Visual Object Recognition

State-of-the-art techniques and performance evaluation

Lecture 1: Introduction

- **Visual Object Recognition**
  what's the problem?

- **Terminology and Taxonomy**
  Classification, Categorisation, Detection, Pose estimation, Articulation, Expression.

- **About this course**
  lecture format, projects, exam

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).

- For 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).

- For evolutionary reasons.

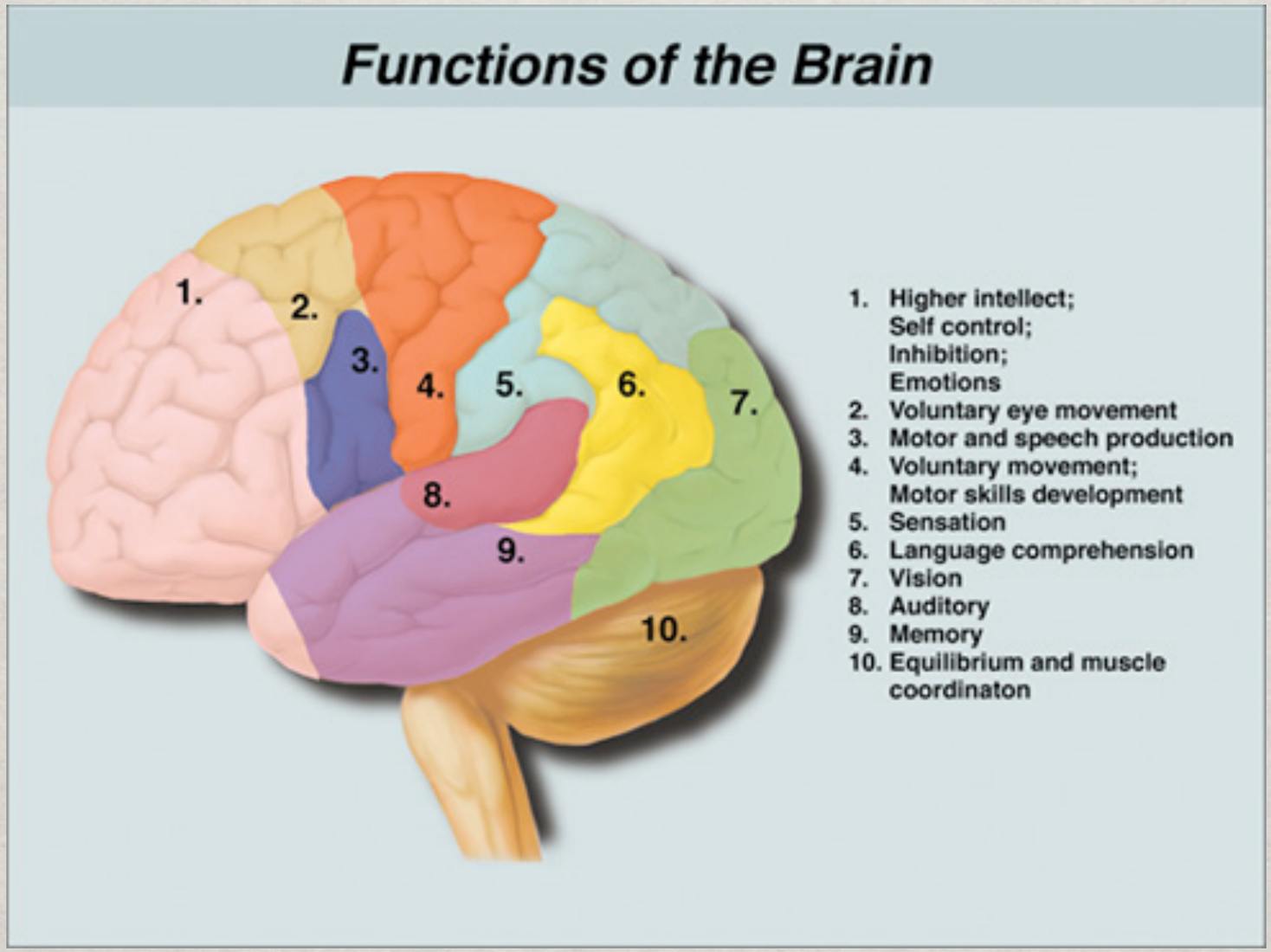
- Since it is a pre-conscious process in our brains, we do not intuitively think of object recognition as being difficult.

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.

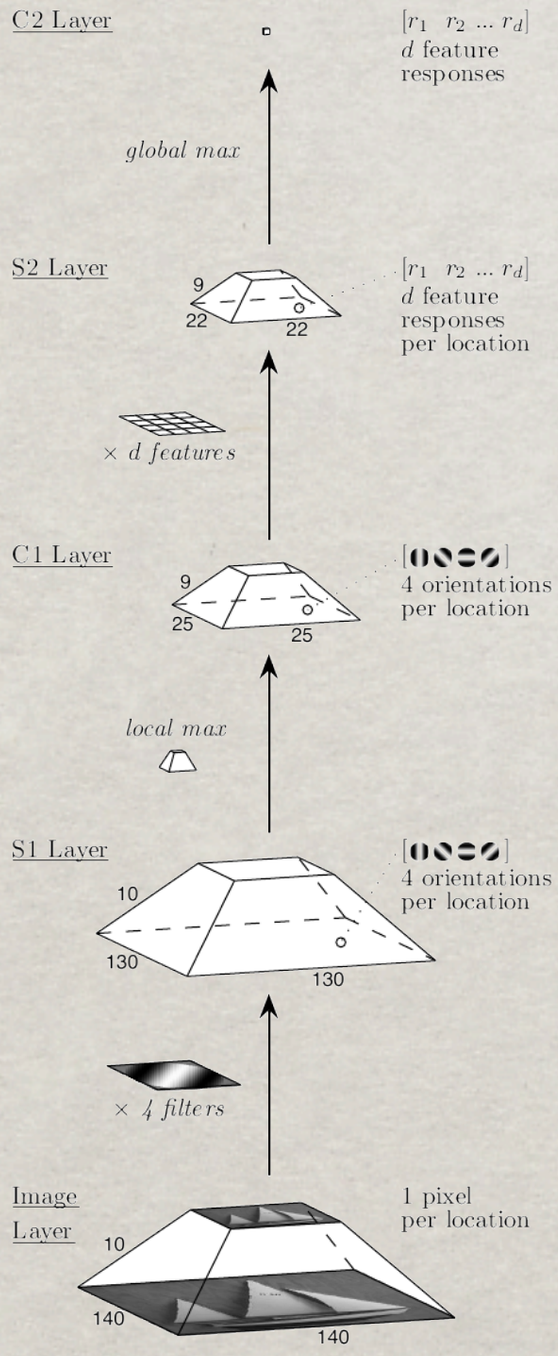
Visual Object Recognition

- von Neumann vs. Brain architectures



www.braininjurylegalguide.com

Visual Object Recognition



Mutch&Lowe CVPR'06

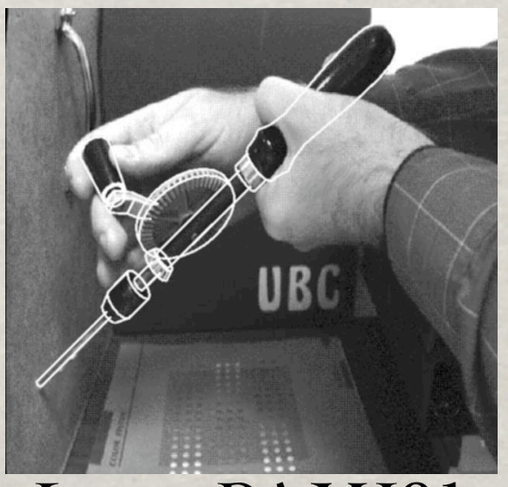
- The "Standard Model", Riesenhuber&Poggio, Nature Neuroscience vol.2 no.11, 1999
- Alternating *template matching* and *local max* operations.
- Decreasing spatial resolution, increasing number of feature types
- Perception only, no motor functions (head&eye movements)

What do we mean by Visual Recognition?

- The same **object instance**?
- The same **class**? **category**?
- The same **pose**?

- The same **articulation**? **expression**?

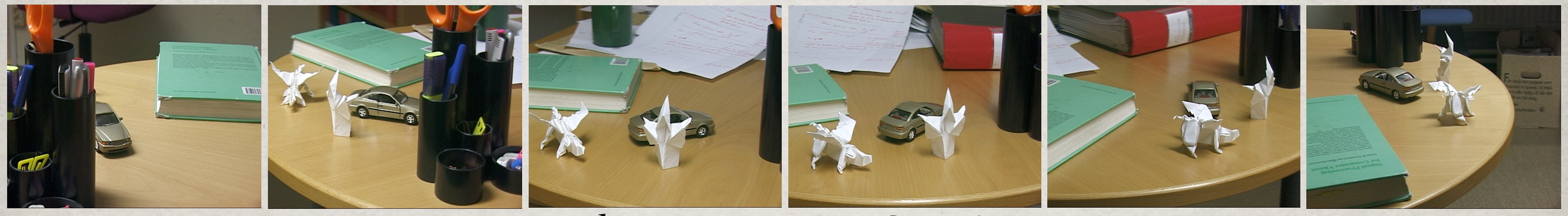
Johansson&Moe CRV05



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

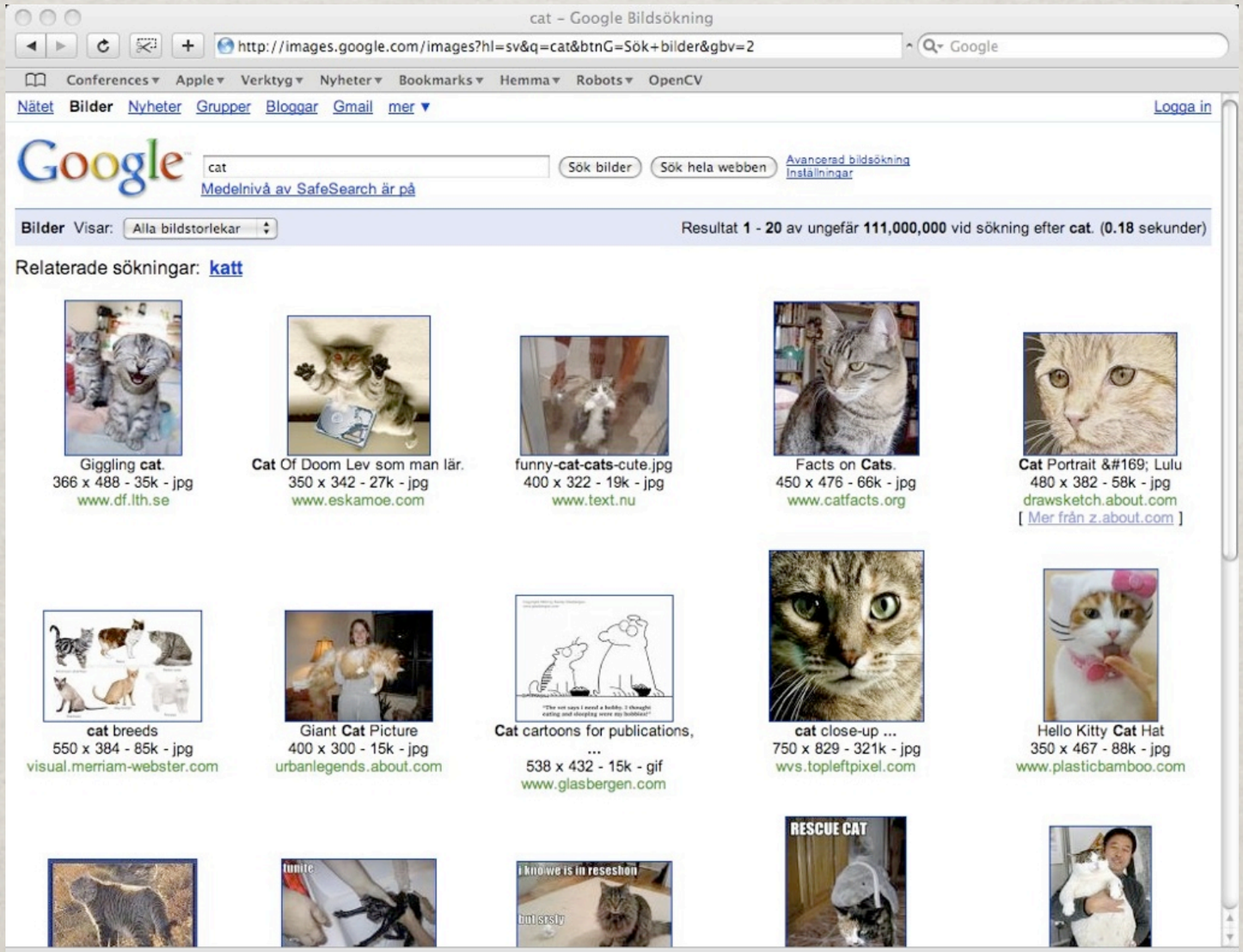
http://www.braintech.com/videos-rbp.php

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:



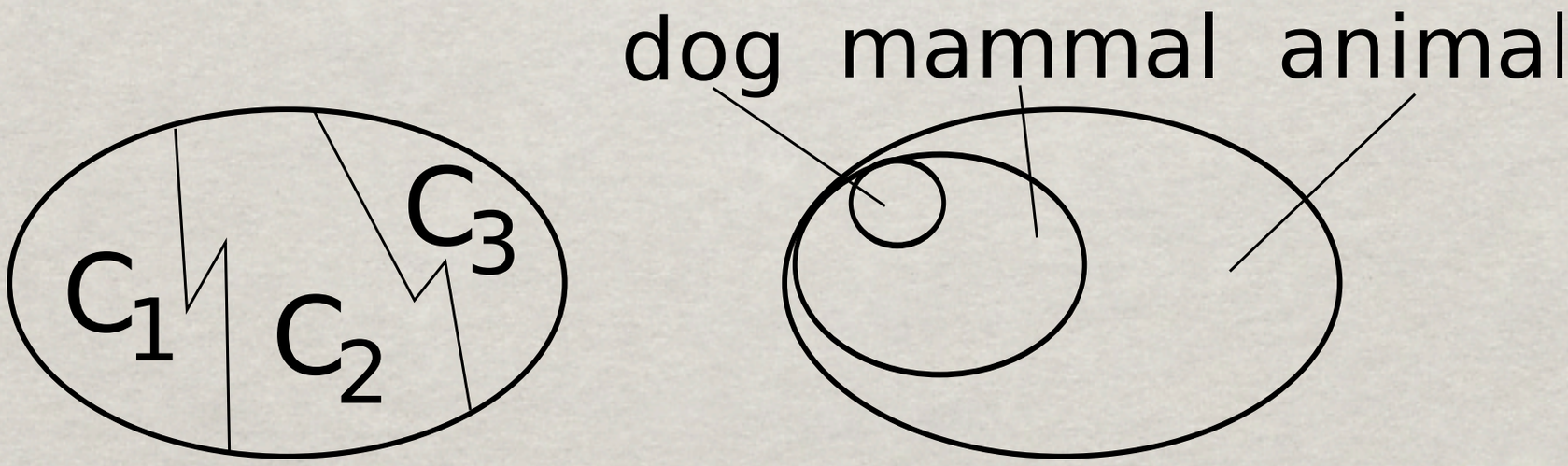
Object Class recognition

- Other application "semantic robot vision":

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.

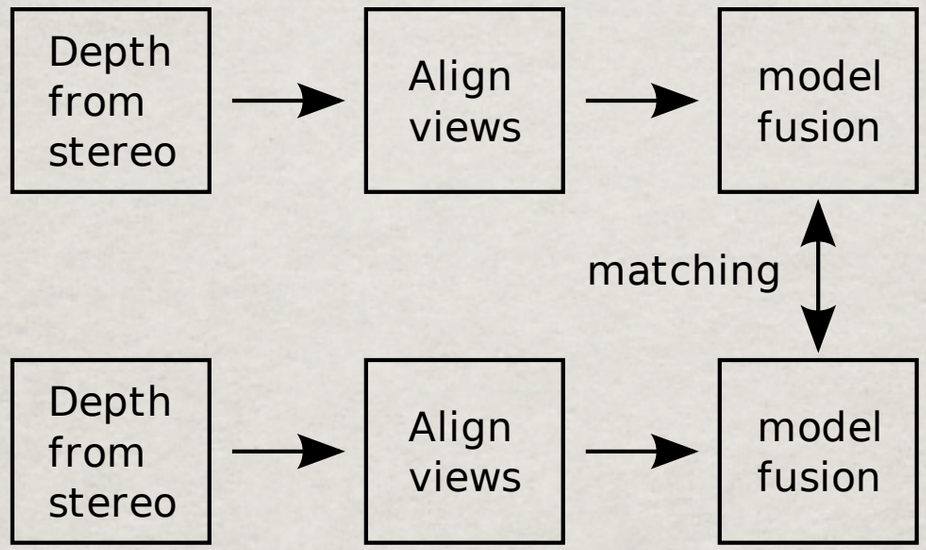
- Category membership is not a binary decision.

- Lakoff, George, "*Women, Fire, and Dangerous Things - what categories reveal about the mind*". University of Chicago Press. 1987

Taxonomy of Recognition Approaches

- **Full 3D modelling**+recognition (old school)
- Appearance based methods:
  - **Global appearance** (pattern matching)
    - Good for e.g. silhouettes and faces
- **Local feature methods**
  - can handle **occlusion** and **articulation**
  - make 3D models **after** recognition

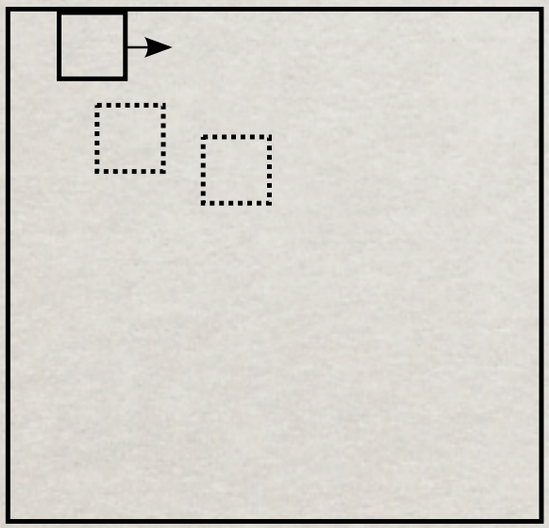
Full 3D Modelling

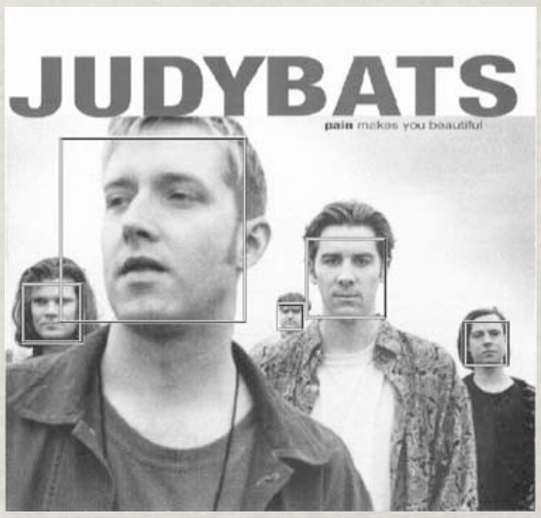


- First generate a 3D model, then match the 3D model to memory
- Bad as it adds new error sources.
- “The world is it’s own best model”, R. Brooks “Elephants don’t play chess”, RAS 6, 1990

Global Appearance

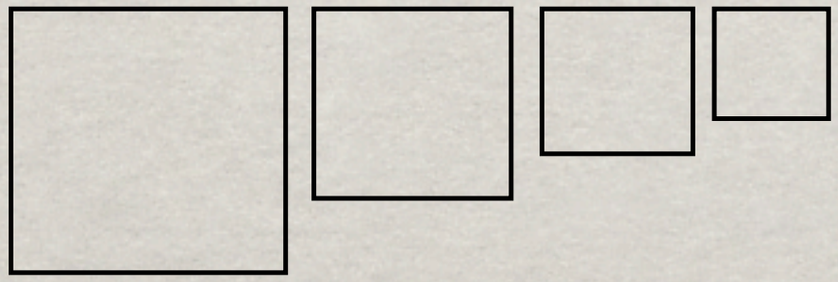
- Run a fast pattern recognition algorithm as a sliding window detector.





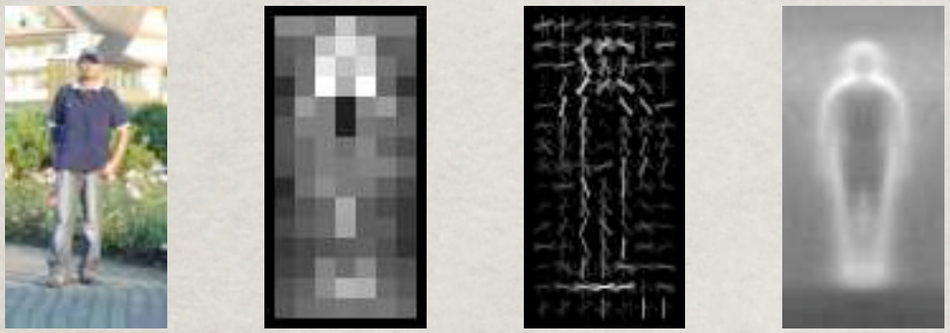
Viola&Jones IJCV04

Scale pyramid for scale invariance



Global Appearance

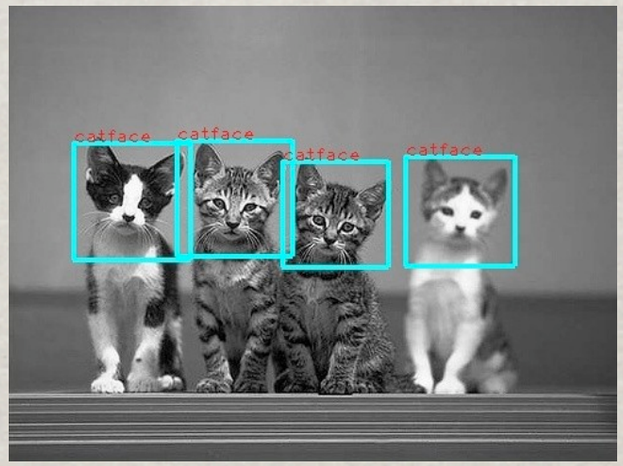
- Pedestrian detection



Dalal&Triggs, CVPR'05

- Cascaded face detection

Viola&Jones IJCV'04



Ivan Laptev

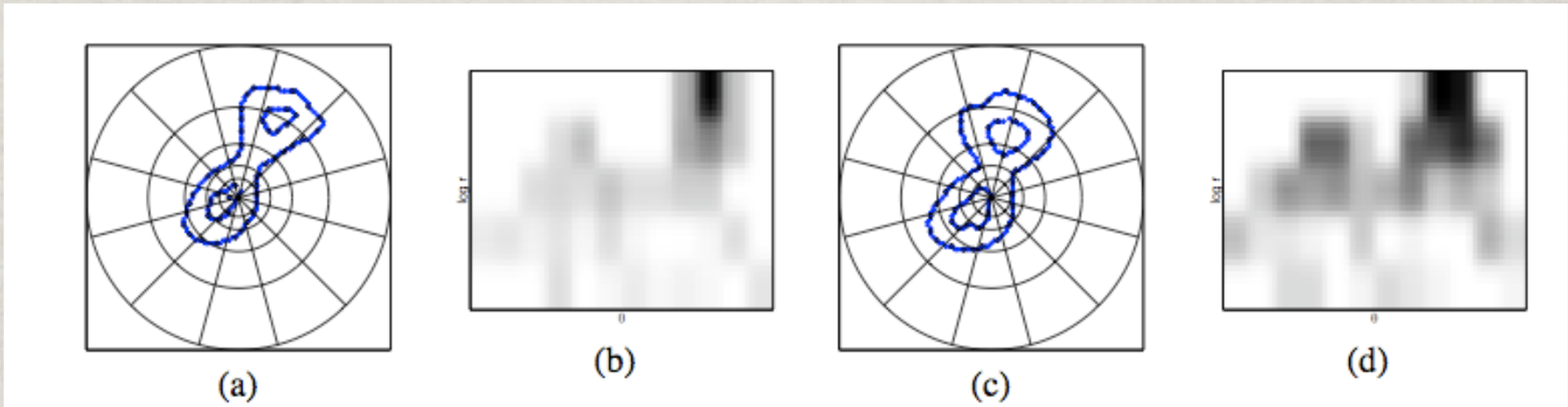
Global Appearance

- Application: Cascaded face detection for autofocus

www.adorama.com review of Fujifilm finepics F40fd

Global Appearance

- Global Elastic Contour models



e.g. Zhang & Malik, CVPR2003

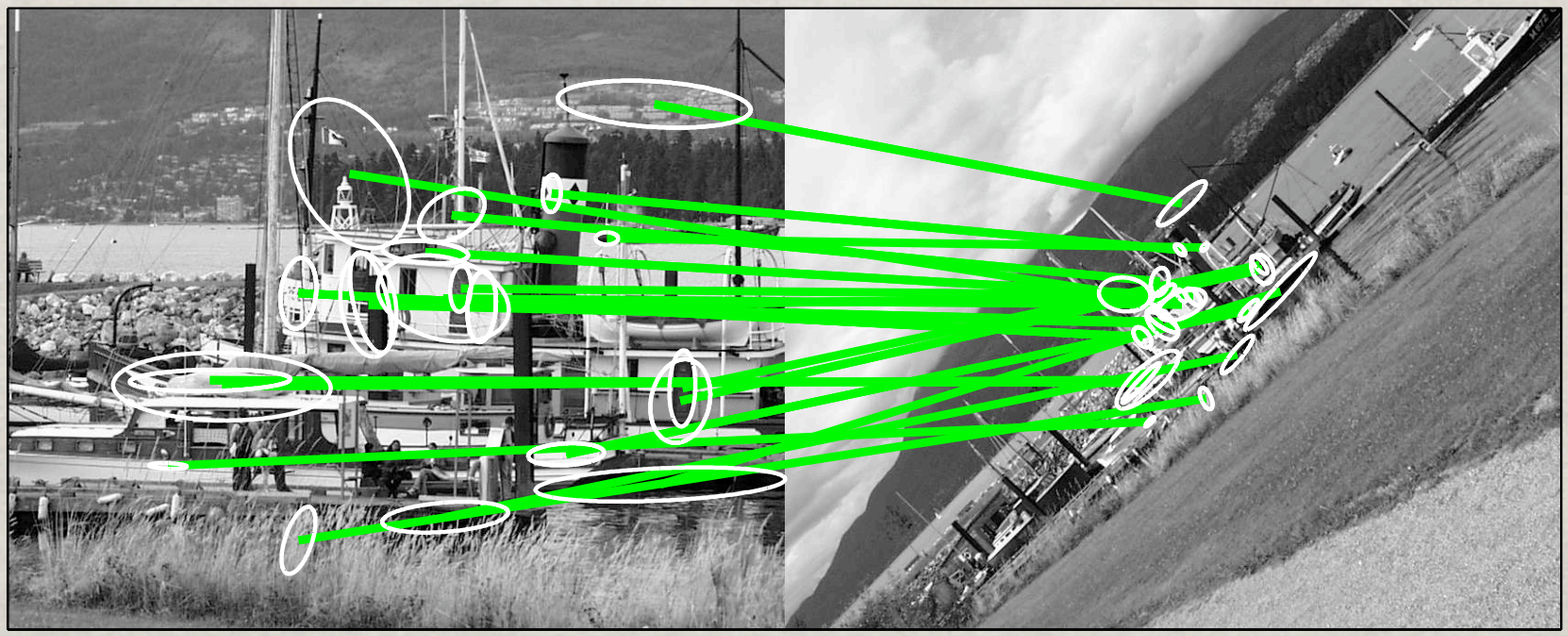
- Useful for e.g., human and animal silhouettes and hand drawn symbols.

Global Appearance

- 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.

Local Appearance models

- 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 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
- The focus of this course.

Local Appearance models

- Lecture 2: Image formation and invariances
- Lecture 3: Detection of canonical frames
- Lecture 4: Descriptor construction
- Lecture 5: Metrics for comparison
- Lecture 6: Tree search and hashing
- Lecture 7: Performance evaluation

Course Format

- **Two options**
  - If you are not a doctoral student you may skip the project, and just follow the lectures. You should still read the papers of course.
  - PhD students should also do the project, and the final exam.

Course Format

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

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

Course Format

- **The Lectures, preparation**
  - Read the paper thoroughly
  - Make notes of related questions and issues you want to discuss
  - 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.