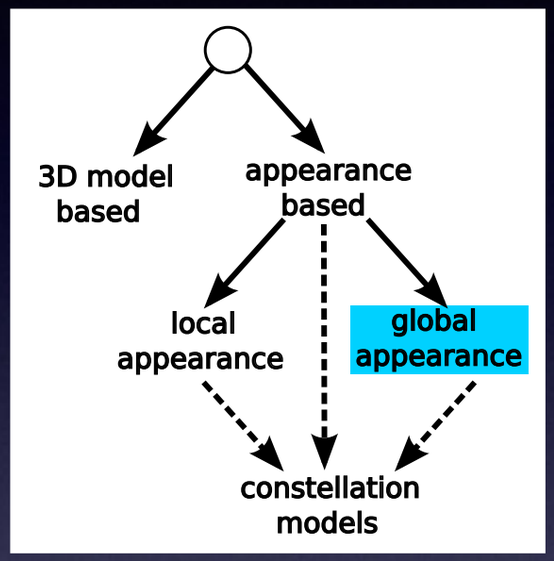


- Application: Cascaded face **detection** for autofocus

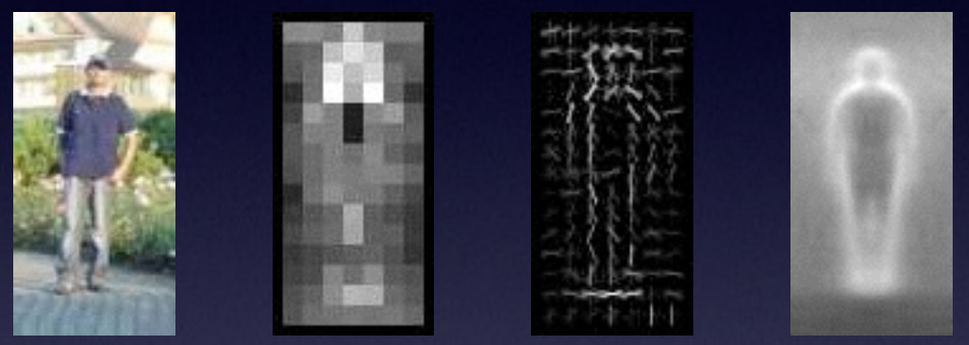
www.adorama.com review of Fujifilm finepics F40fd

- Also preprocessing step in face **recognition**

Global appearance

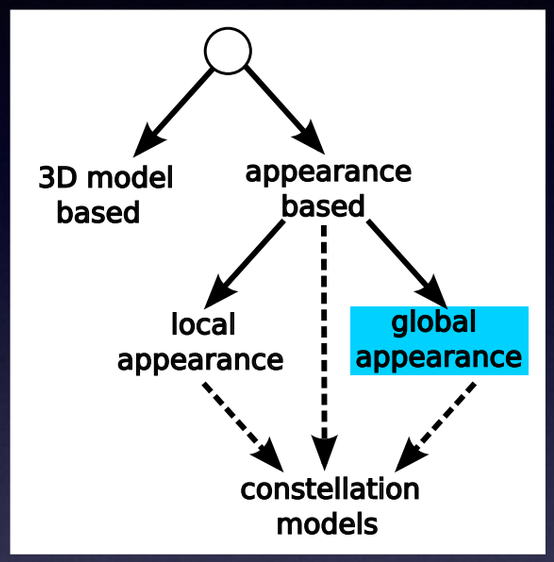


- Pedestrian detection

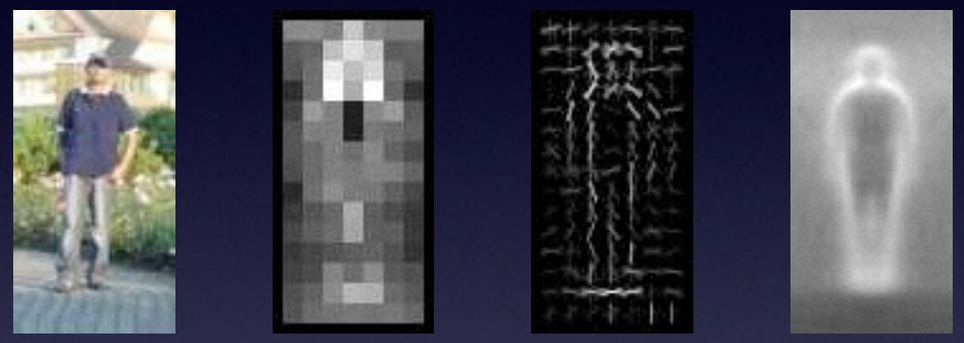


Dalal&Triggs, CVPR'05
Histograms of Oriented Gradients for Human Detection

Global appearance



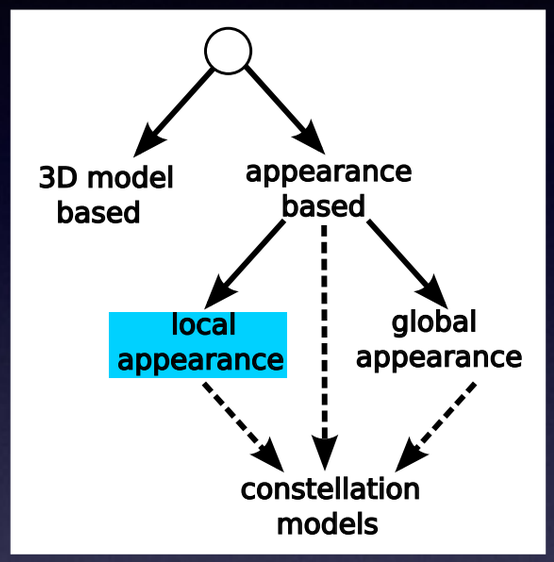
- Pedestrian detection



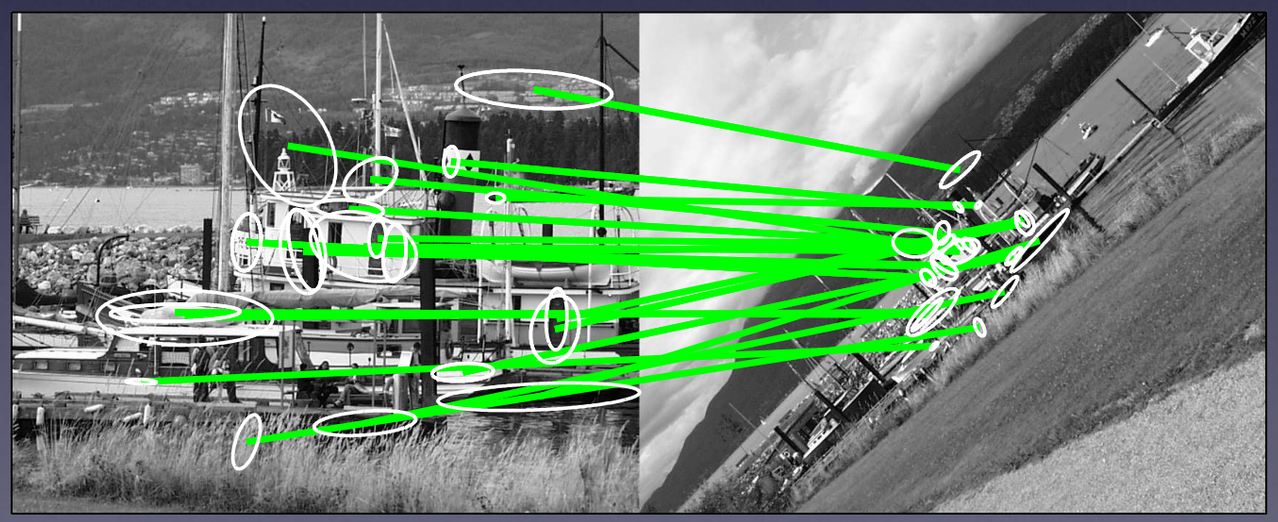
Dalal&Triggs, CVPR'05
Histograms of Oriented Gradients for Human Detection

- Nowadays constellation models are used E.g. Felzenswalb et al., CVPR'08, "A discriminatively trained, multiscale, deformable part model"

Local appearance

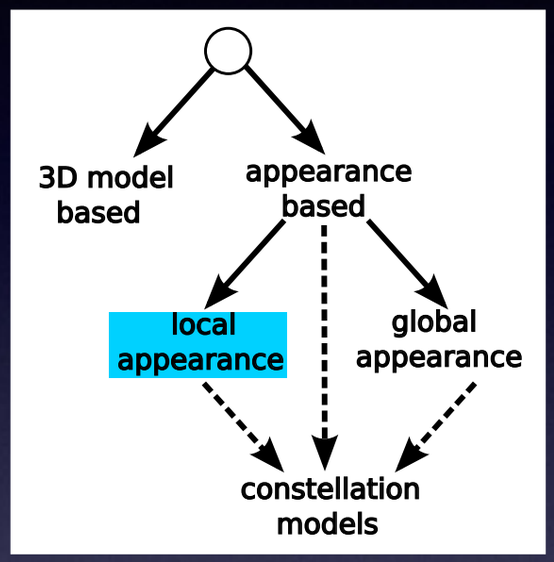


- Detect local invariant frames and cut out many patches.
- Try to match all patches in image to all patches in memory.



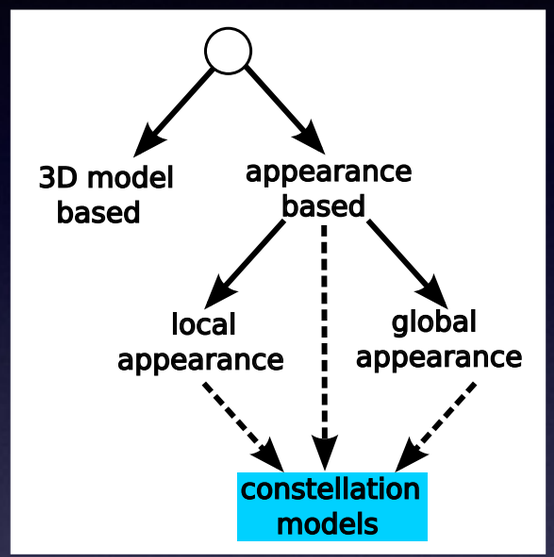
Forssén, Lowe ICCV07

Local appearance

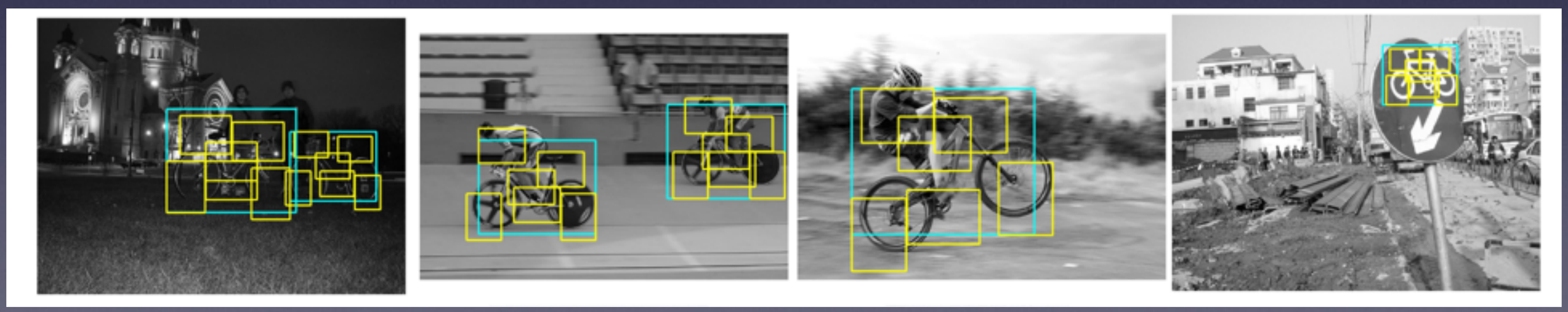


- Bag-of-features / Bag-of-words
- Crude, so often combined with verification to remove false matches.
- Can handle occlusion and articulation
- Scales to video.
- 3D models may be built ***after*** recognition

Constellation models



- A combination of local and global methods.
  1. A coarse global model
  2. A fixed number of part models with flexible spatial arrangement.



Felzenswalb et al., CVPR'08, "A discriminatively trained, multiscale, deformable part model"

Global appearance models

- Good for non-articulate objects and objects with small articulation.
- Pose can be dealt with by running one detector for each pose.
  E.g. for faces: frontal, left side, right side.
- Handles lower resolution than local appearance models.
- Occlusion is problematic.
- Large training data sets are needed.

Constellation models

- Improves on global models to better handle articulation and moderate pose changes.
- Handles lower resolution than local appearance models.
- Occlusion is problematic.
- Large training data sets are needed.

Local appearance models

- Can handle occlusion
- Deals with rotations, scale changes, and affine distortions.
- Can handle large view changes, 25-60deg, depending on what is imaged
- Requires higher resolution than sliding window approach

Course Format

- Two options

    1. For 6 PhD course credits you need to participate actively in the paper discussion, and do the final exam. If you also do the project you get an extra 2hp.

    2. If you are not a doctoral student you skip the project, and the paper discussion. Just follow the lectures. You should still read the papers of course.

Course Format

- The papers:
  - Each lecture has an associated paper, chosen both for content and readability.
  - The paper should be read in advance
  - PDFs of papers will be available on the course web page:

    http://www.cvl.isy.liu.se/Education/Graduate/VOR14/articles/

Course Format

- The Lectures, preparation

  1. Read the paper thoroughly

  2. Make notes of related questions and issues you want to discuss

  3. Each participant should prepare at least two issues/questions for each lecture

Course Format

- The Project
  - For 8hp you are also expected to do a small programming project
  - You are encouraged to suggest your own project.
  - A list of possible other project topics will be handed out later.
  - Duration should be approx 2 weeks including the writing of a small report.

Course Format

- The Exam
  - The course will end with a written exam
  - If 4 people or fewer, possibly an oral exam
  - Be prepared to answer questions about concepts and algorithms introduced in the course.

Schedule

- Is shedule after Christmas OK?
- Paper to read for LE2 is:

  M. Brown, D. Lowe, "Invariant Features from Interest Point Groups", BMVC 2002

- Paper for LE3 will be decided before LE2.

Summary

- Recognition is matching between observations and memory
- Visual recognition works with a local to global principle
- Most recognition approaches are view based / appearance based. NOT 3D
- Depending on application, recognition means different things