Understanding Visual Scenes

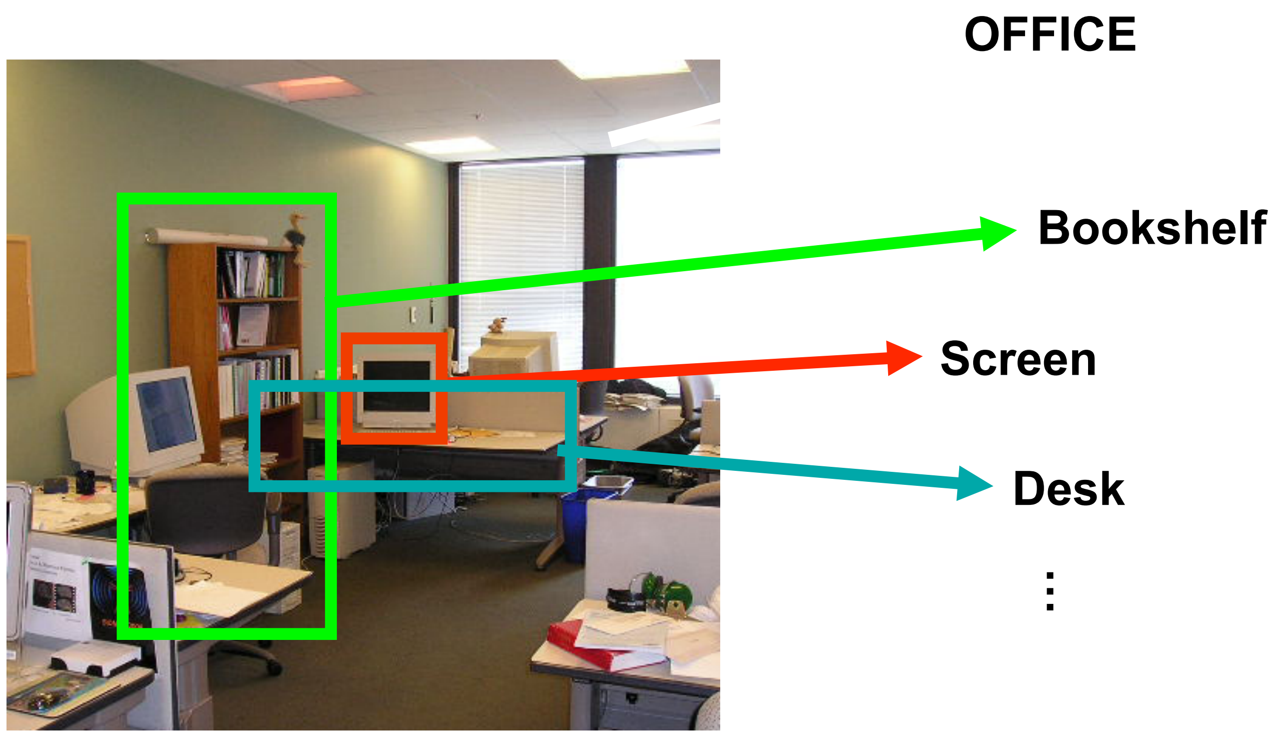
Antonio Torralba

Computer Science and Artificial Intelligence Laboratory (CSAIL)
Department of Electrical Engineering and Computer Science

Testimonials: “since I attended this class, I can recognize all the objects that I see”

A computer vision goal

Recognize many different objects under many viewing conditions in unconstrained settings.



Object recognition in 60+ minutes

Why is object recognition a hard task?

Challenges 1: view point variation

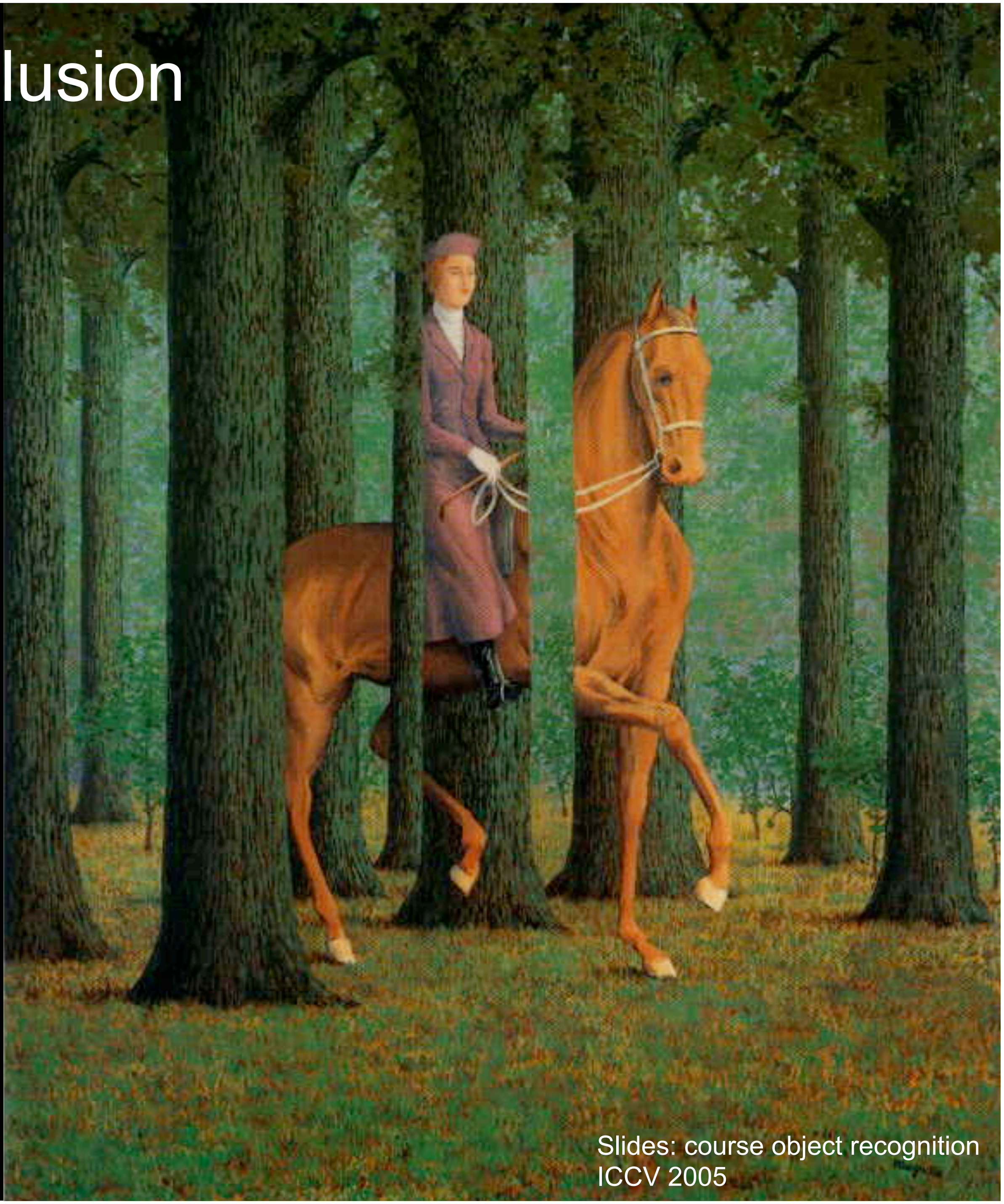
Michelangelo 1475-1564

Slides: course object recognition
ICCV 2005

Challenges 2: illumination

slide credit: S. Ullman

Challenges 3: occlusion



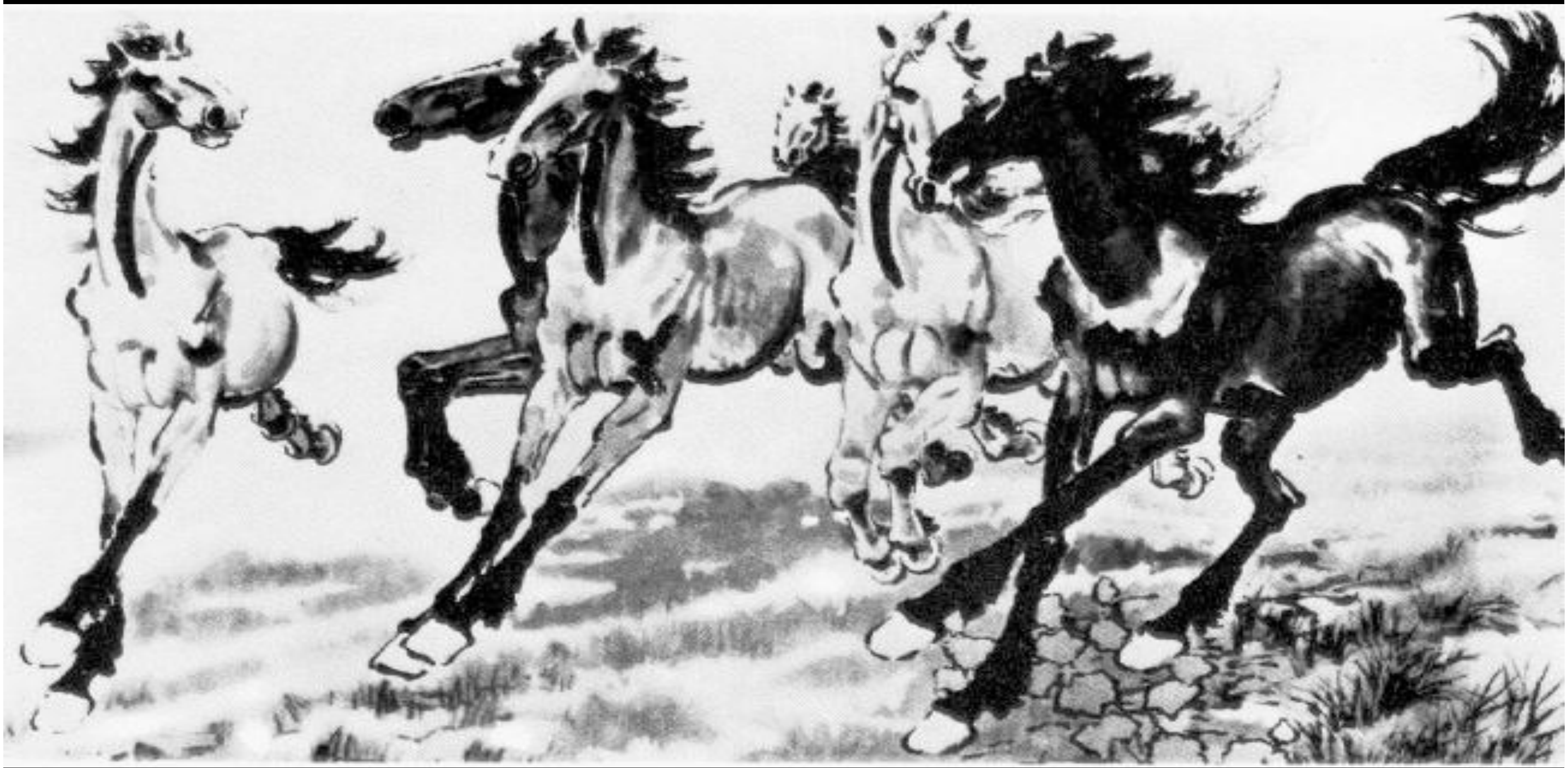
Magritte, 1957

Slides: course object recognition
ICCV 2005

Challenges 4: scale

Slides: course object recognition
ICCV 2005

Challenges 5: deformation



Slides: course object recognition
ICCV 2005

Xu, Beihong 1943

Challenges 6: intra-class variation






Slides: course object recognition
ICCV 2005

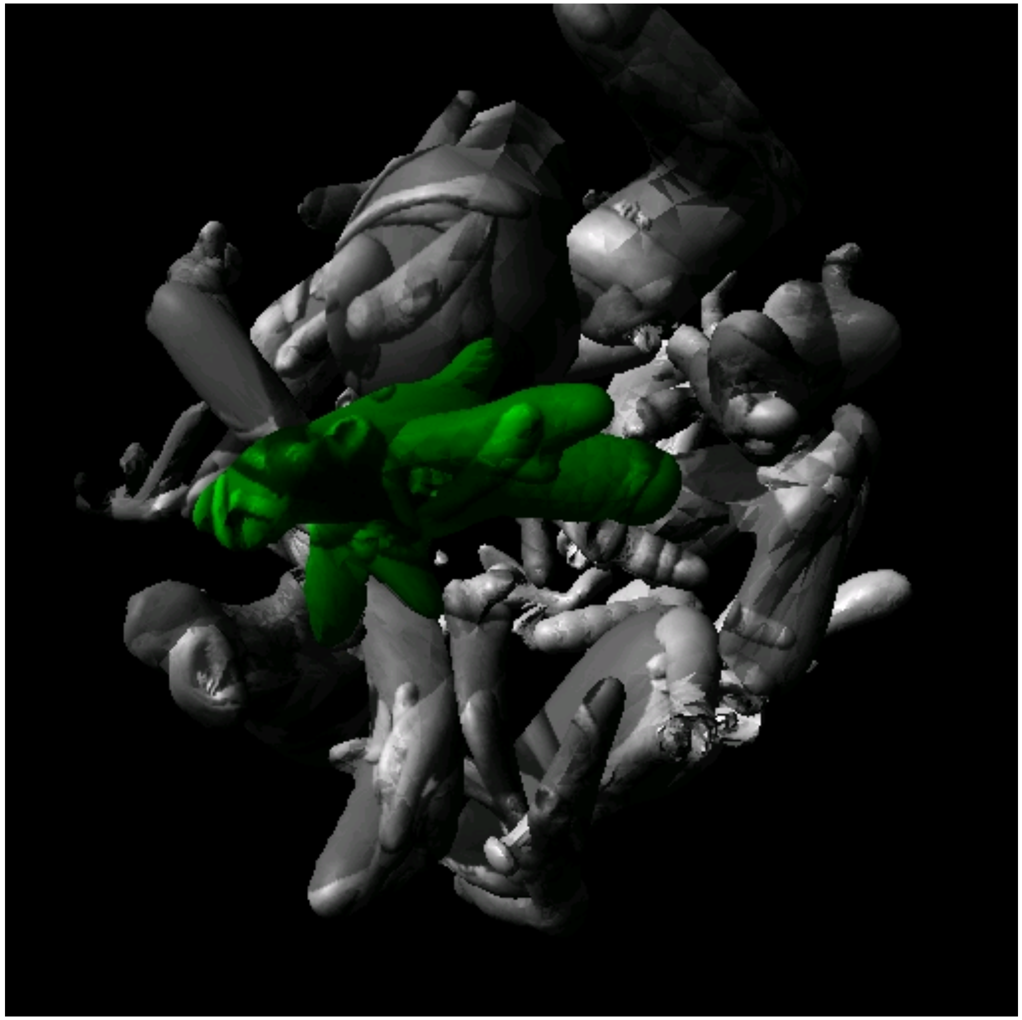
Challenges 7: background clutter

Brady, M. J., & Kersten, D. (2003). Bootstrapped learning of novel objects. J Vis, 3(6), 413-422

your visual system is amazing

your visual system is amazing?

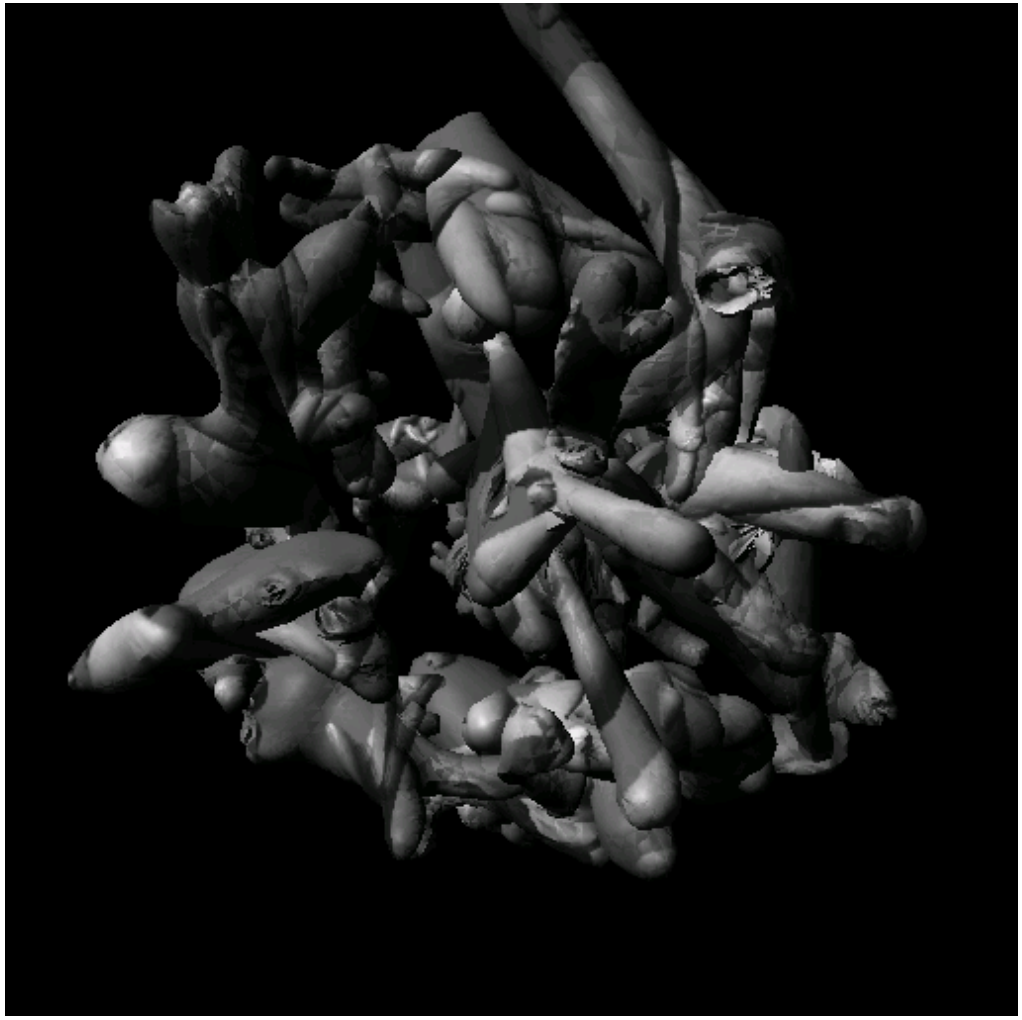
Discover the camouflaged object

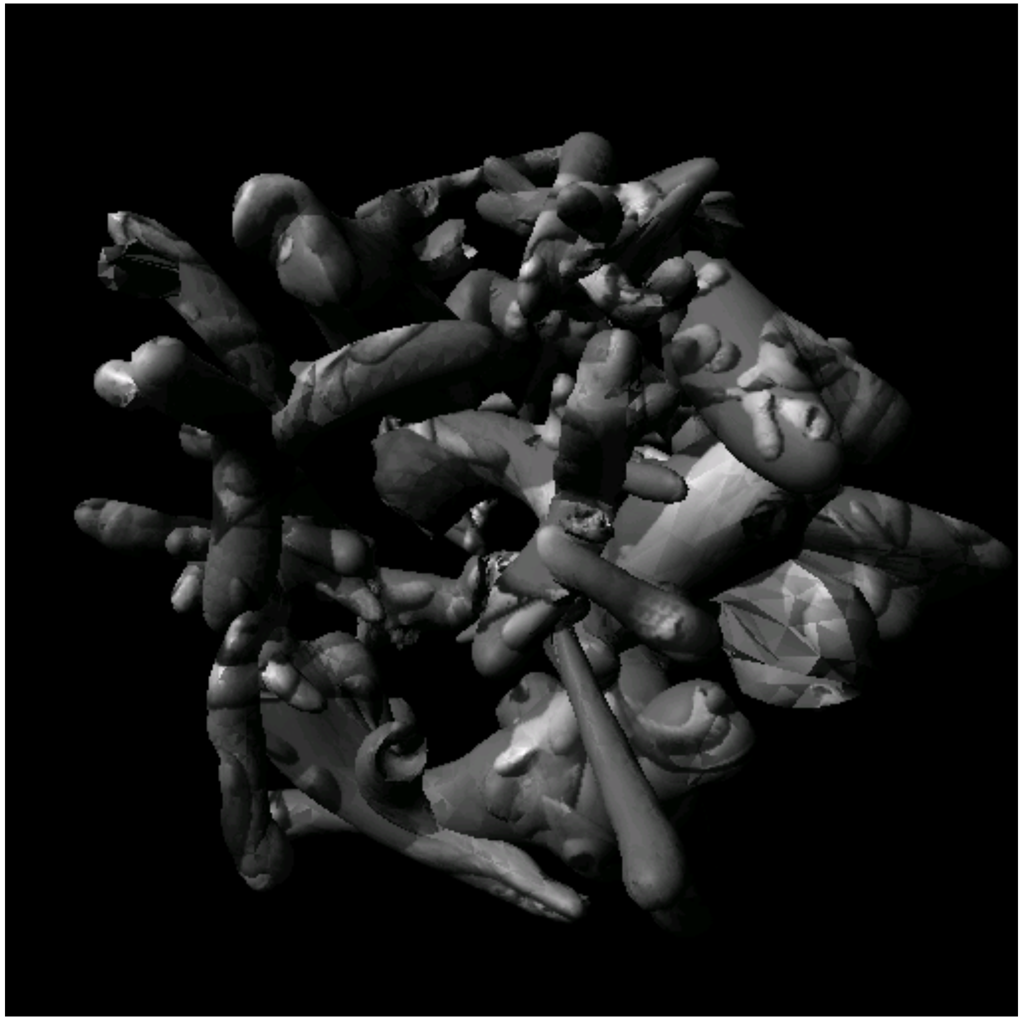


Brady, M. J., & Kersten, D. (2003). Bootstrapped learning of novel objects. J Vis, 3(6), 413-422

Discover the camouflaged object

Brady, M. J., & Kersten, D. (2003). Bootstrapped learning of novel objects. J Vis, 3(6), 413-422





Any guesses?

Why do we care about recognition?

Perception of function: We can perceive the 3D shape, texture, material properties, without knowing about objects. But, the concept of category encapsulates also information about what can we do with those objects.

"We therefore include the perception of function as a proper –indeed, crucial- subject for vision science", *from Vision Science, chapter 9, Palmer*.

The perception of function

- Direct perception (affordances): Gibson

Flat surface
Horizontal
Knee-high
…

→ Sittable upon

- Mediated perception (Categorization)

Flat surface
Horizontal
Knee-high
…

→ Chair → Sittable upon

Direct perception

Some aspects of an object function can be perceived directly

- Functional form: Some forms clearly indicate to a function (“sittable-upon”, container, cutting device, …)

Direct perception

Some aspects of an object function can be perceived directly

- Observer relativity: Function is observer dependent

Limitations of Direct Perception

Objects of similar structure might have very different functions

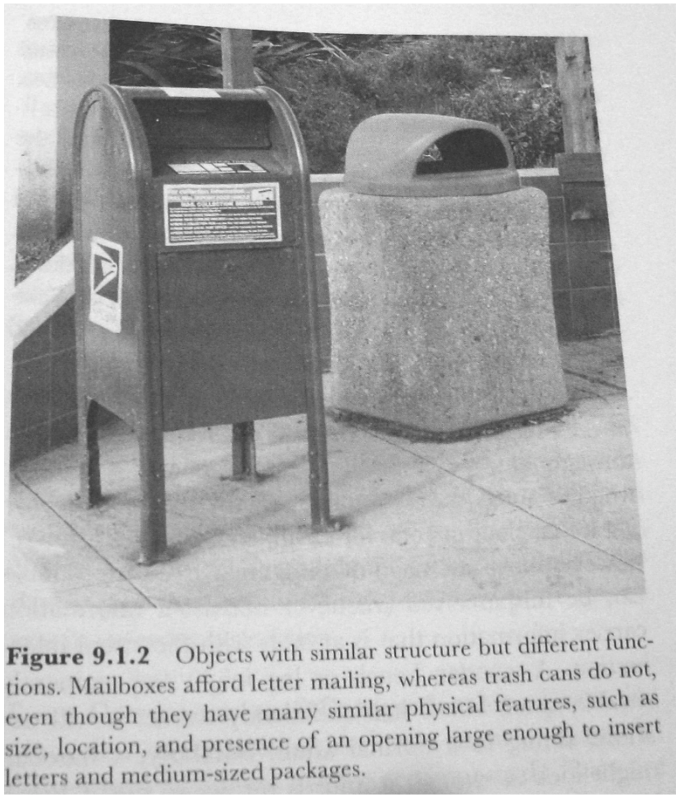


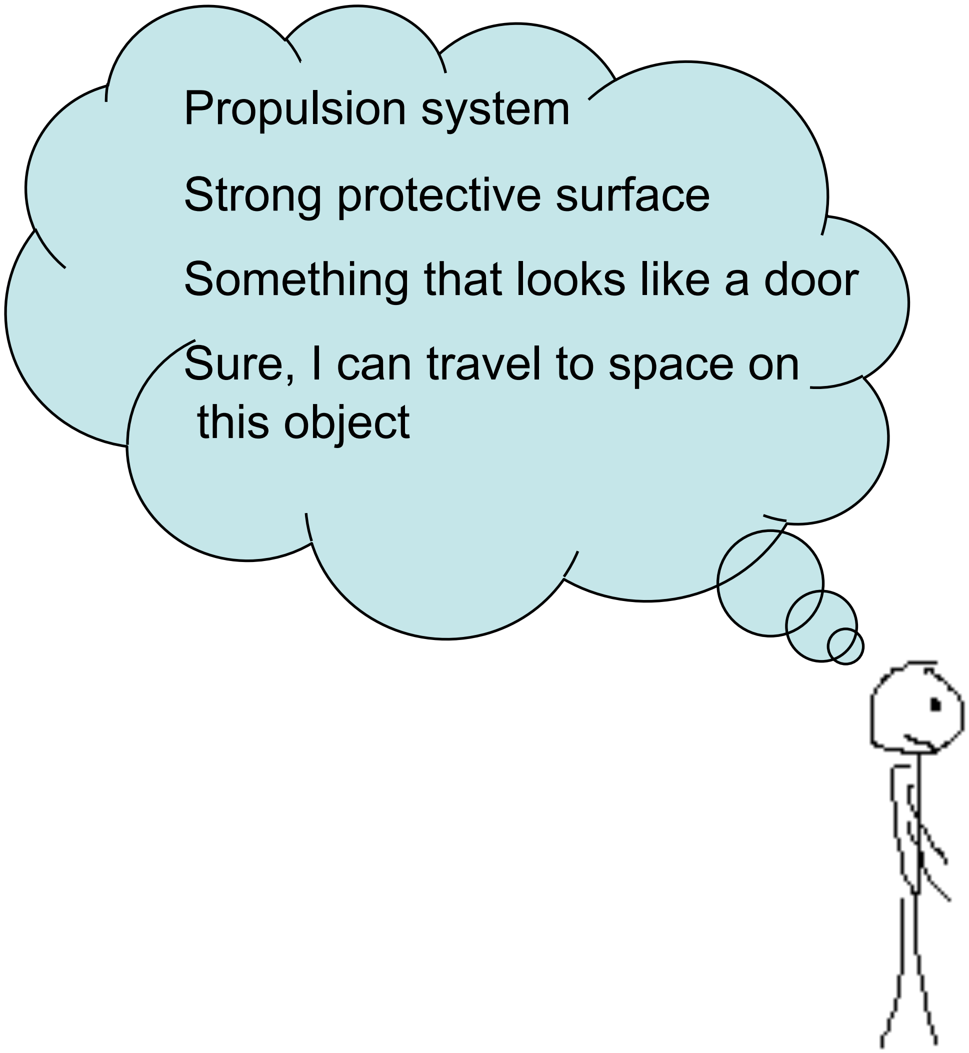
Figure 9.1.2 Objects with similar structure but different functions. Mailboxes afford letter mailing, whereas trash cans do not, even though they have many similar physical features, such as size, location, and presence of an opening large enough to insert letters and medium-sized packages.

Not all functions seem to be available from direct visual information only.

The functions are the same at some level of description: we can put things inside in both and somebody will come later to empty them. However, we are not expected to put inside the same kinds of things…

Limitations of Direct Perception

Visual appearance might be a very weak cue to function

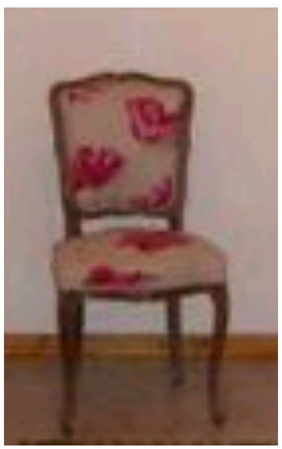


How do we achieve Mediated perception?

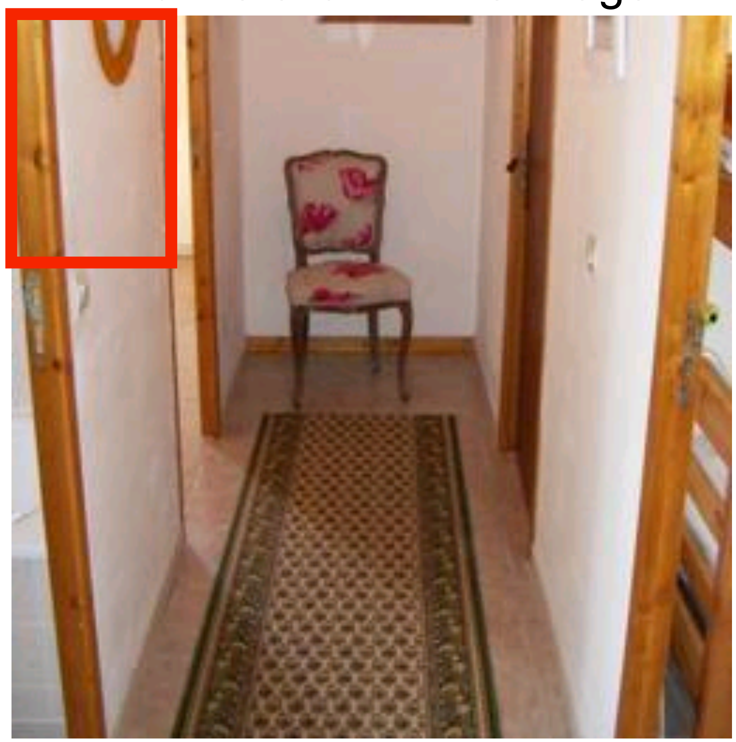
Well… this requires object recognition (for more details, see entire course)

Object recognition
Is it really so hard?

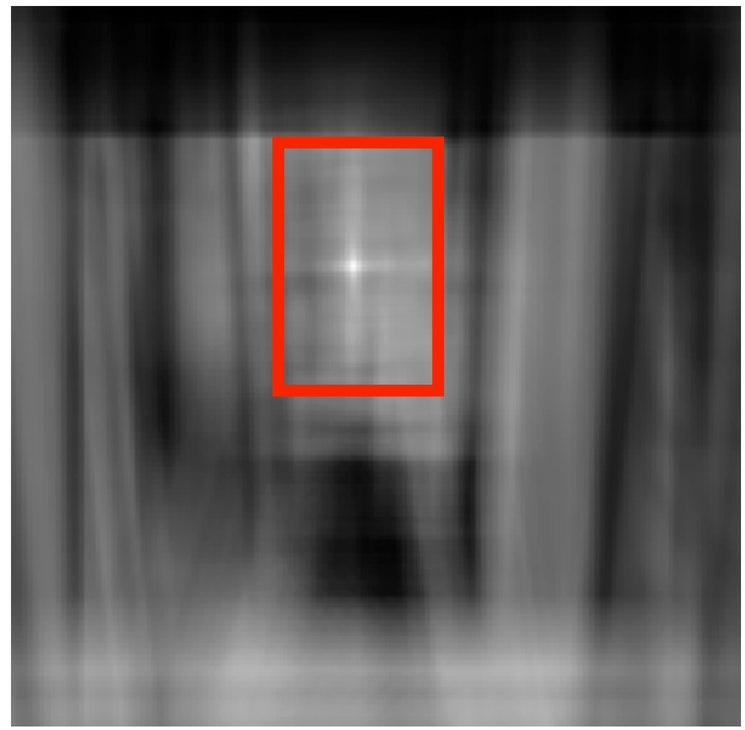
This is a chair

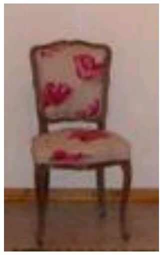


Find the chair in this image



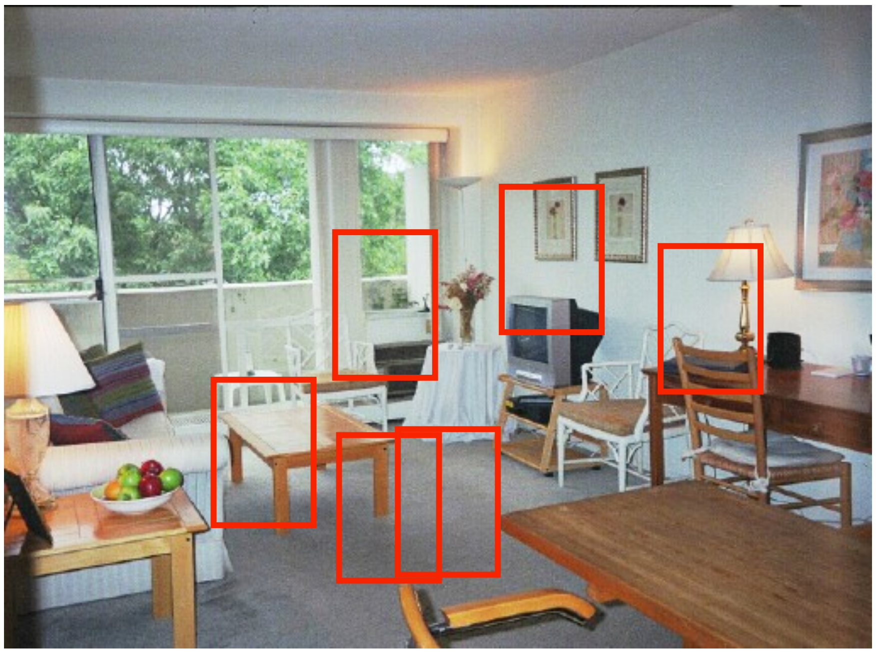
Output of normalized correlation

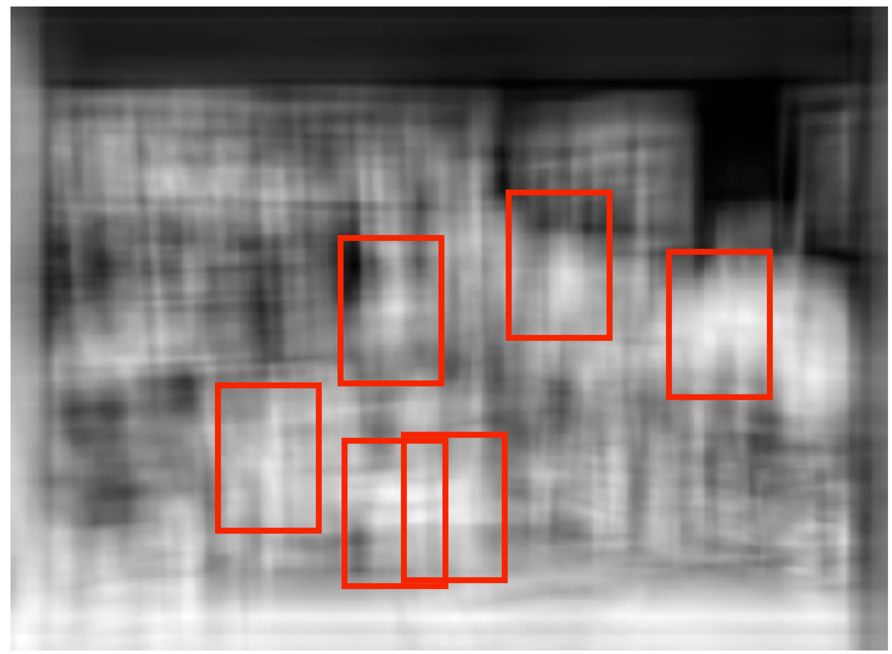




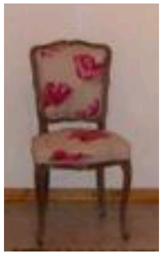
Object recognition
Is it really so hard?

Find the chair in this image





Pretty much garbage
Simple template matching is not going to make it

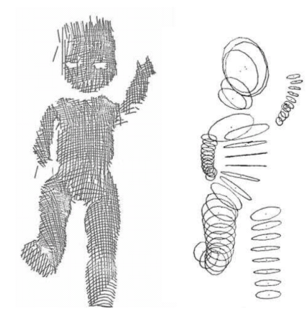


Object recognition
Is it really so hard?

Find the chair in this image

A “popular method is that of template matching, by point to point correlation of a model pattern with the image pattern. These techniques are inadequate for three-dimensional scene analysis for many reasons, such as occlusion, changes in viewing angle, and articulation of parts.” Nivatia & Binford, 1977.

A short story of object recognition



So, let's make the problem simpler: Block world

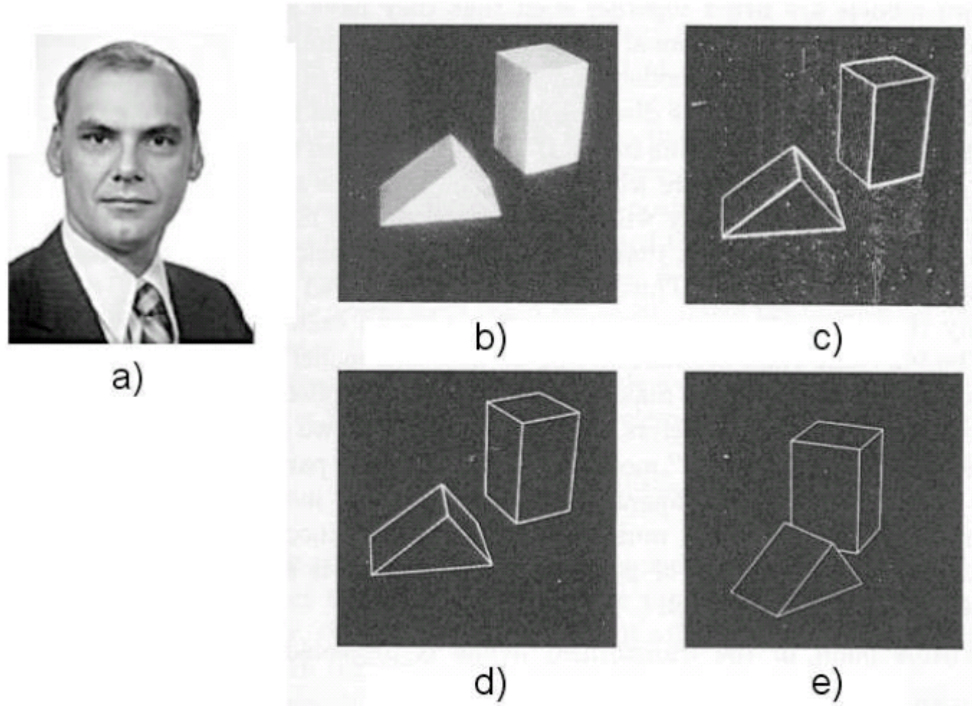
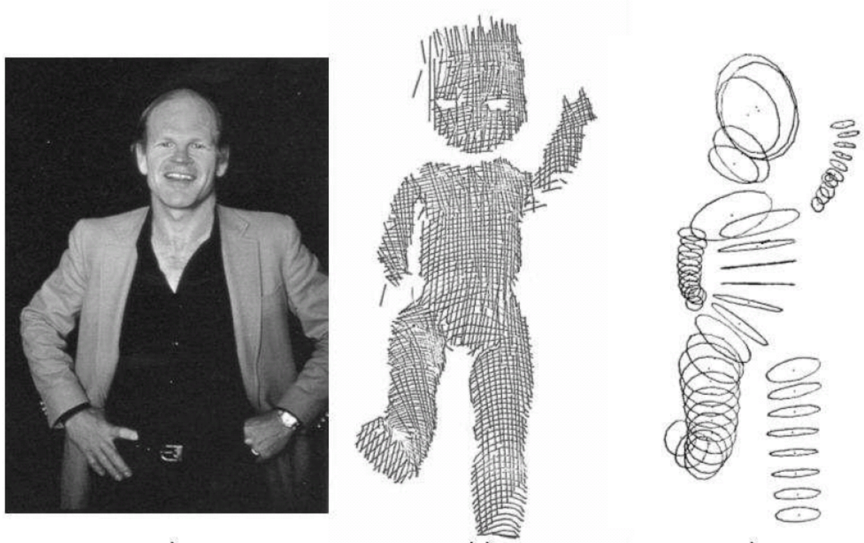


Fig. 1. A system for recognizing 3-d polyhedral scenes. a) L.G. Roberts. b)A blocks world scene. c)Detected edges using a 2x2 gradient operator. d) A 3-d polyhedral description of the scene, formed automatically from the single image. e) The 3-d scene displayed with a viewpoint different from the original image to demonstrate its accuracy and completeness. (b) - e) are taken from [64] with permission MIT Press.)

Nice framework to develop fancy math, but too far from reality…

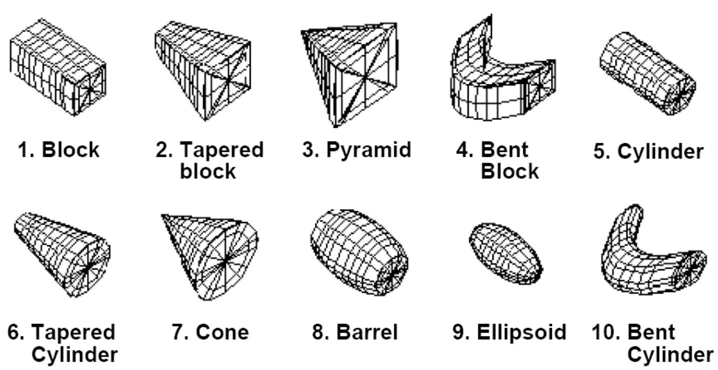
Object Recognition in the Geometric Era:
a Retrospective. Joseph L. Mundy. 2006

Binford and generalized cylinders



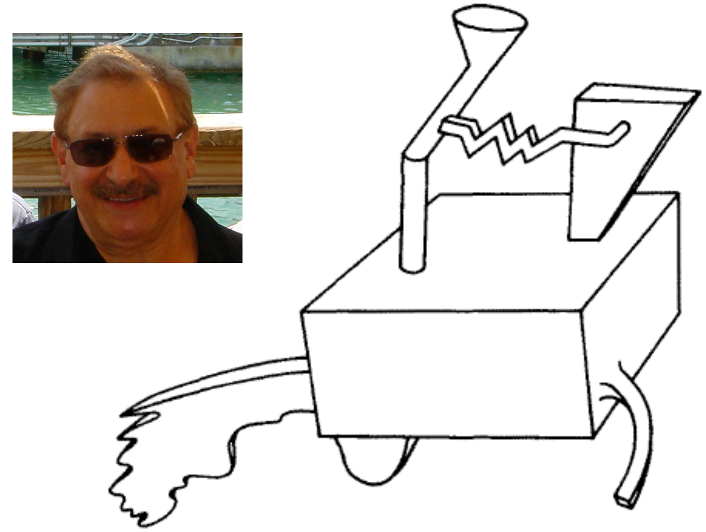
a) b) c)

Fig. 3. The representation of objects by assemblies of generalized cylinders. a) Thomas Binford. b) A range image of a doll. c) The resulting set of generalized cylinders. (b) and c) are taken from Agin [1] with permission.)



Introduced in computer vision by A. Pentland, 1986.

Recognition by components



Irving Biederman
Recognition-by-Components: A Theory of Human Image Understanding.
Psychological Review, 1987.

Object Recognition in the Geometric Era: a Retrospective. Joseph L. Mundy. 2006

Parts and Structure approaches

With a different perspective, these models focused more on the geometry than on defining the constituent elements:

- Fischler & Elschlager 1973
- Yuille ‘91
- Brunelli & Poggio ‘93
- Lades, v.d. Malsburg et al. ‘93
- Cootes, Lanitis, Taylor et al. ‘95
- Amit & Geman ‘95, ‘99
- Perona et al. ‘95, ‘96, ’98, ’00, ’03, ‘04
- Felzenszwalb & Huttenlocher ’00, ’04
- Crandall & Huttenlocher ’05, ’06
- Leibe & Schiele ’03, ’04
- Many papers since 2000

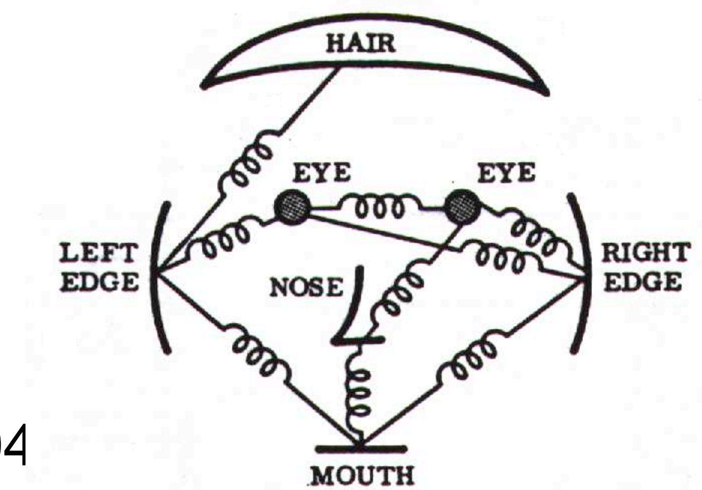
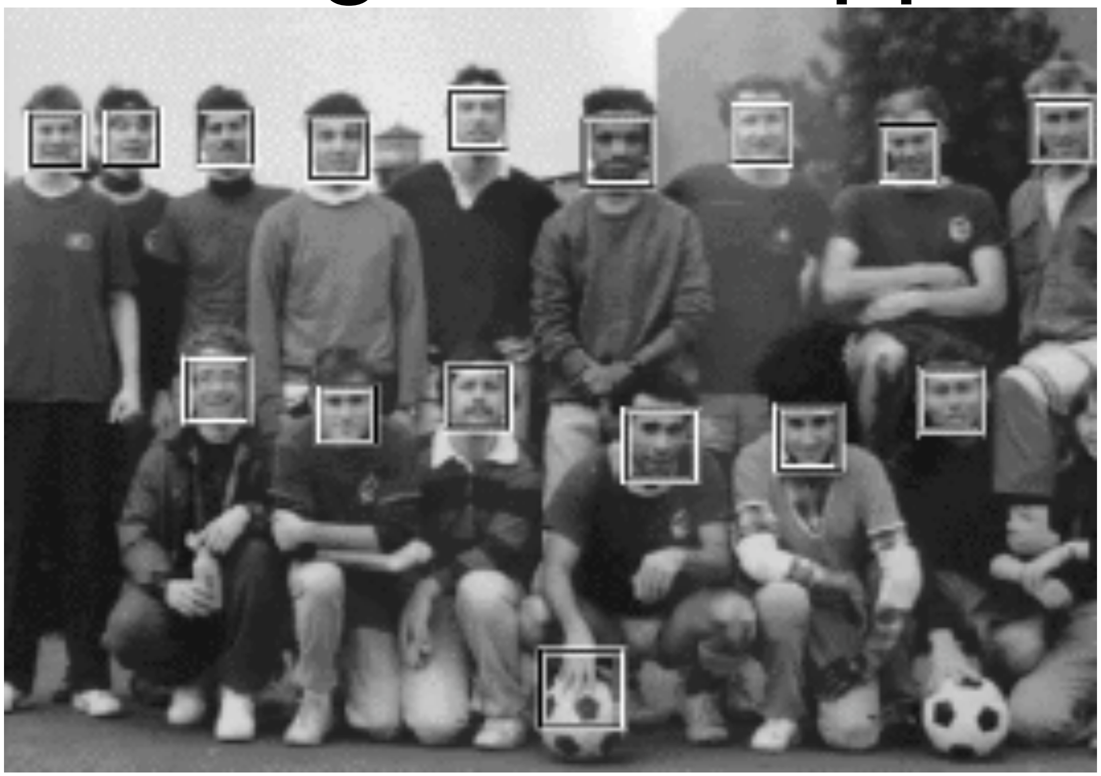


Figure from [Fischler & Elschlager 73]

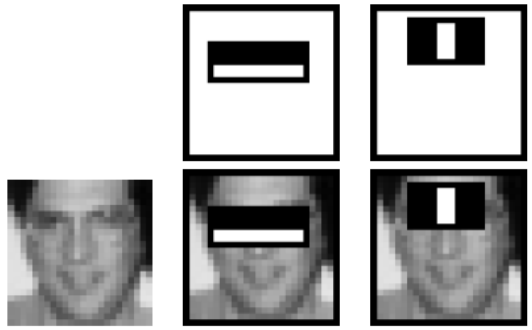
But, despite promising initial results…things did not work out so well (lack of data, processing power, lack of reliable methods for low-level and mid-level vision)

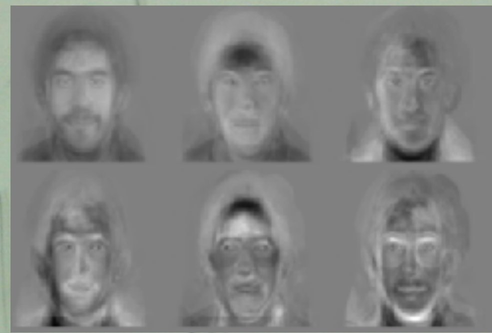
Instead, a different way of thinking about object detection started making some progress: learning based approaches and classifiers, which ignored low and mid-level vision.

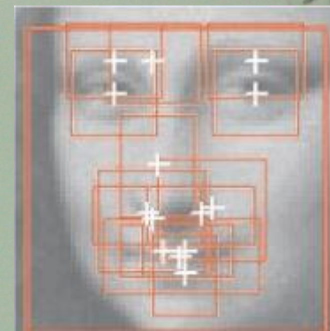
Face detection and the success of learning based approaches

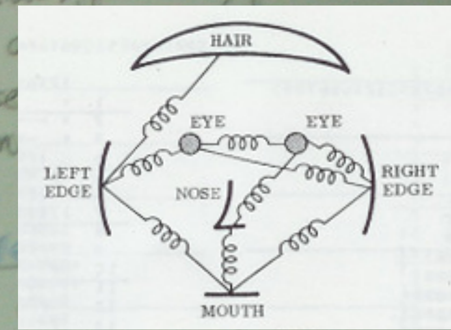


- The representation and matching of pictorial structures Fischler, Elschlager (1973).
- Face recognition using eigenfaces M. Turk and A. Pentland (1991).
- Human Face Detection in Visual Scenes - Rowley, Baluja, Kanade (1995)
- Graded Learning for Object Detection - Fleuret, Geman (1999)
- Robust Real-time Object Detection - Viola, Jones (2001)
- Feature Reduction and Hierarchy of Classifiers for Fast Object Detection in Video Images - Heisele, Serre, Mukherjee, Poggio (2001)
- ….



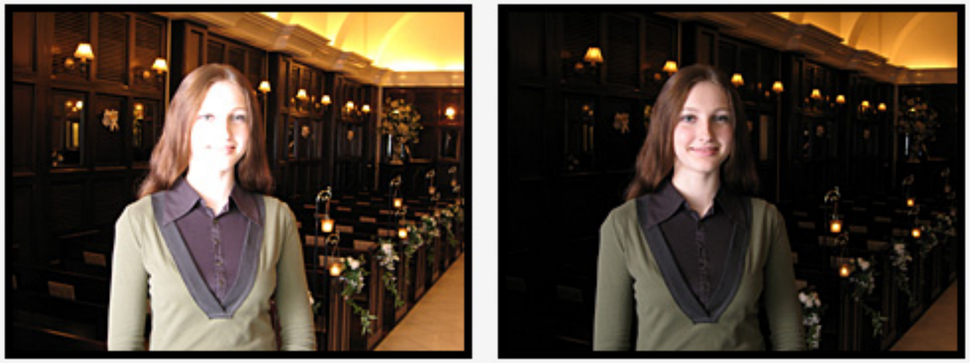






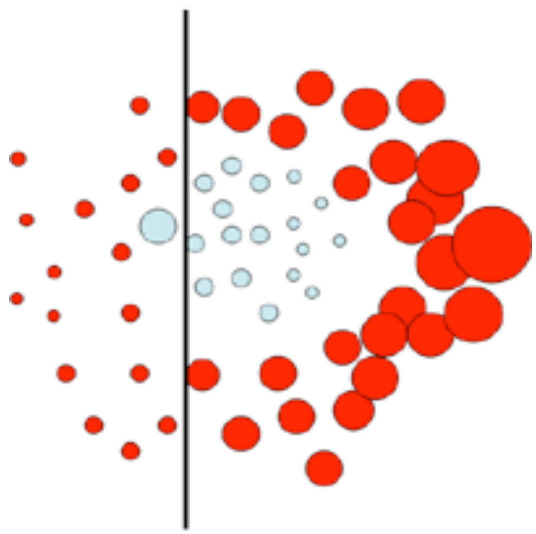
- The representation and matching of pictorial structures Fischler, Elschlager (1973)
- Face recognition using eigenfaces M. Turk and A. Pentland (1991).
- Human Face Detection in Visual Scenes - Rowley, Baluja, Kanade (1995)
- Graded Learning for Object Detection - Fleuret, Geman (1999)
- Robust Real-time Object Detection - Viola, Jones (2001)
- Feature Reduction and Hierarchy of Classifiers for Fast Object Detection in Video Images - Heisele, Serre, Mukherjee, Poggio (2001)
- ….

Face detection



[Face priority AE] When a bright part of the face is too bright

A simple object detector



• Simple but contains some of same basic elements of many state of the art detectors.

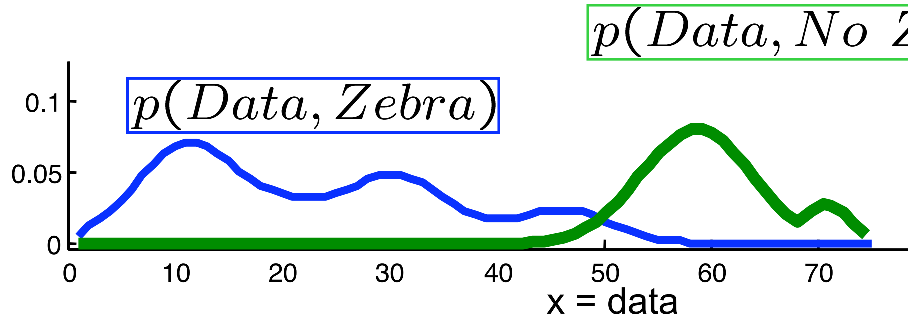
• Based on boosting which makes all the stages of the training and testing easy to understand.

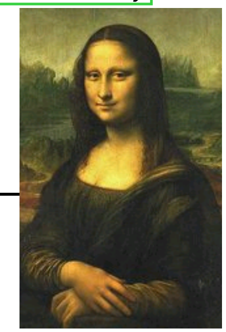
Most of the slides are from the ICCV 05 short course
http://people.csail.mit.edu/torralba/shortCourseRLOC/

Discriminative vs. generative

- Generative model

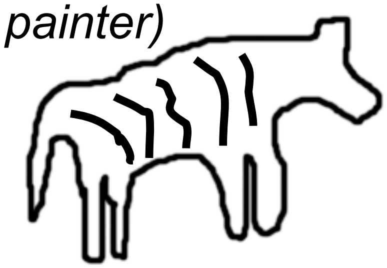
  (*The artist*)

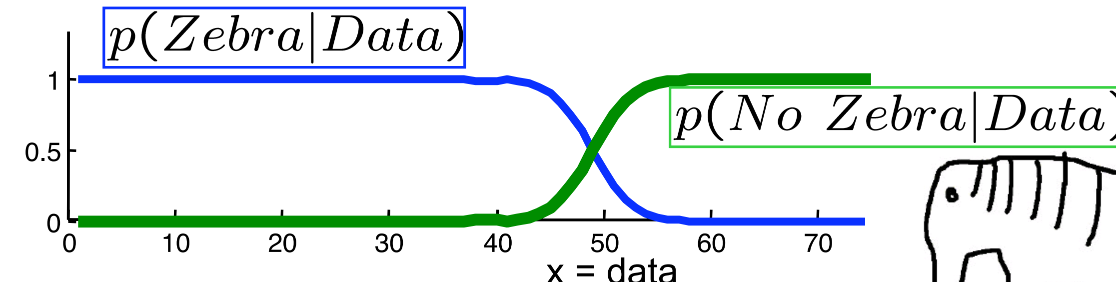




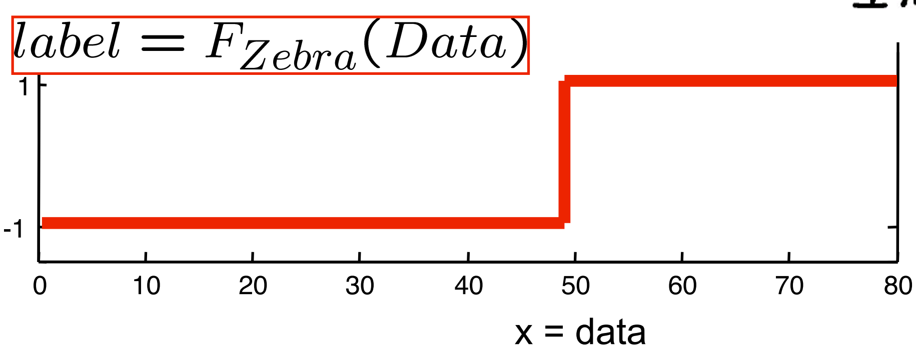
- Discriminative model

  (The lousy painter)





- Classification function

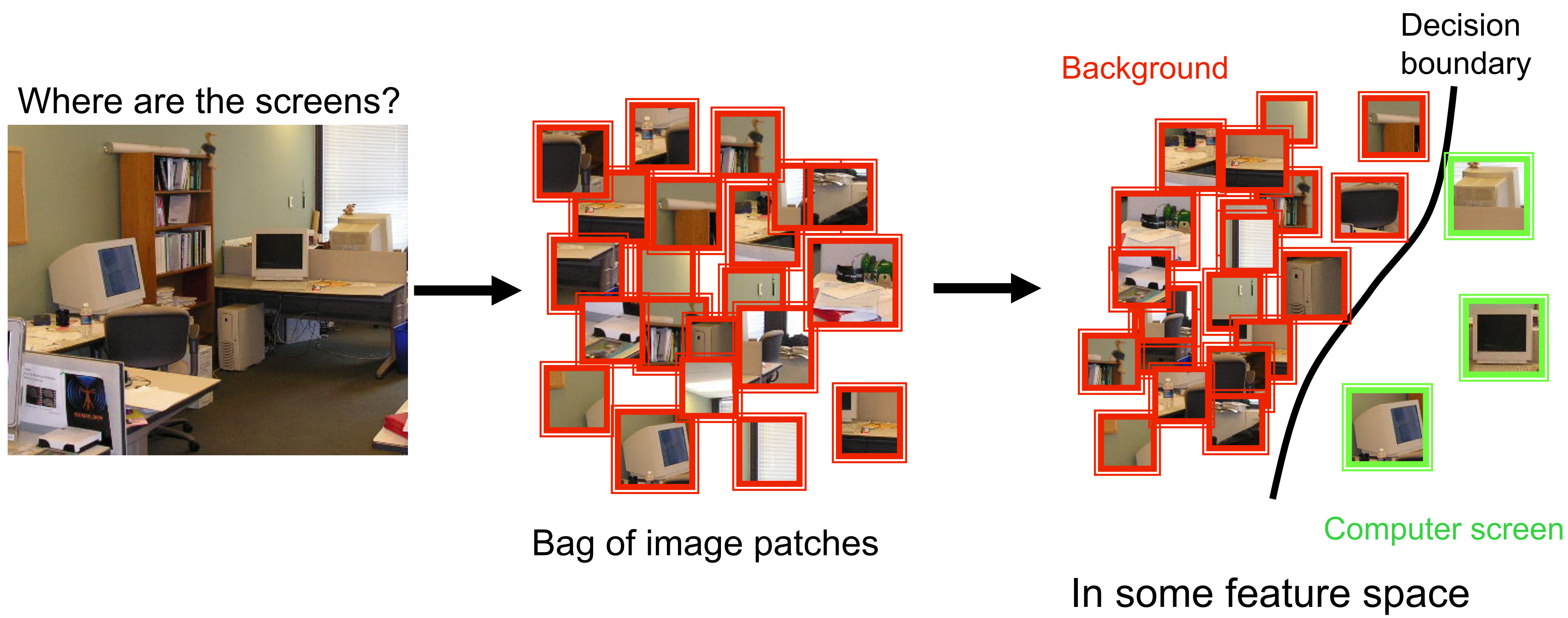


Discriminative methods

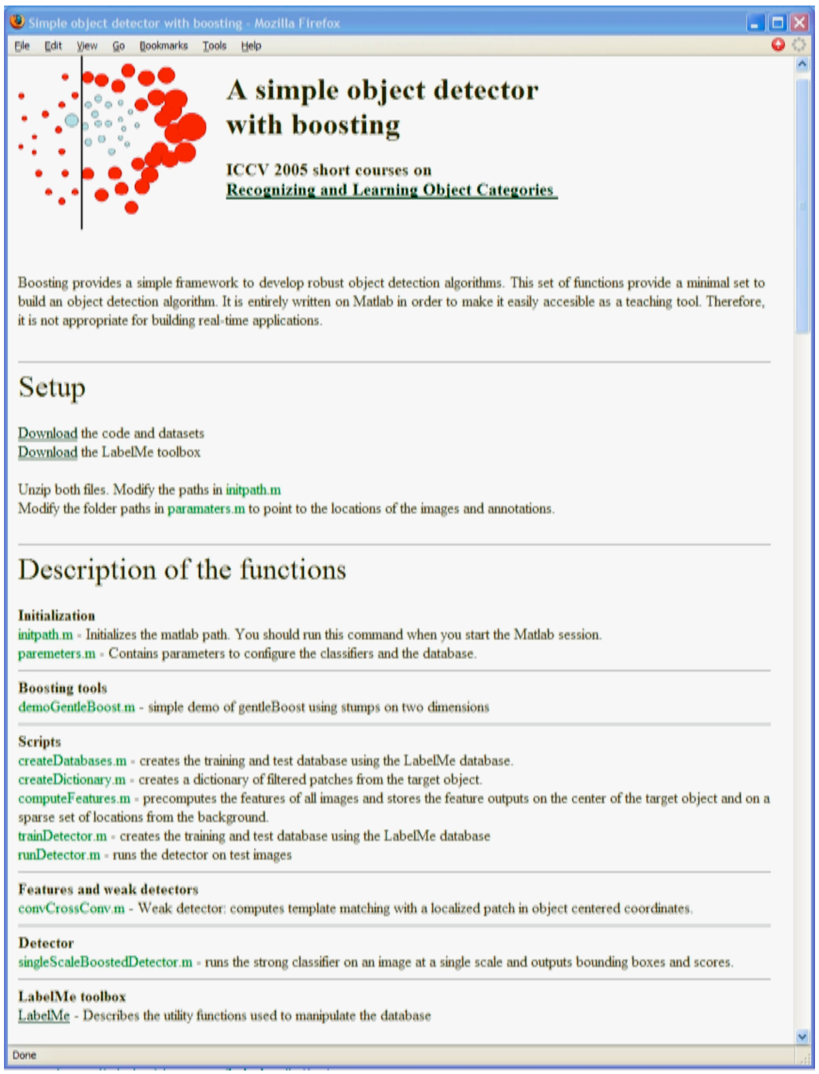
Object detection and recognition is formulated as a classification problem.

The image is partitioned into a set of overlapping windows

… and a decision is taken at each window about if it contains a target object or not.



A simple object detector with Boosting



http://people.csail.mit.edu/torralba/iccv2005/

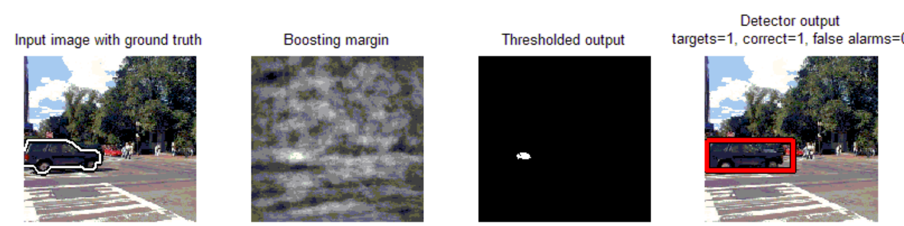
Download

- Toolbox for manipulating dataset
- Code and dataset

Matlab code

- Gentle boosting
- Object detector using a part based model

Dataset with cars and computer monitors



Why boosting?

- A simple algorithm for learning robust classifiers
  - Freund & Shapire, 1995
  - Friedman, Hastie, Tibshhirani, 1998

- Provides efficient algorithm for sparse visual feature selection
  - *Tieu & Viola, 2000*
  - *Viola & Jones, 2003*

- Easy to implement, not requires external optimization tools.

Boosting

- Defines a classifier using an additive model:

$$\underbrace{F}_{\text{Strong classifier}}(\underbrace{x}_{\text{Features vector}}) = \underbrace{\alpha_1}_{\text{Weight}} \underbrace{f_1}_{\text{Weak classifier}}(x) + \alpha_2 f_2(x) + \alpha_3 f_3(x) + \ldots$$

Boosting

- Defines a classifier using an additive model:

$$F(x) = \alpha_1 f_1(x) + \alpha_2 f_2(x) + \alpha_3 f_3(x) + \dots$$

Strong classifier

Weak classifier

Weight

Features vector

- We need to define a family of weak classifiers

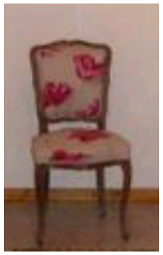
$f_k(x)$ from a family of weak classifiers

From images to features: Weak detectors

We will now define a family of visual features that can be used as weak classifiers ("weak detectors")

$$\longrightarrow h_i(I, x, y) \longrightarrow$$

Takes image as input and the output is binary response.
The output is a weak detector.



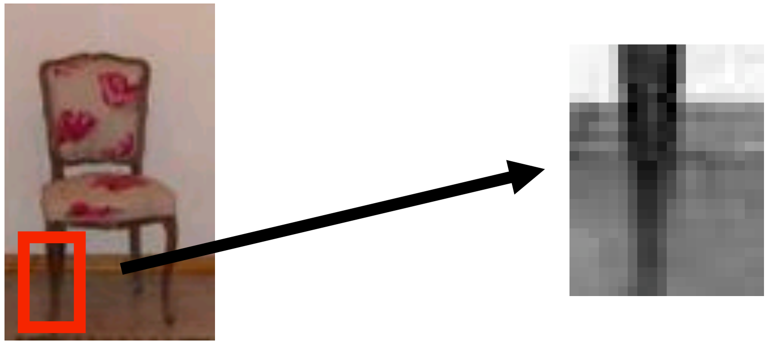
Object recognition
Is it really so hard?

Find the chair in this image

But what if we use smaller patches? Just a part of the chair?

Parts

But what if we use smaller patches? Just a part of the chair?

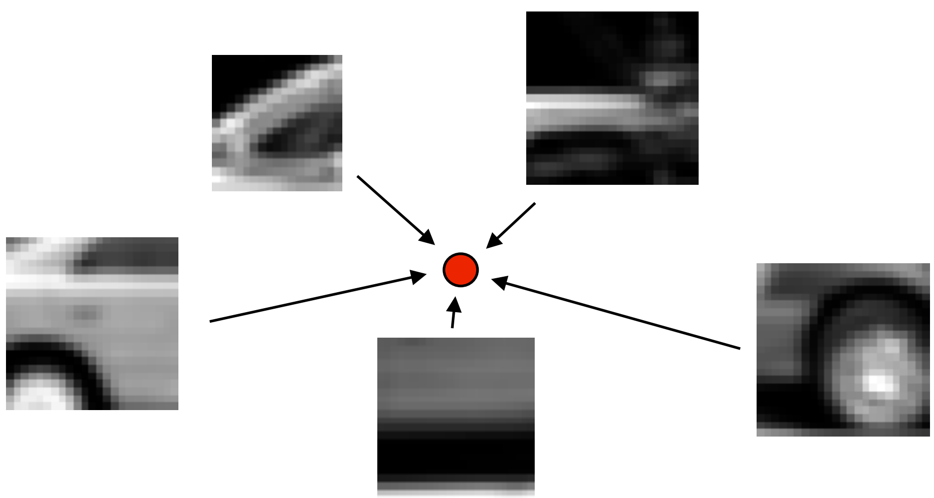


Find a chair in this image

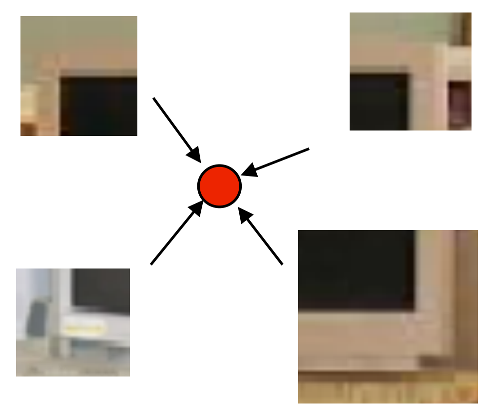
Seems to fire on legs… not so bad

Weak detectors

Part based: similar to part-based generative models. We create weak detectors by using parts and voting for the object center location



Car model



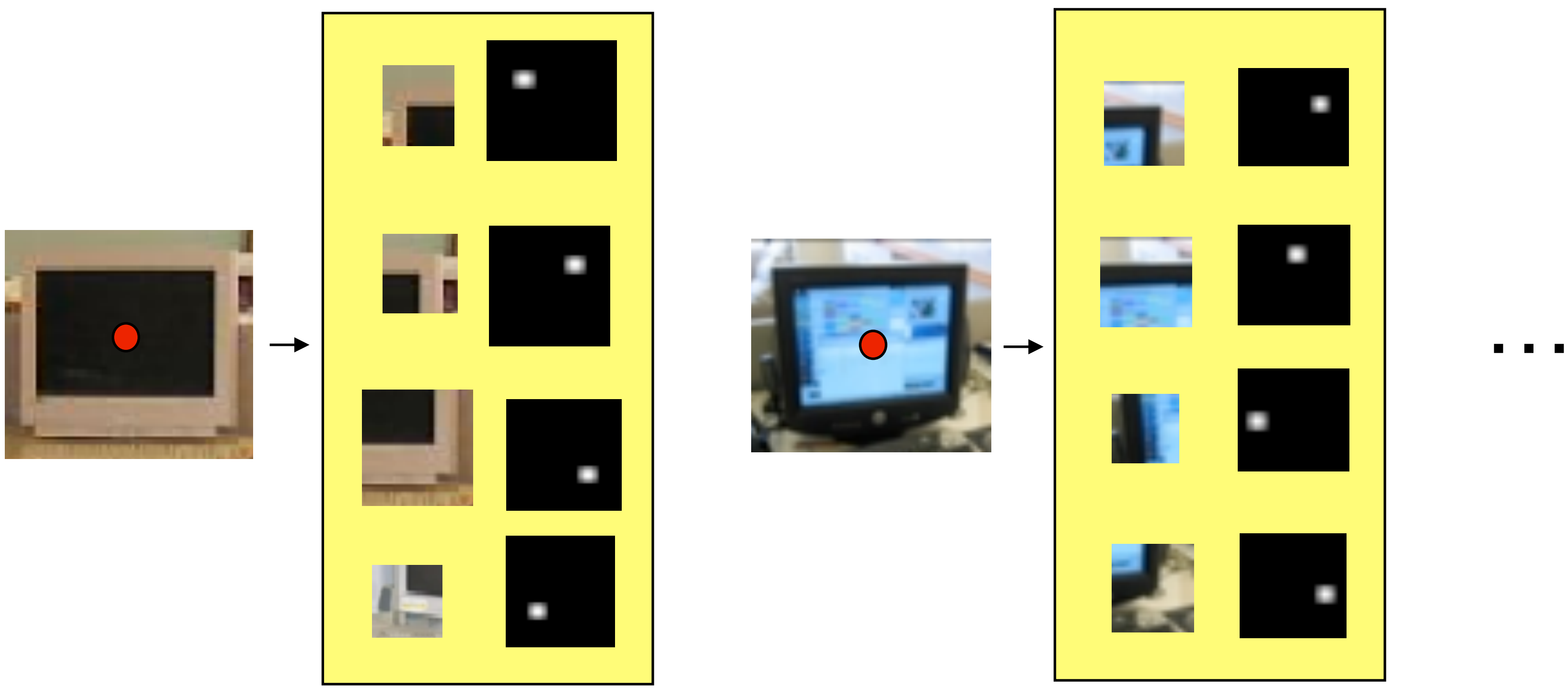
Screen model

These features are used for the detector on the course web site.

Weak detectors

First we collect a set of part templates from a set of training objects.

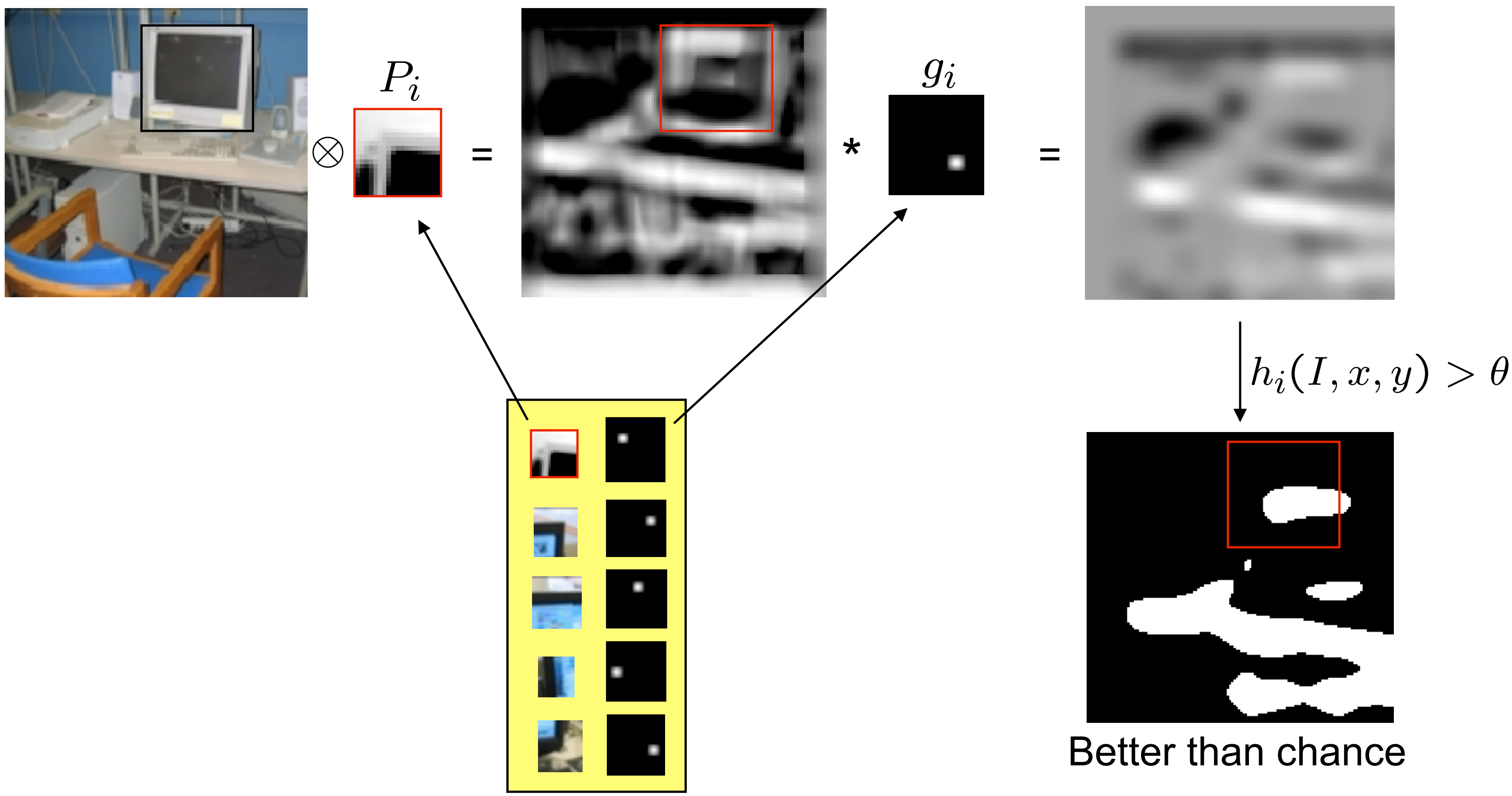
Vidal-Naquet, Ullman (2003)



Weak detectors

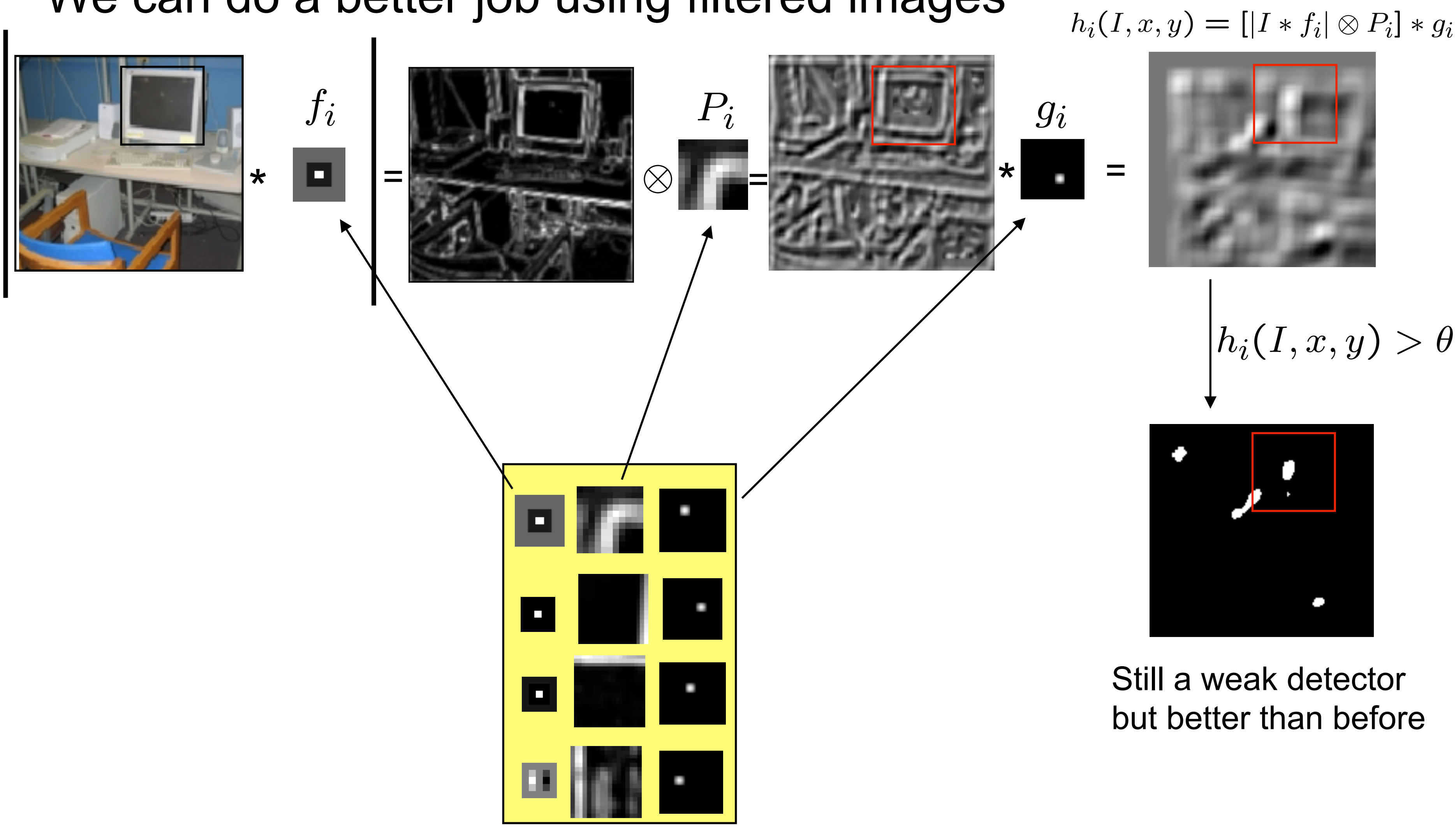
We now define a family of "weak detectors" as:

$$h_i(I, x, y) = [I \otimes P_i] * g_i$$



Weak detectors

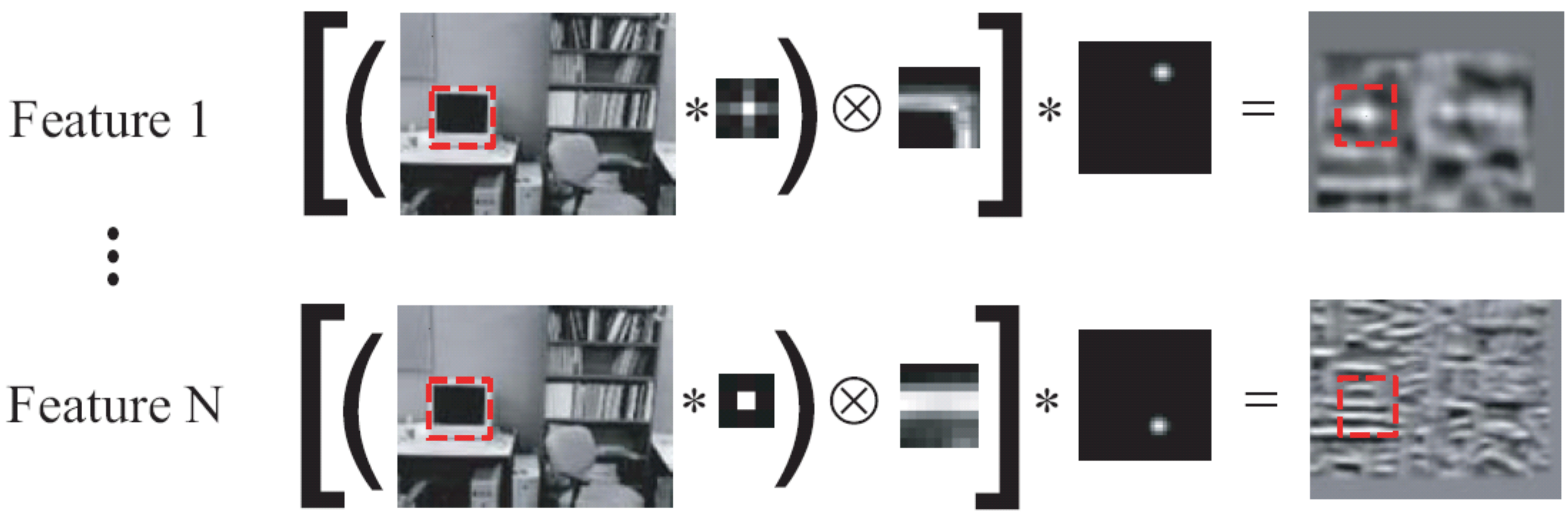
We can do a better job using filtered images



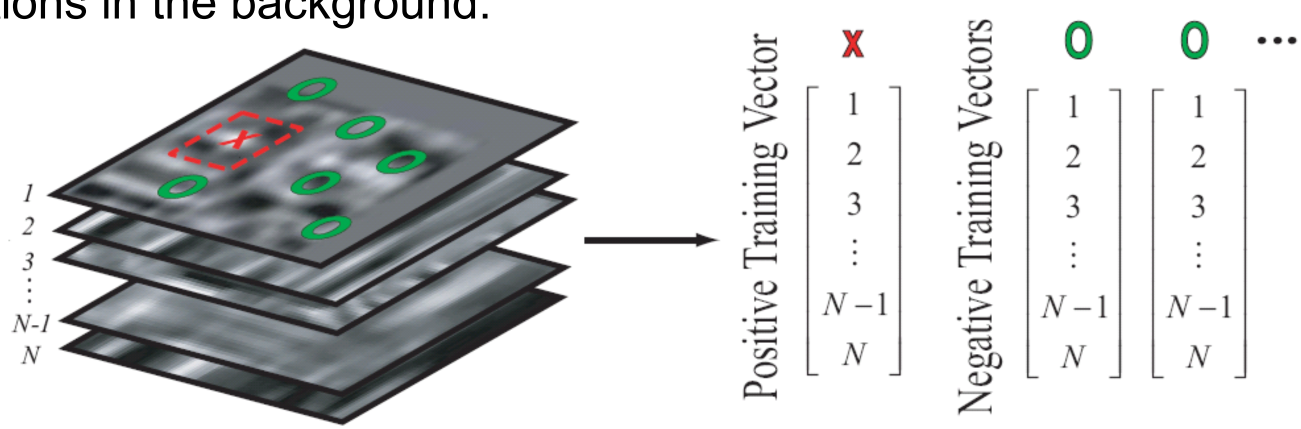
Still a weak detector
but better than before

Training

First we evaluate all the N features on all the training images.

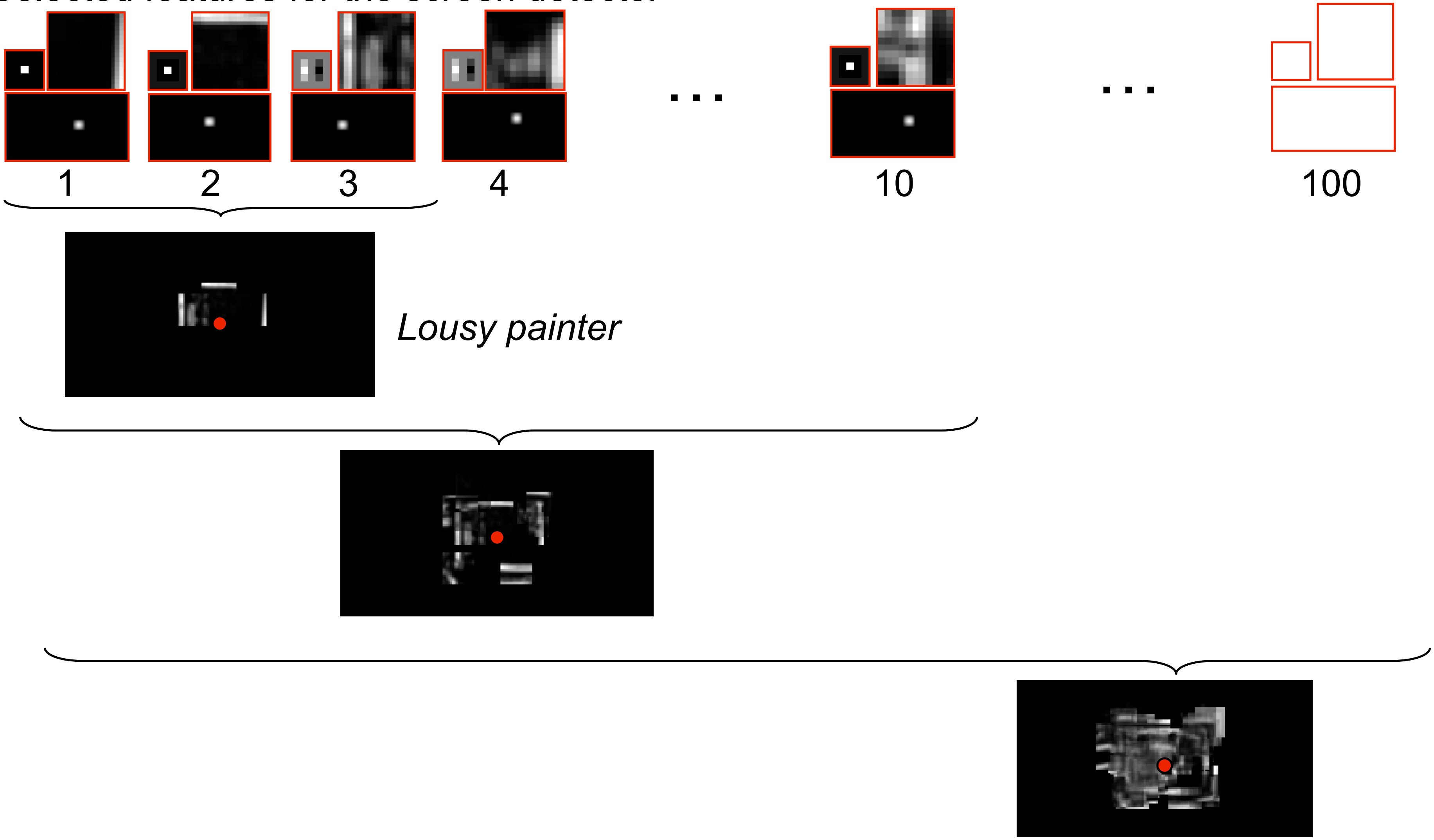


Then, we sample the feature outputs on the object center and at random locations in the background:



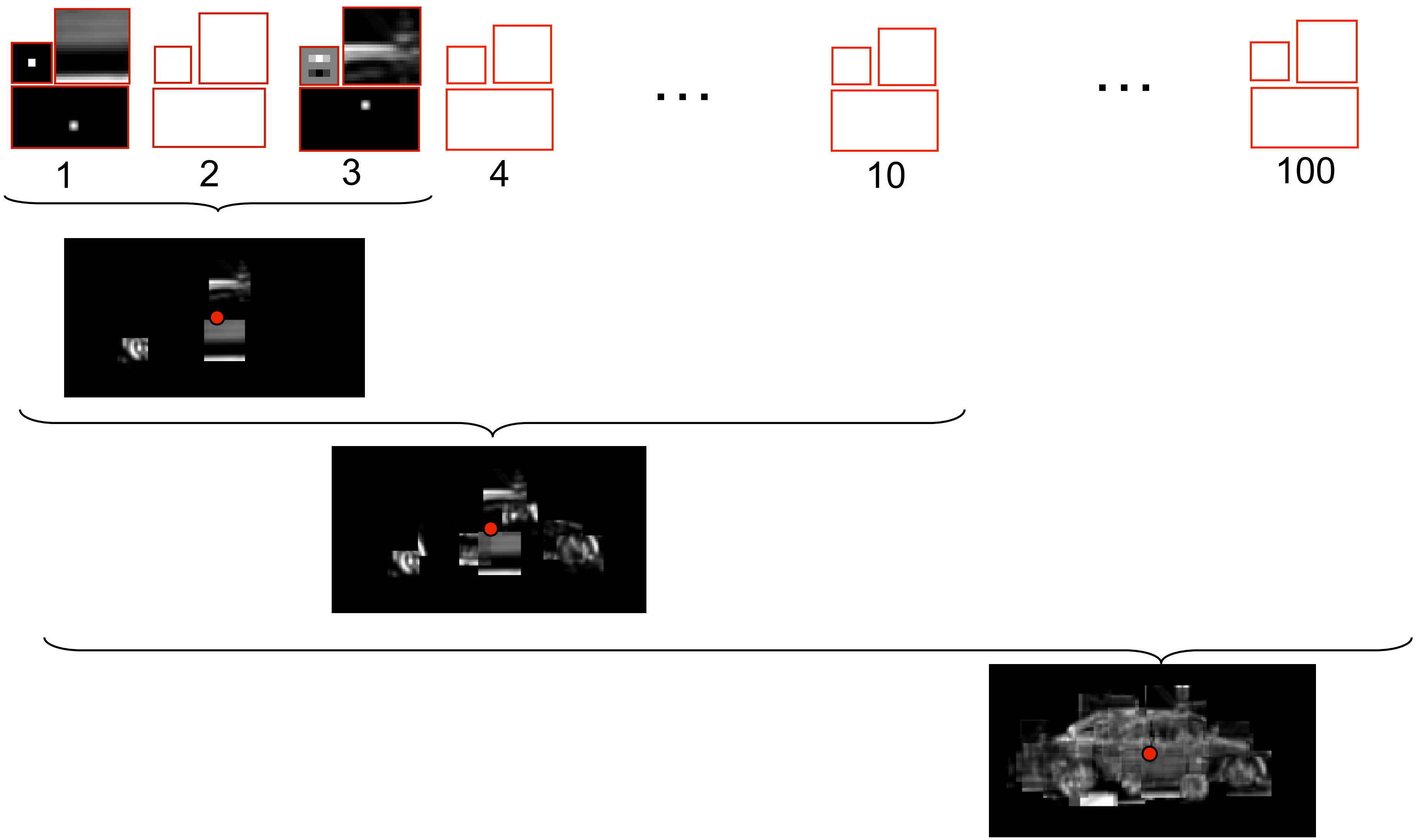
Representation and object model

Selected features for the screen detector

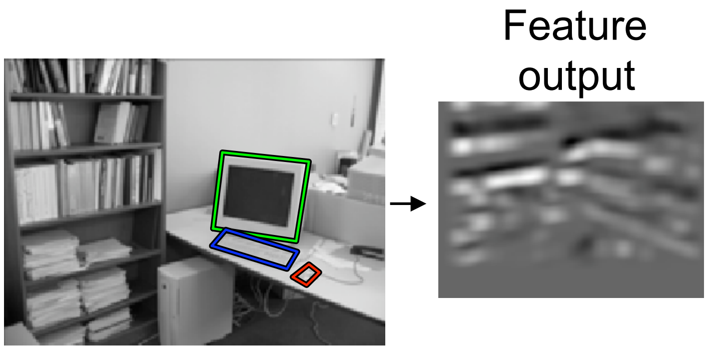


Representation and object model

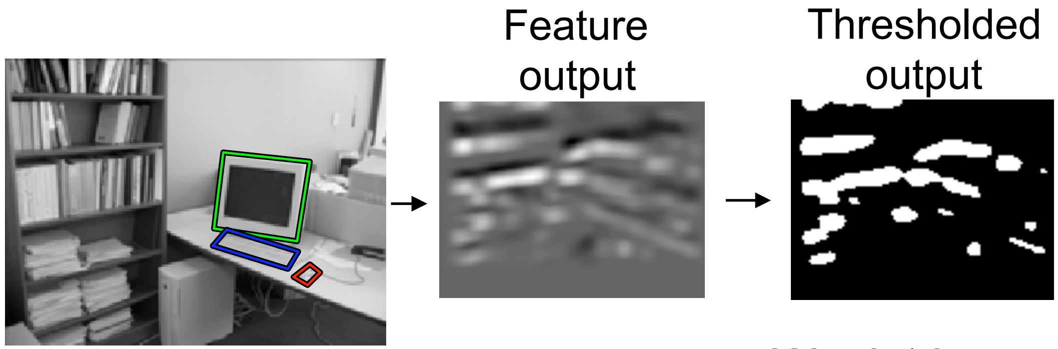
Selected features for the car detector



Example: screen detection

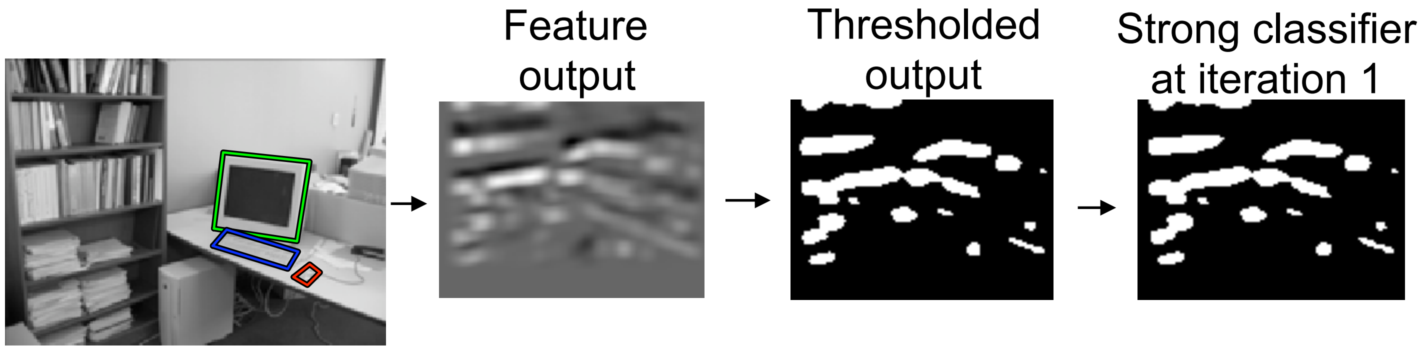


Example: screen detection

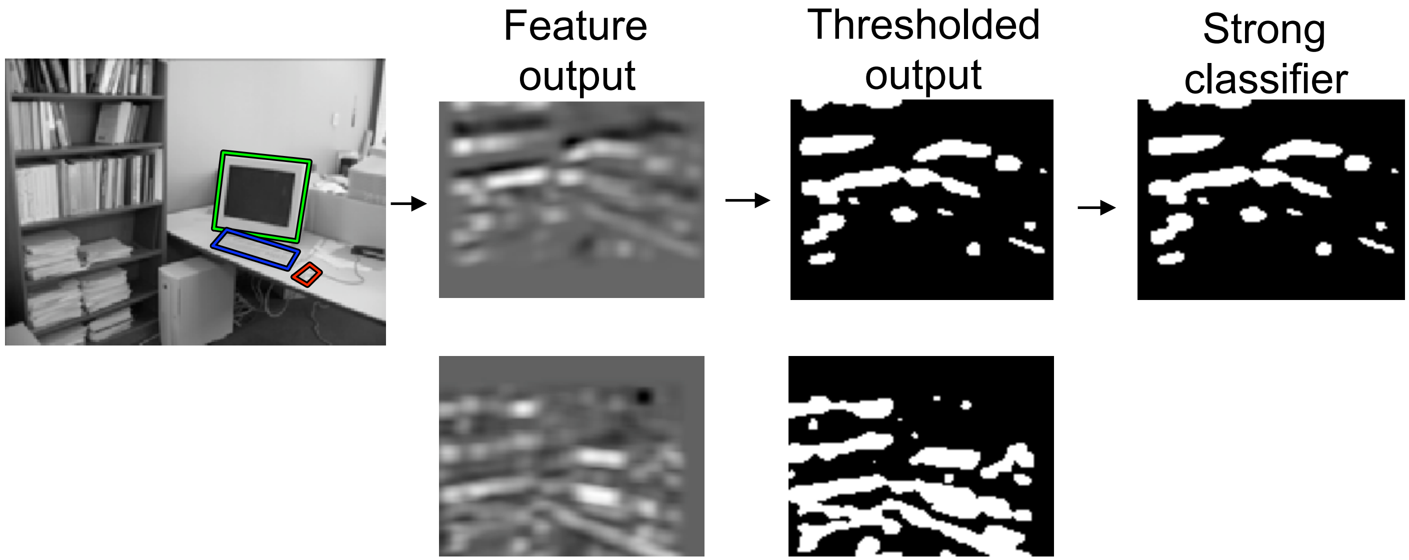


Weak 'detector'
Produces many false alarms.

Example: screen detection



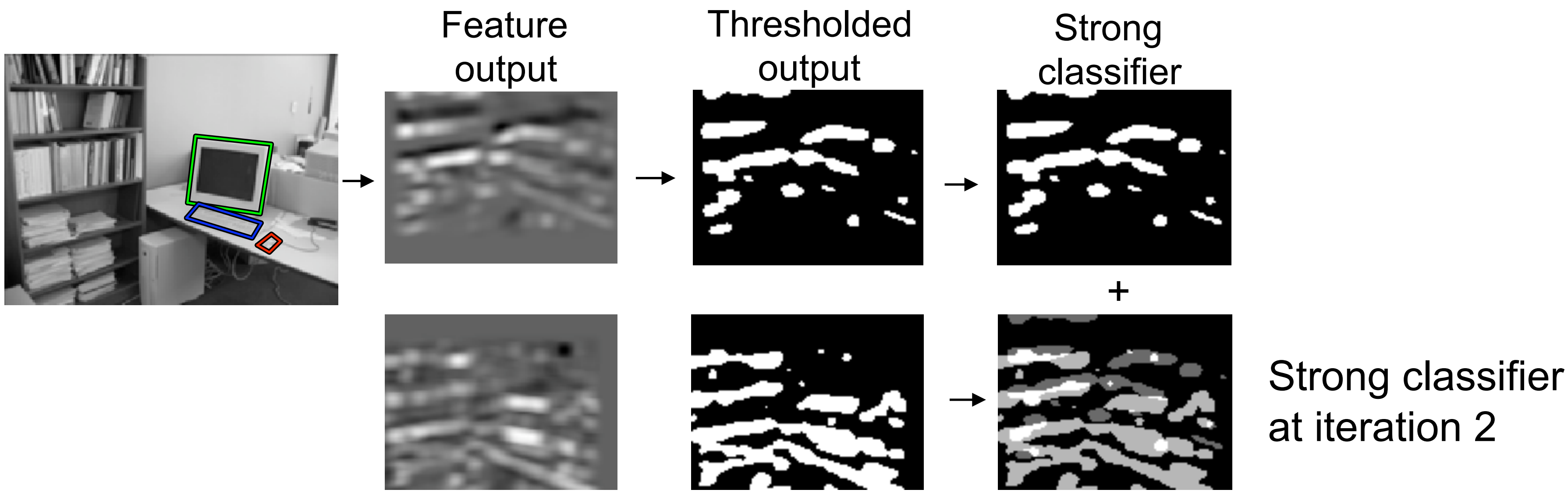
Example: screen detection



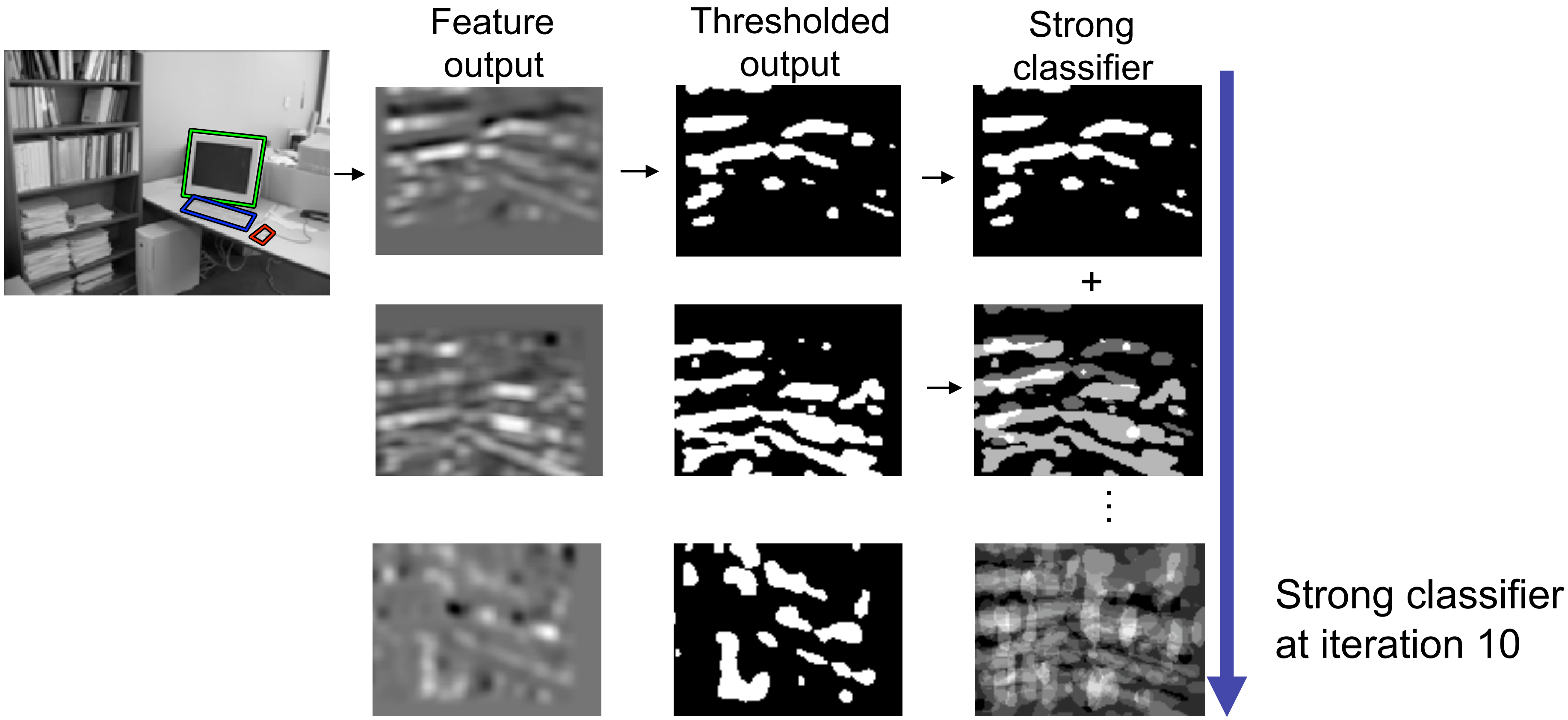
Second weak ‘detector’

Produces a different set of false alarms.

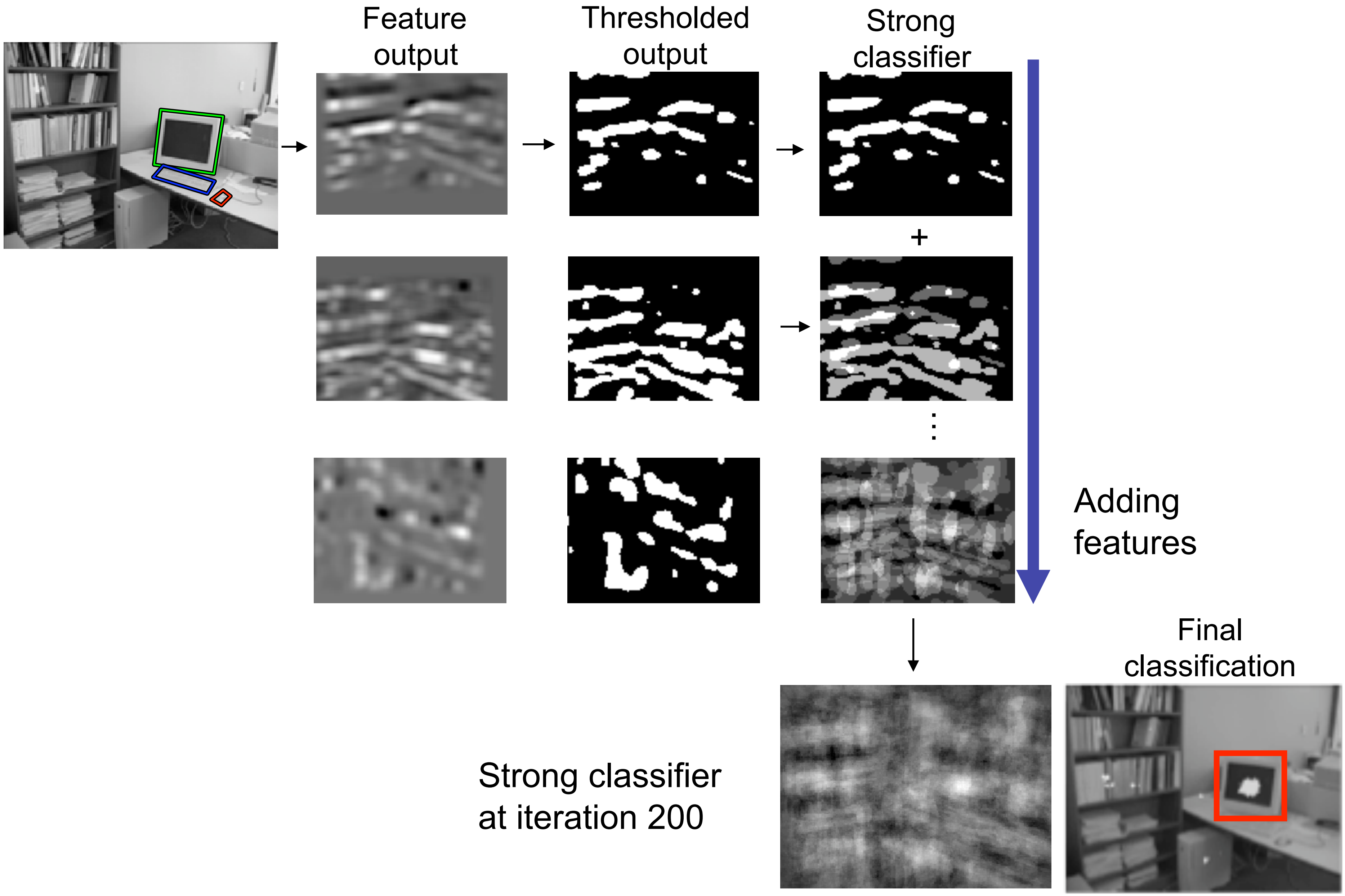
Example: screen detection



Example: screen detection



Example: screen detection

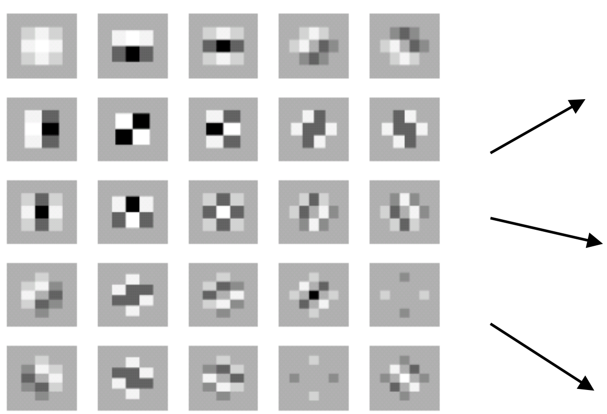


Weak detectors

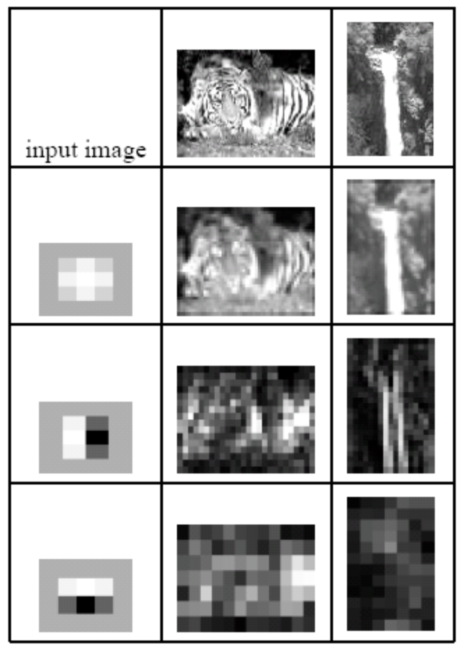
Textures of textures

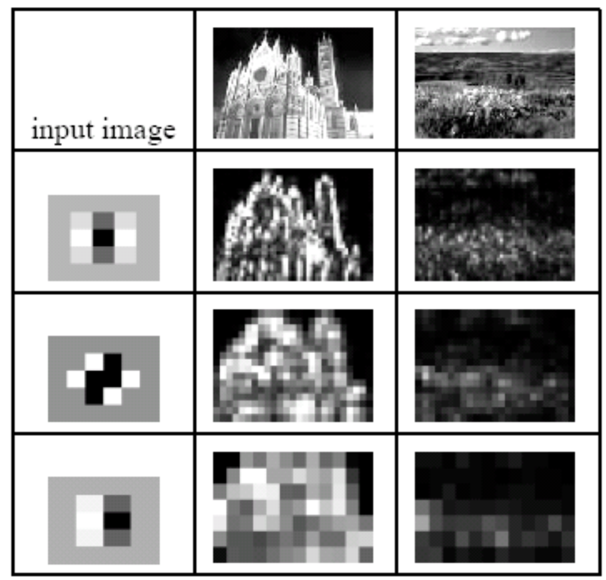
Tieu and Viola, CVPR 2000. One of the first papers to use boosting for vision.

$$g_{i,j,k} = \sum_{pixels} ||I * f_i| \downarrow_2 * f_j| \downarrow_2 * f_k$$



Every combination of three filters generates a different feature



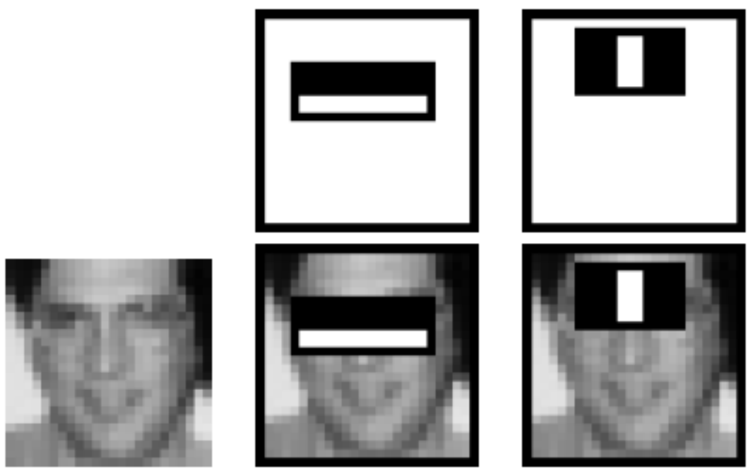


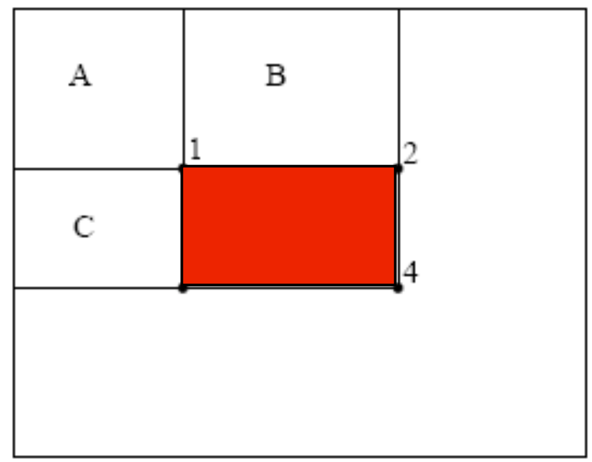
This gives thousands of features. Boosting selects a sparse subset, so computations on test time are very efficient. Boosting also avoids overfitting to some extend.

Weak detectors

Haar filters and integral image

Viola and Jones, ICCV 2001

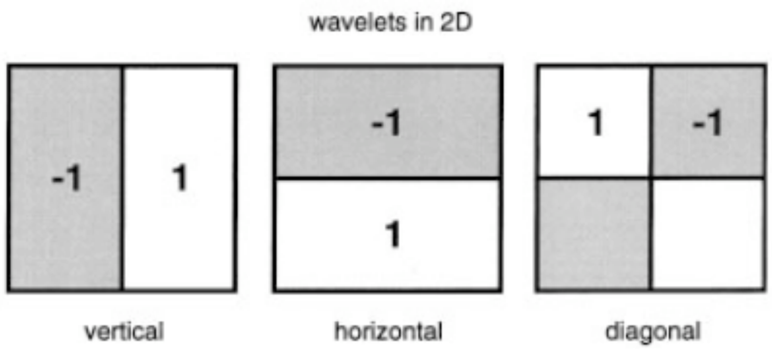


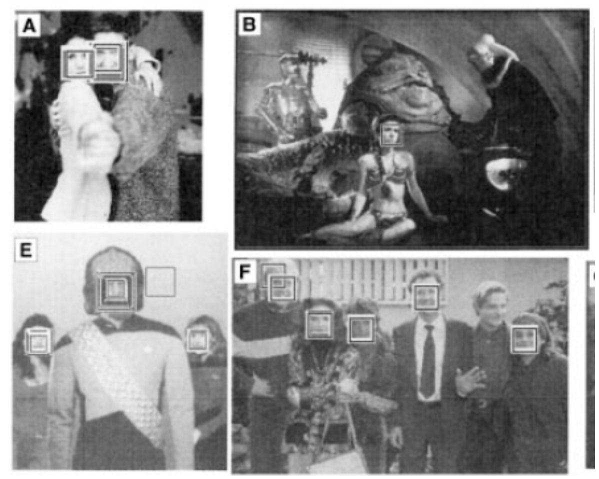


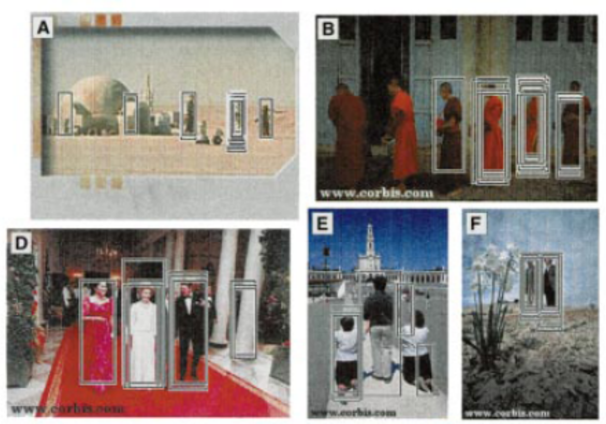
The average intensity in the block is computed with four sums independently of the block size.

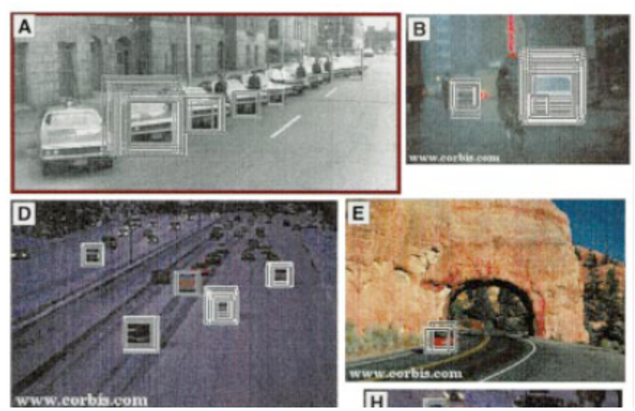
Haar wavelets

Papageorgiou & Poggio (2000)



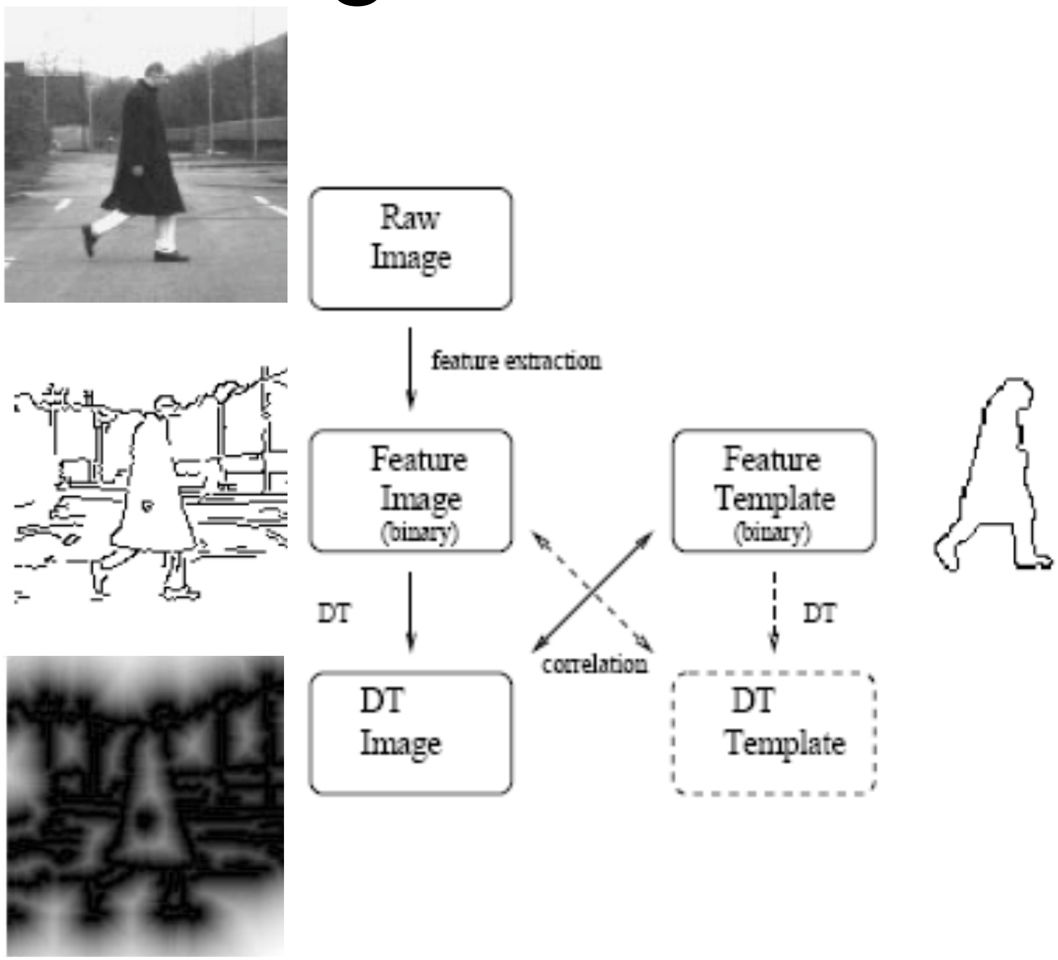


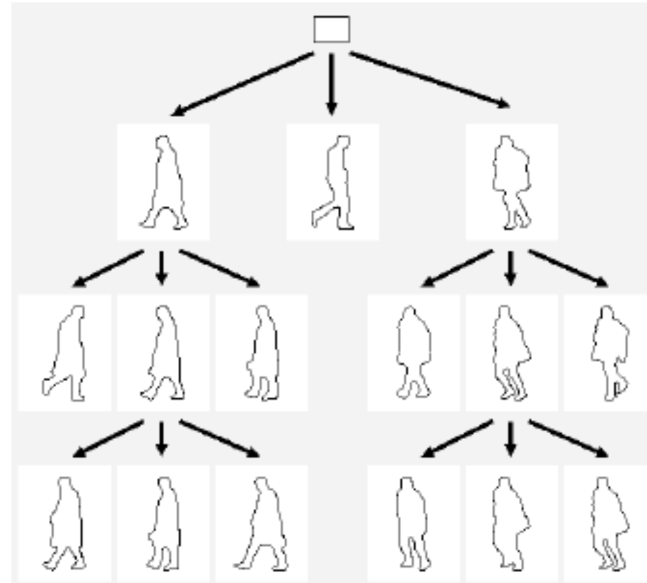




Polynomial SVM

Edges and chamfer distance





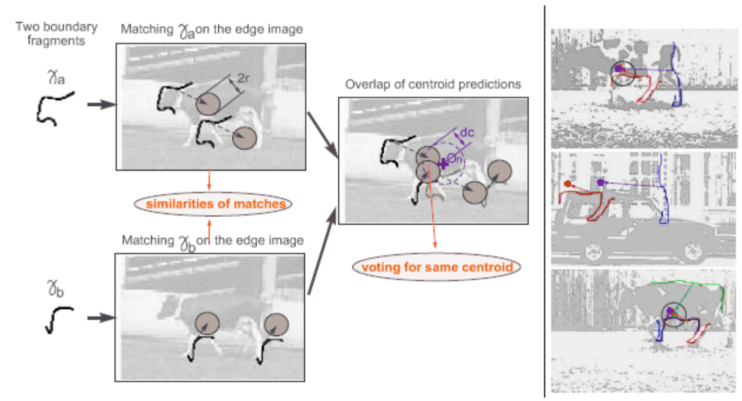
Gavrila, Philomin, ICCV 1999

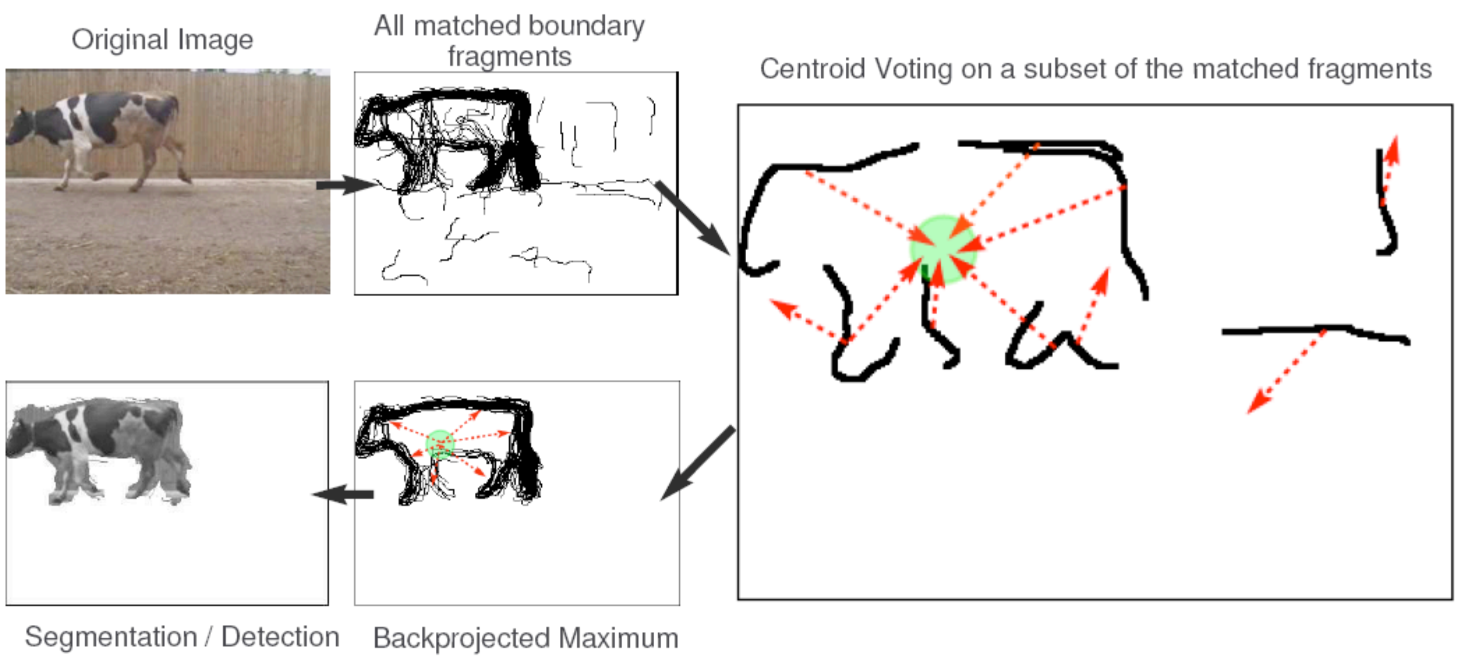
Edge fragments

J. Shotton, A. Blake, R. Cipolla.
Multi-Scale Categorical Object Recognition
Using Contour Fragments. In *IEEE Trans. on PAMI*, 30(7):1270-1281, July 2008.

Fig. 1. **Object recognition using contour fragments.** Our innate biological vision system is able to interpret spatially arranged local fragments of contour to recognize the objects present. In this work we show that an automatic computer vision system can also successfully exploit the cue of contour for object recognition.

Opelt, Pinz, Zisserman, ECCV 2006



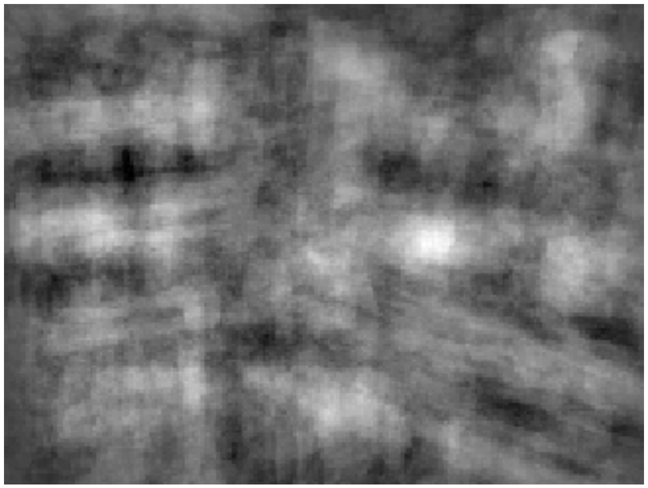


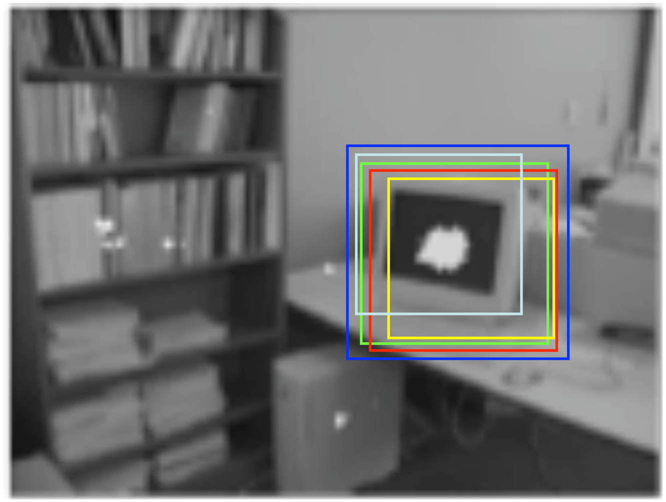
Weak detectors

Other weak detectors:

- Carmichael, Hebert 2004
- Yuille, Snow, Nitzbert, 1998
- Amit, Geman 1998
- Papageorgiou, Poggio, 2000
- Heisele, Serre, Poggio, 2001
- Agarwal, Awan, Roth, 2004
- Schneiderman, Kanade 2004
- …

Maximal suppression



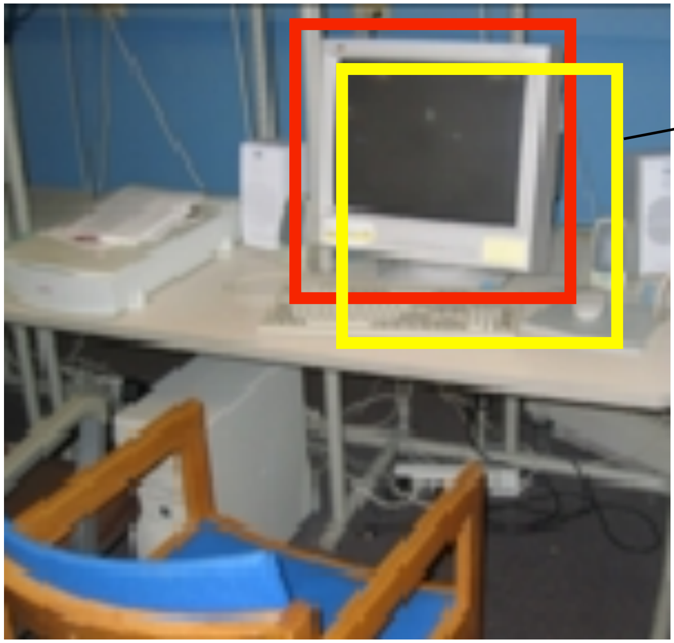


Detect local maximum of the response. We are only allowed detecting each object once. The rest will be considered false alarms.

This post-processing stage can have a very strong impact in the final performance.

Evaluation

When do we have a correct detection?



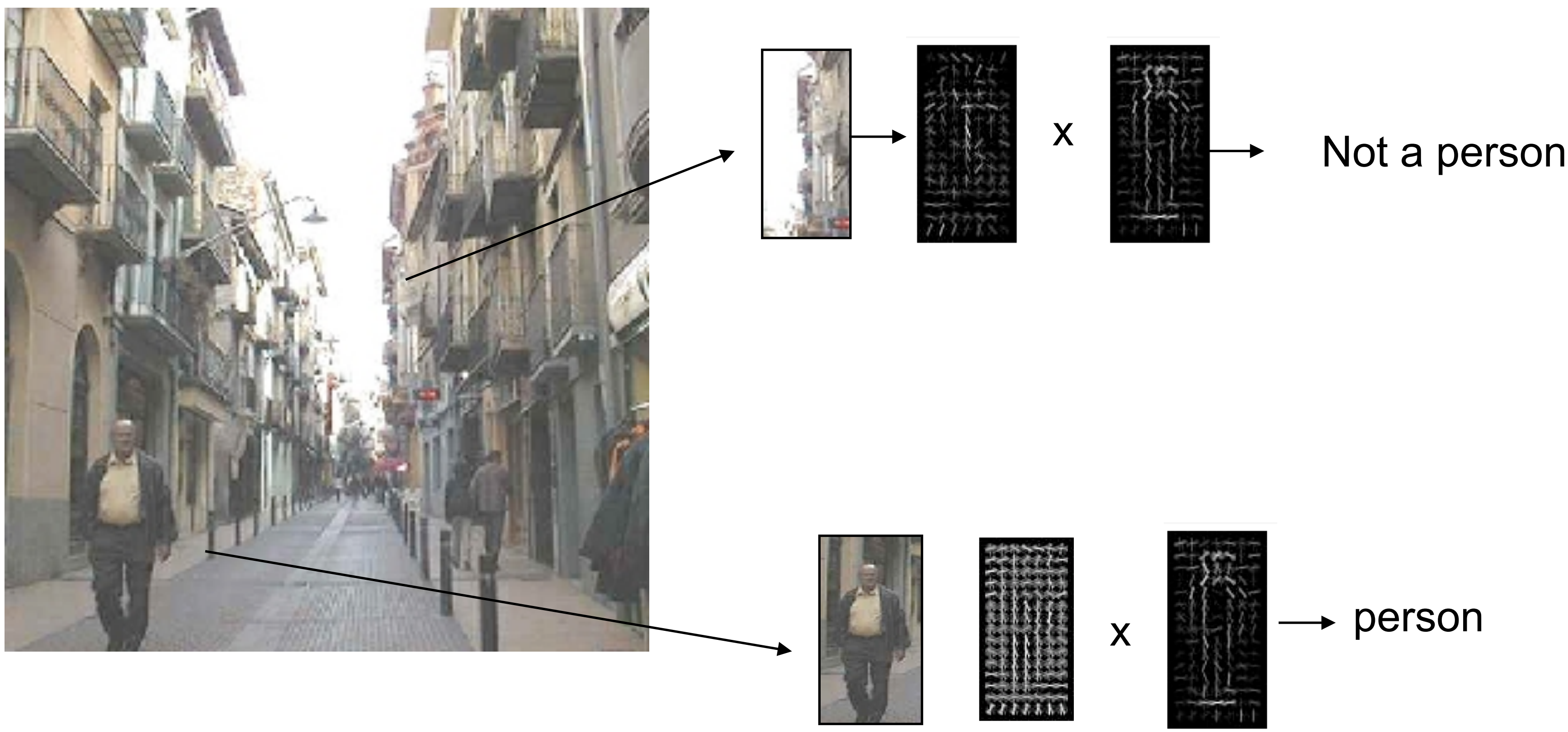
Is this correct?

$$\frac{\text{Area intersection}}{\text{Area union}} > 0.5$$

- ROC
- Precision-recall

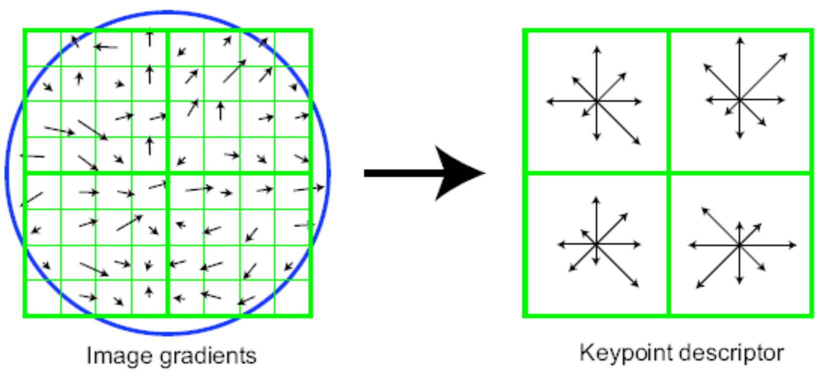
Histograms of oriented gradients

Dalal & Trigs, 2006



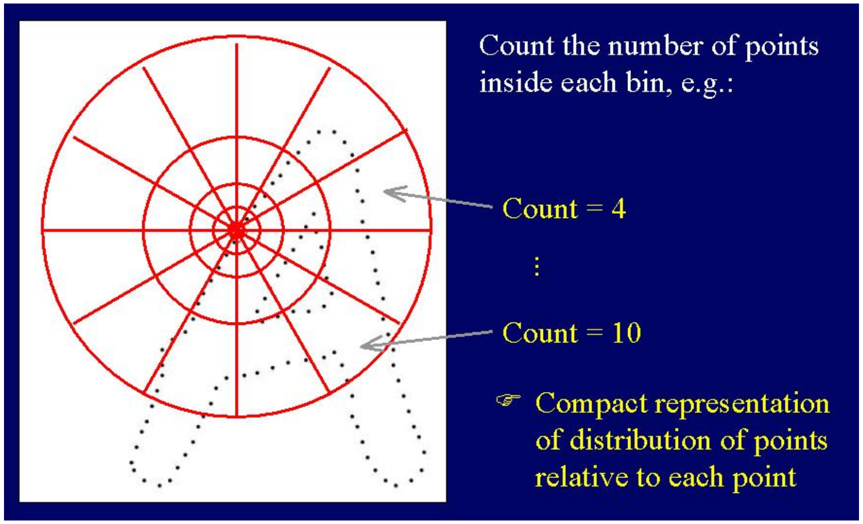
Histograms of oriented gradients

- SIFT, D. Lowe, ICCV 1999

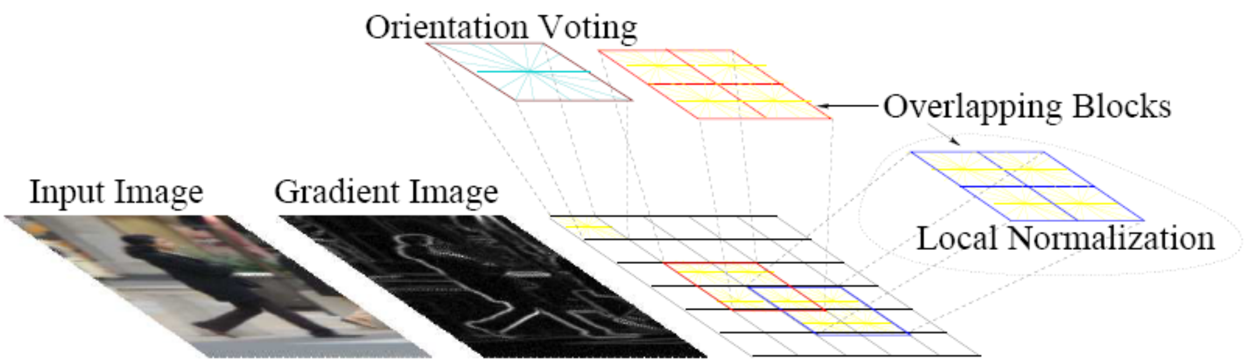


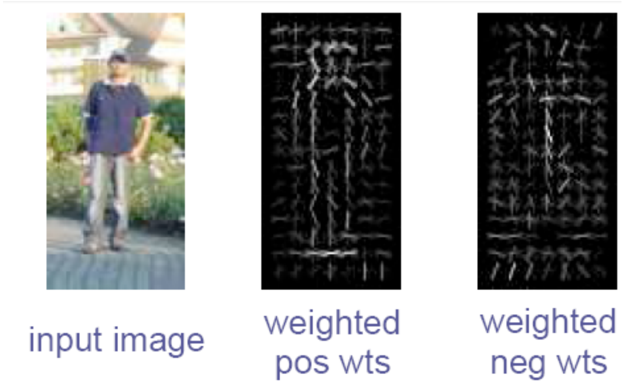
- Shape context

Belongie, Malik, Puzicha, NIPS 2000



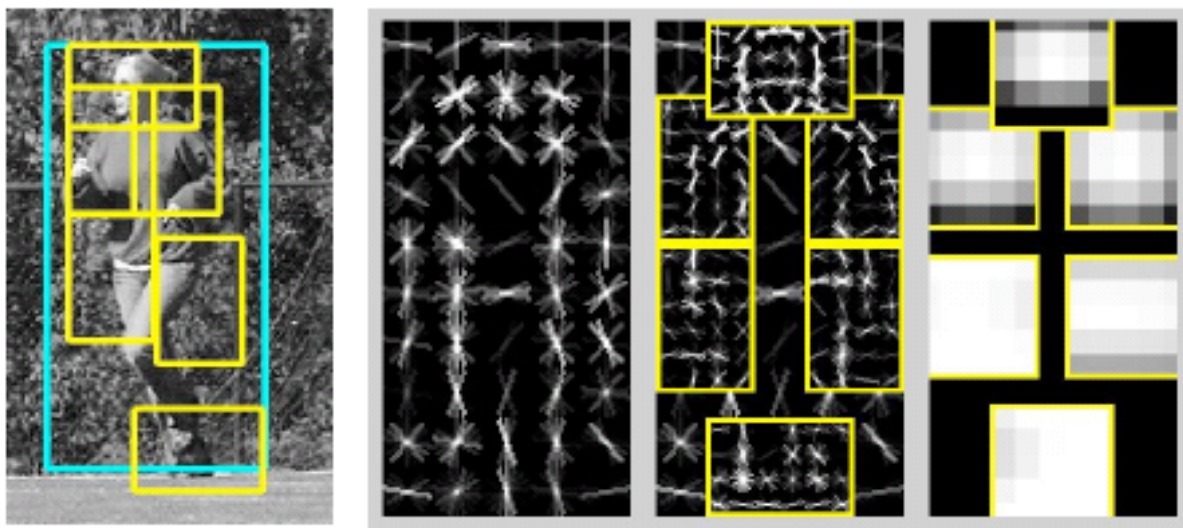
- Dalal & Trigs, 2006

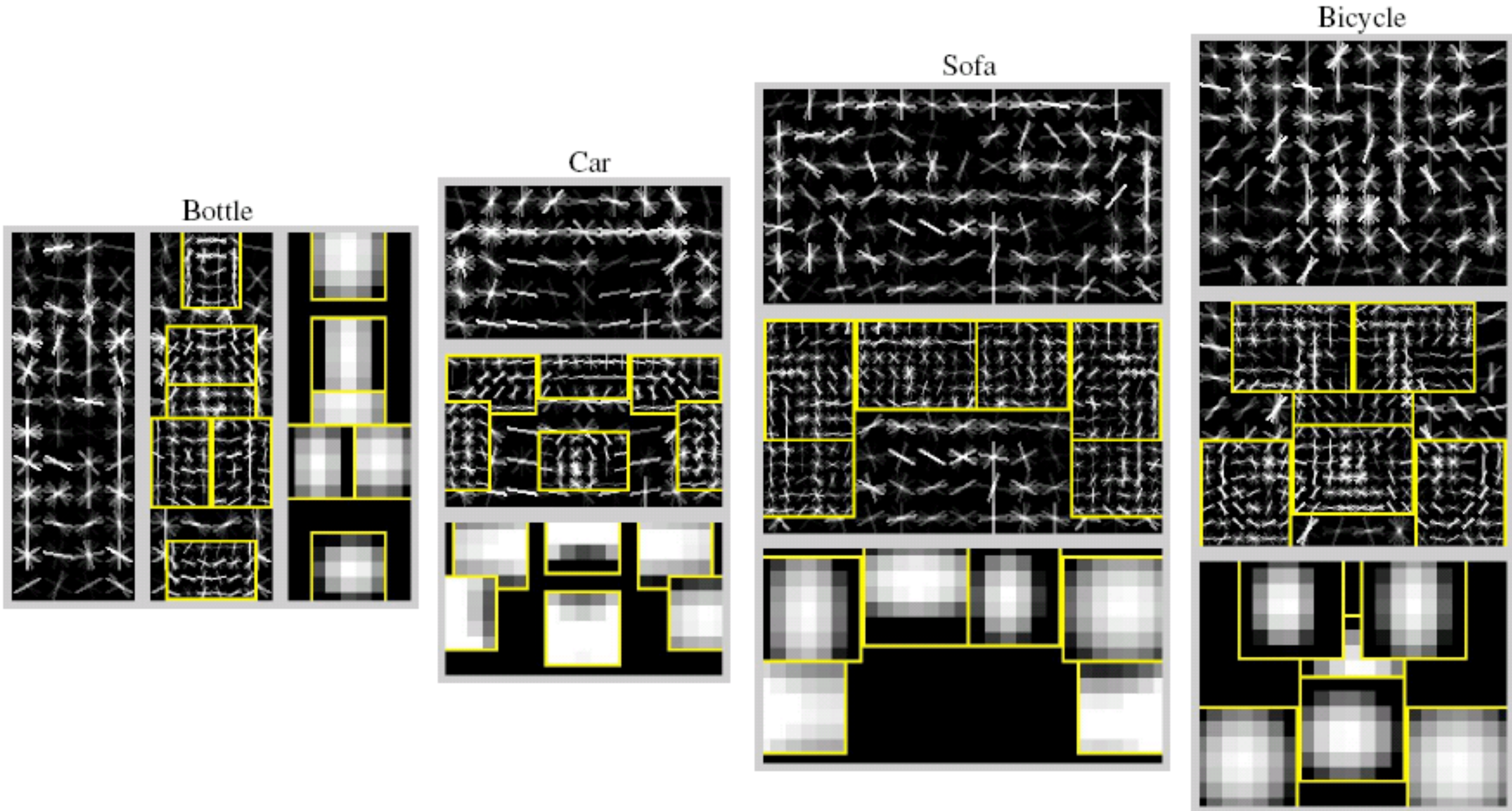




Adding parts

Felzenszwalb, McAllester, Ramanan. 2008.

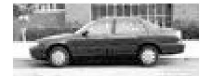


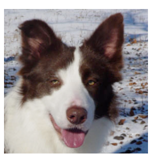


Felzenszwalb, McAllester, Ramanan. 2008.

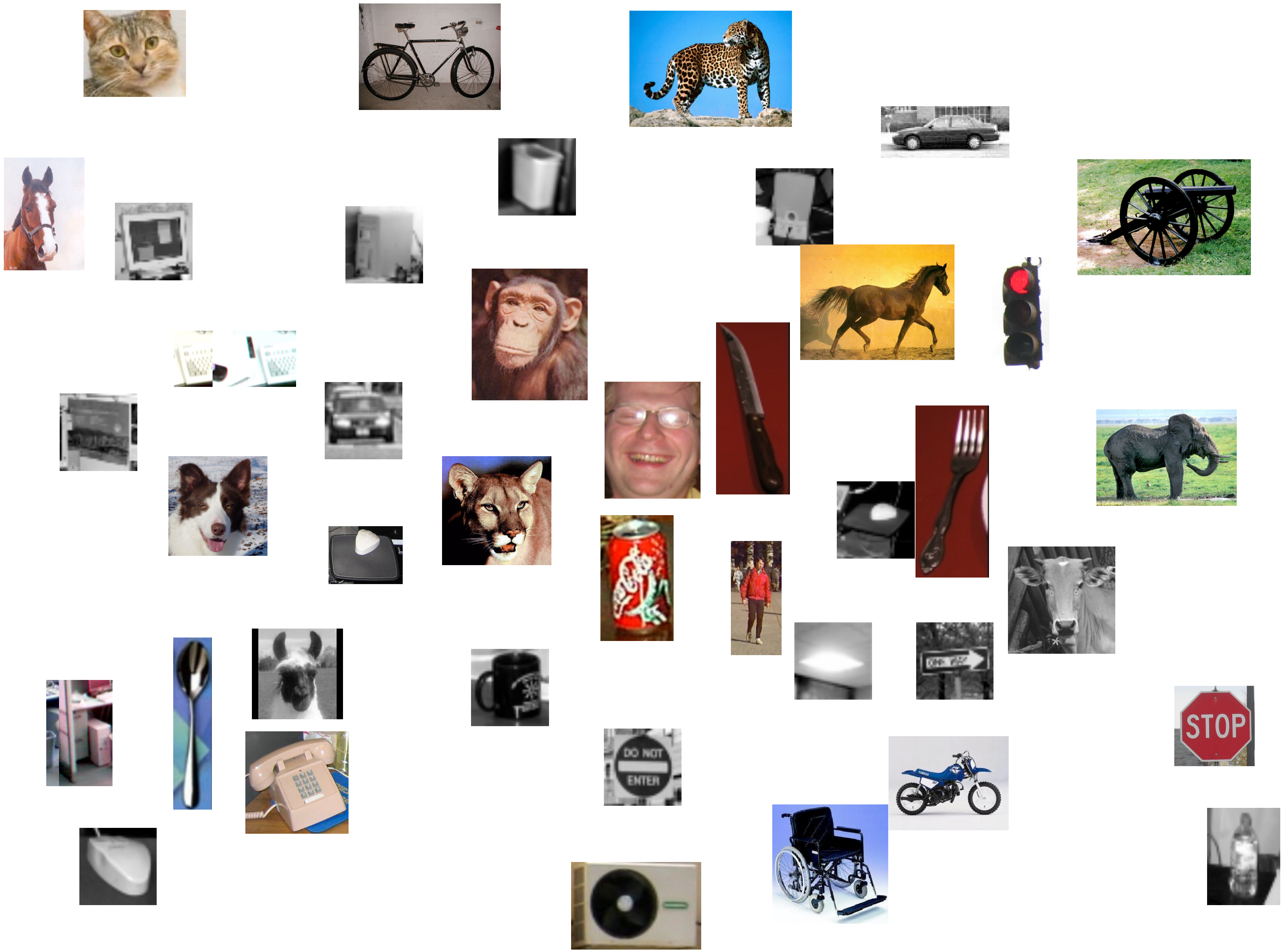
Felzenszwalb, McAllester, Ramanan. 2008.

Beyond single classes

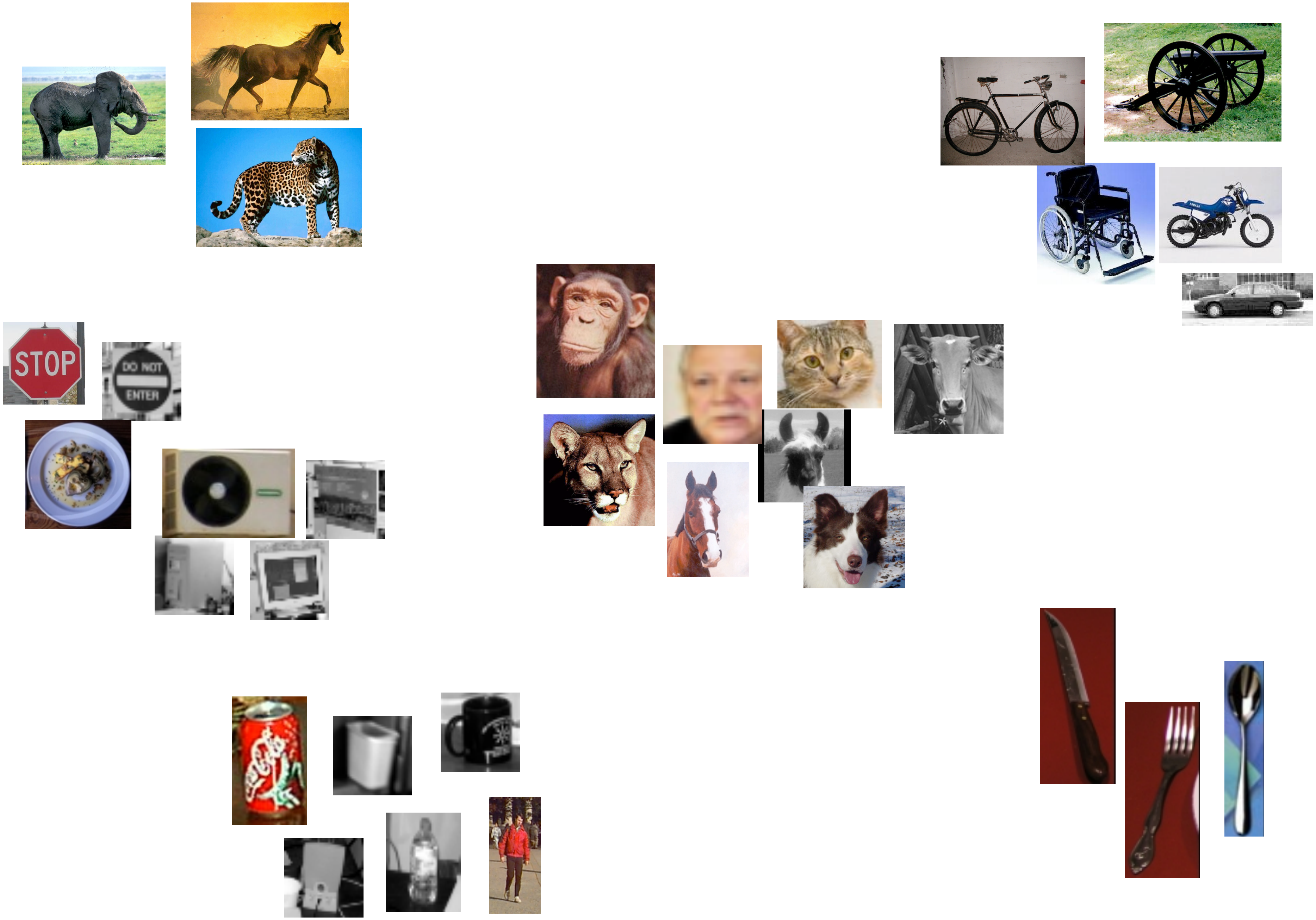




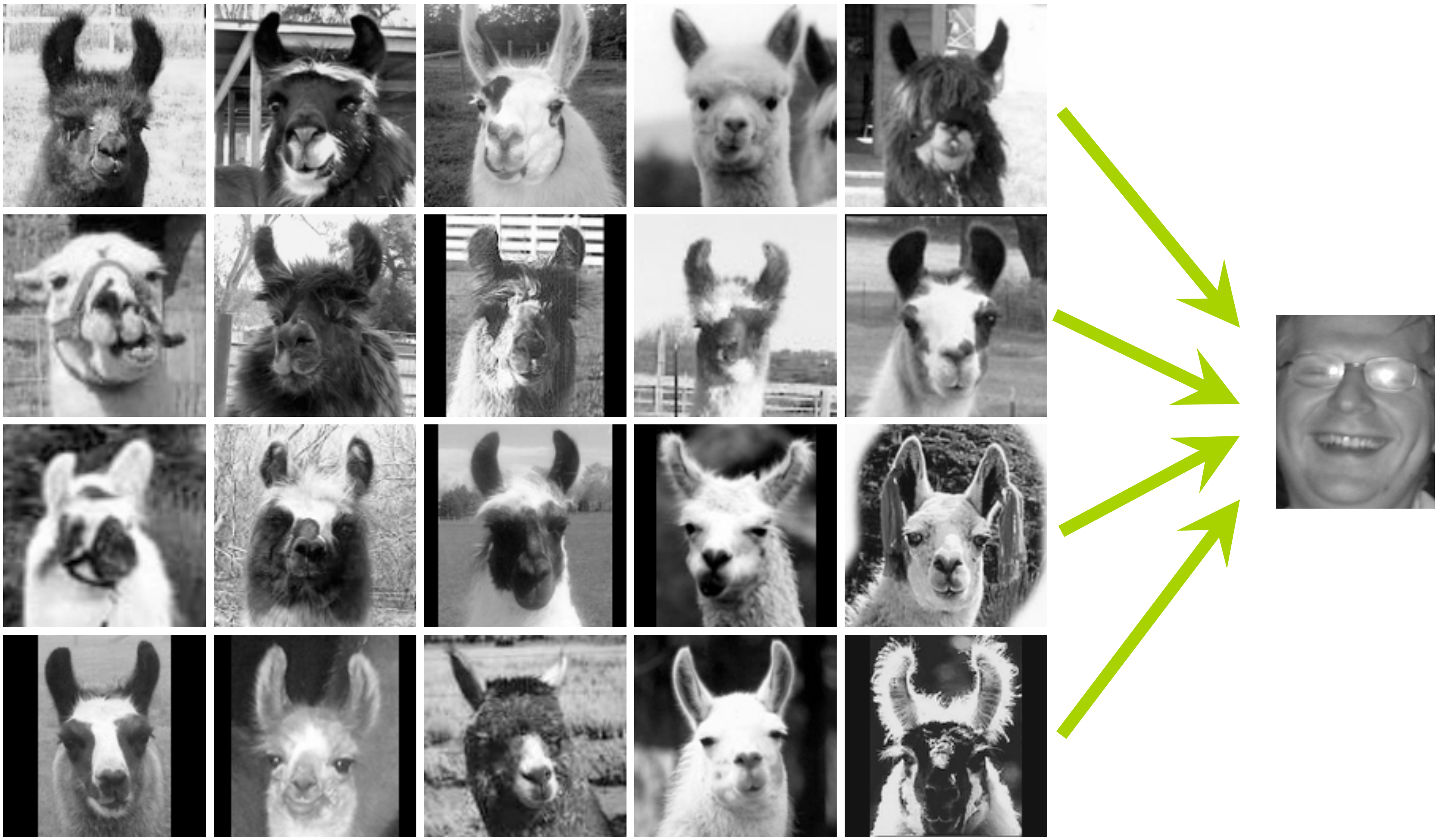

STOP


DO NOT
ENTER
STOP


STOP
DO NOT
ENTER


STOP
DO NOT
ENTER

Generalizing Across Categories



Can we transfer knowledge from one object category to another?

Slide by Erik Sudderth

How many categories?

“Muchas”

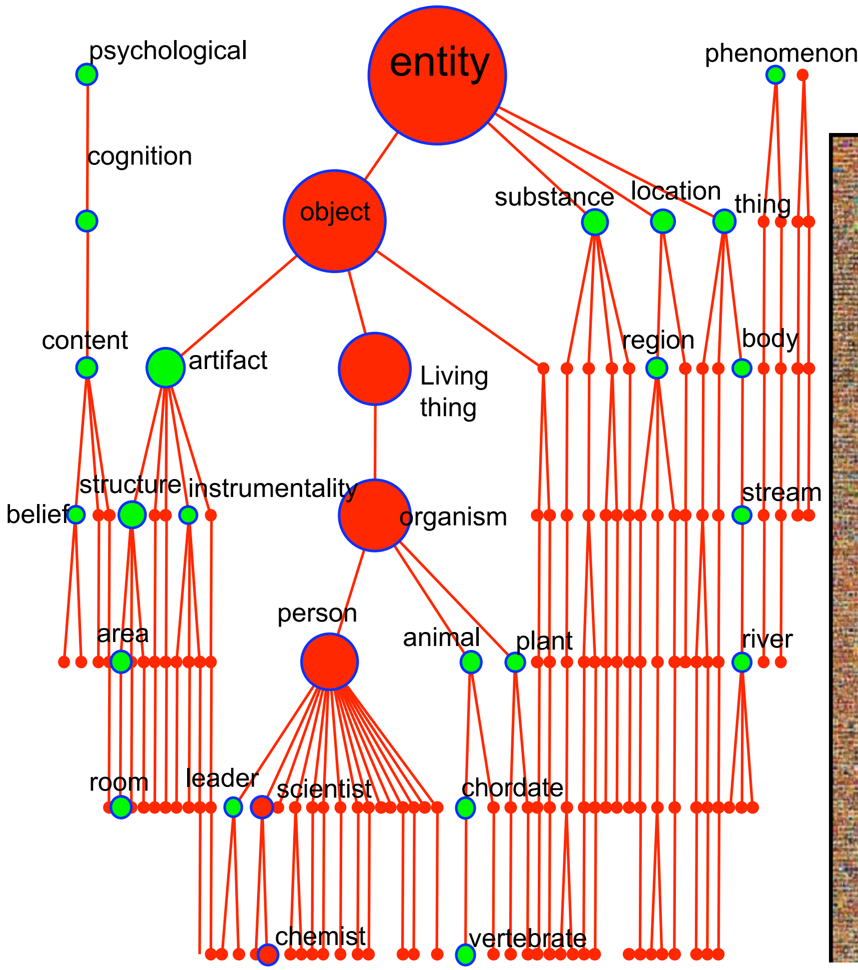
Slide by Aude Oliva

How many object categories are there?

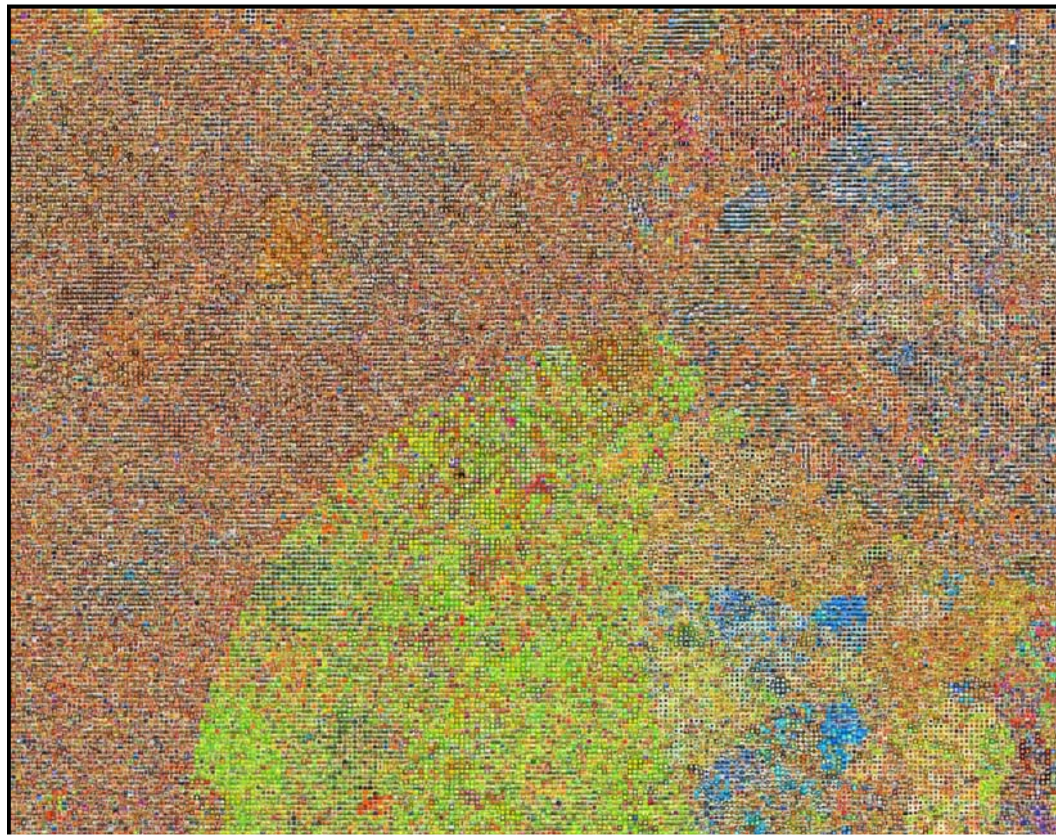
Biederman 1987

Categorical hierarchies

Categories can be organized in hierarchies (tree structures are commonly used)



From Wordnet



Which level of categorization is the right one?

Car is an object composed of:

a few doors, four wheels (not all visible at all times), a roof, front lights, windshield

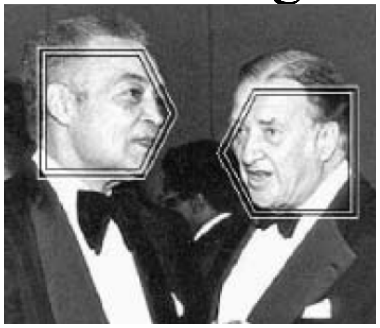
If you are thinking in buying a car, you might want to be a bit more specific about your categorization.

Multiclass object detection
the not so early days

Multiclass object detection
the not so early days

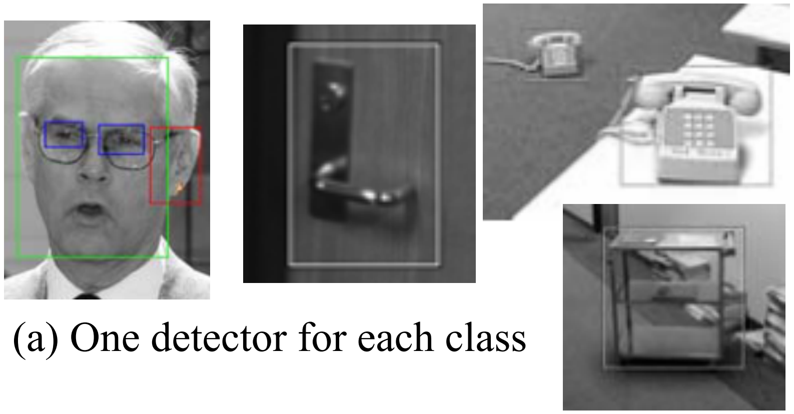
Using a set of independent binary classifiers was a common strategy:

• Viola-Jones extension for dealing with rotations



- two cascades for each view

• Schneiderman-Kanade multiclass object detection



(a) One detector for each class

(b) For cars, classifiers are trained on 8 viewpoints

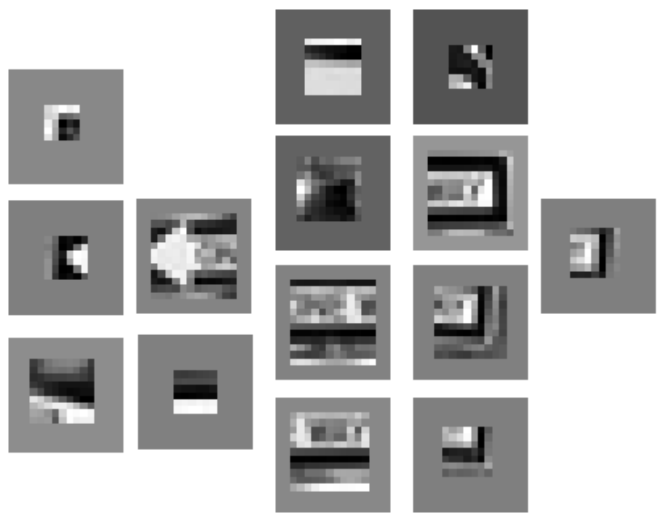
There is nothing wrong with this approach if you have access to lots of training data and you do not care about efficiency.

Some symptoms of one-vs-all multiclass approaches

What is the best representation to detect a traffic sign?

Very regular object: template matching will do the job

Parts derived from training a binary classifier.

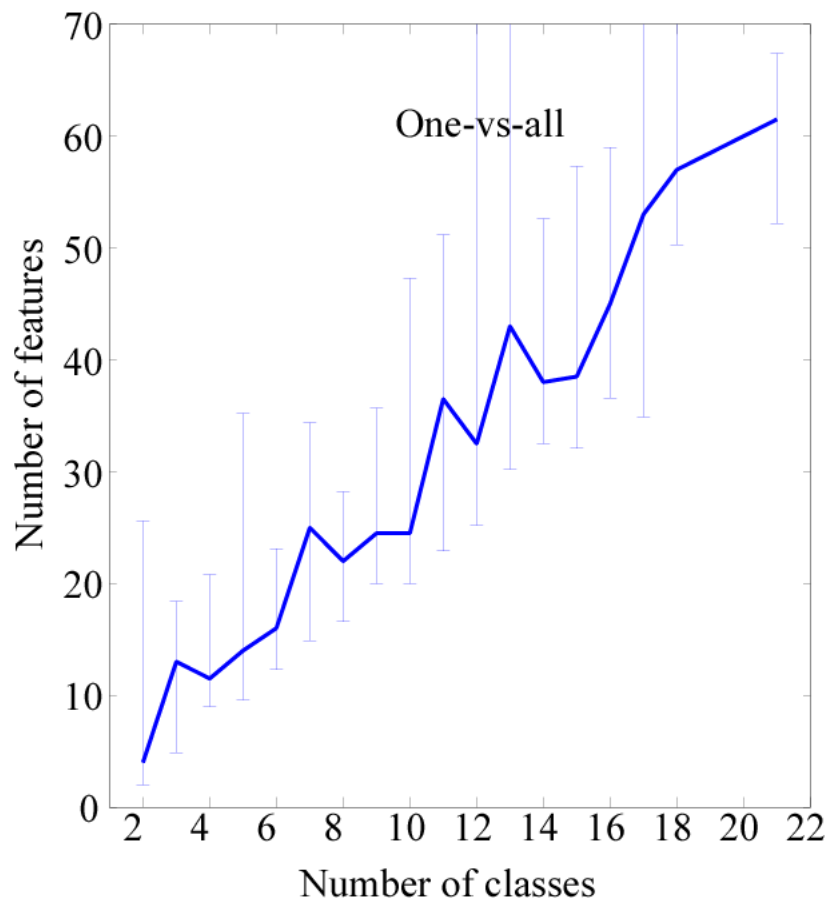


~100% detection rate with 0 false alarms

Some of these parts cannot be used for anything else than this object.

Some symptoms of one-vs-all multiclass approaches

Computational cost grows linearly with Nclasses * Nviews * Nstyles …



Shared features

- Is learning the object class 1000 easier than learning the first?

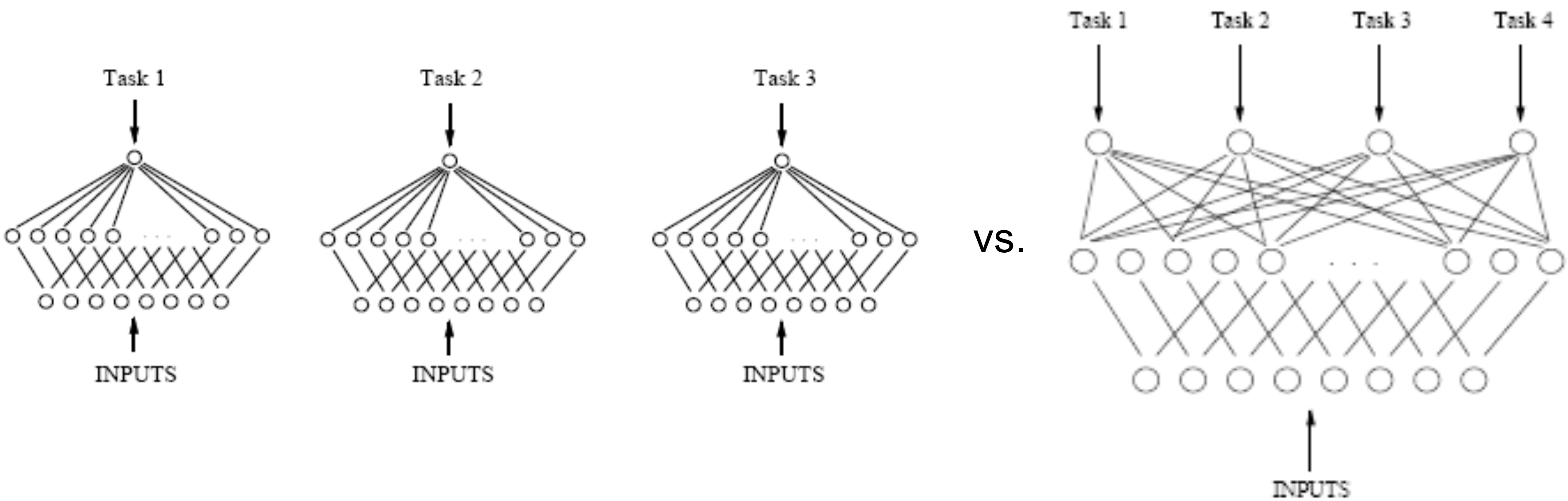
…

- Can we transfer knowledge from one object to another?
- Are the shared properties interesting by themselves?

Multitask learning

R. Caruana. Multitask Learning. ML 1997

"MTL improves generalization by leveraging the domain-specific information contained in the training signals of *related* tasks. It does this by training tasks in parallel while using a shared representation".



Sejnowski & Rosenberg 1986; Hinton 1986; Le Cun et al. 1989; Suddarth & Kergosien 1990; Pratt et al. 1991; Sharkey & Sharkey 1992; …

Multitask learning

R. Caruana. Multitask Learning. ML 1997

Primary task: detect door knobs

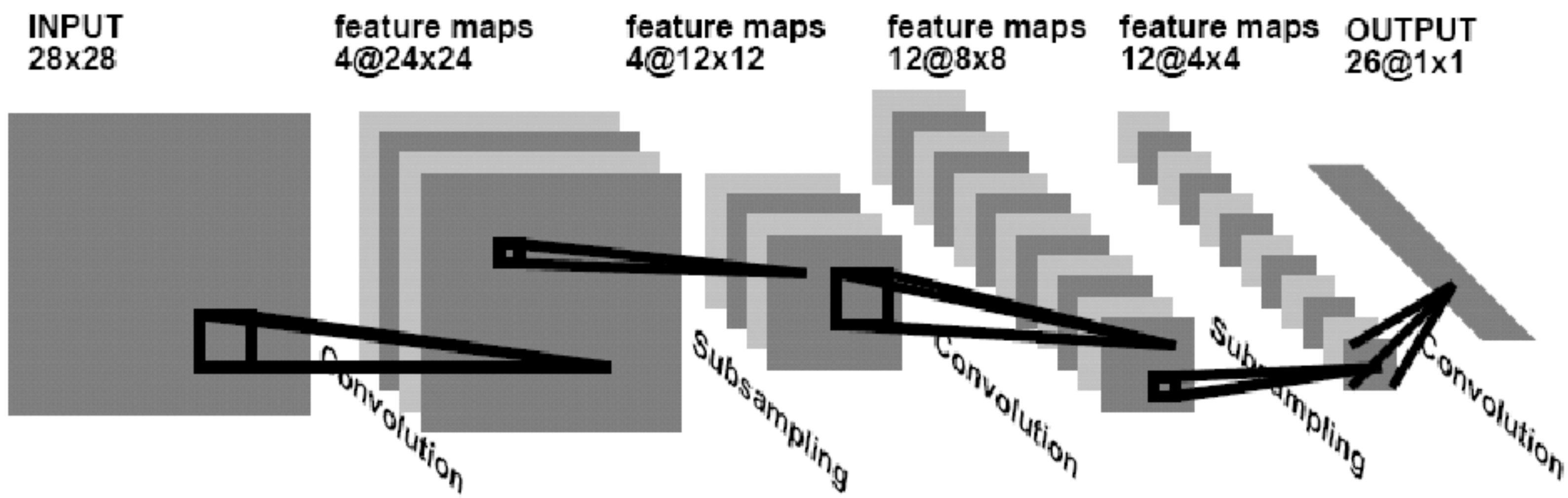
Tasks used:

- horizontal location of doorknob
- single or double door
- horizontal location of doorway center
- width of doorway
- horizontal location of left door jamb
- horizontal location of right door jamb
- width of left door jamb
- width of right door jamb
- horizontal location of left edge of door
- horizontal location of right edge of door

ROOT-MEAN SQUARED ERROR ON TEST SET

TASK	Single Task Backprop (STL)			MTL
	6HU	24HU	96HU	120HU
Doorknob Loc	.085	.082	**.081**	**.062**

Convolutional Neural Network



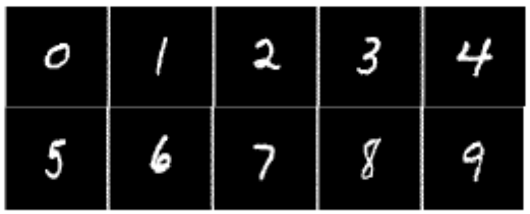
Le Cun et al, 98

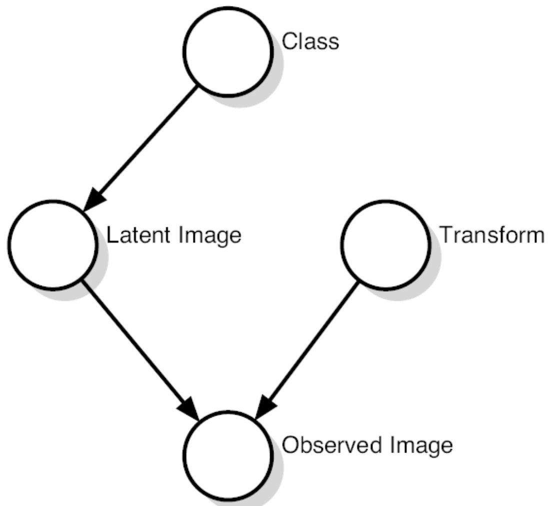
Translation invariance is already built into the network

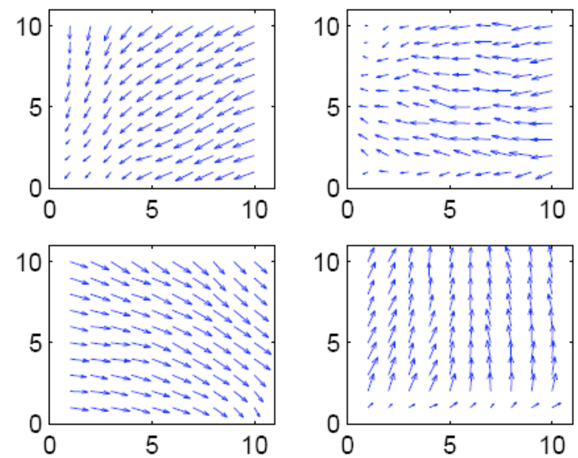
The output neurons share all the intermediate levels

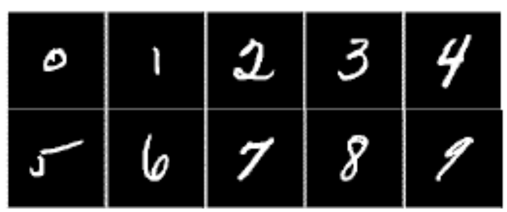
Sharing transformations

Miller, E., Matsakis, N., and Viola, P. (2000). Learning from one example through shared densities on transforms. In IEEE Computer Vision and Pattern Recognition.







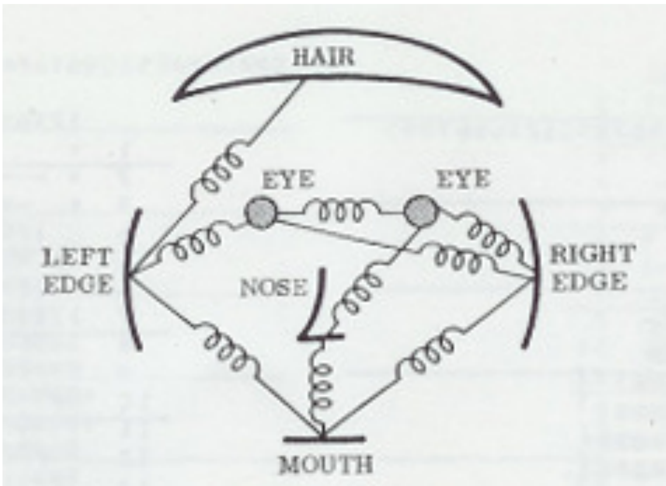


Transformations are shared
and can be learnt from other tasks.

Training Samples	Basic Hausdorff	With Congealing	With Transform Density
1000	92.5%	87.3%	96.4%
1	29.7%	60.0%	89.3%

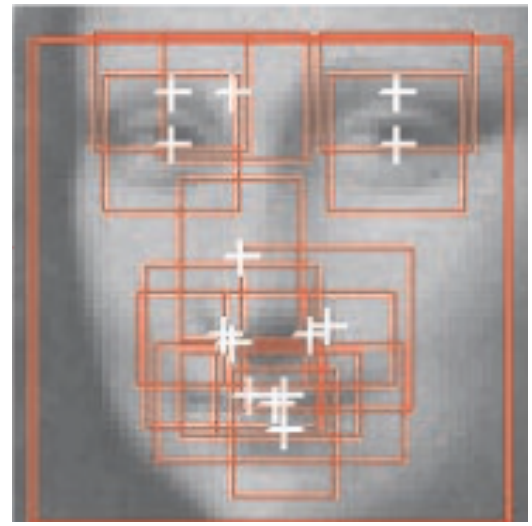
Sharing in constellation models

(next Wednesday)



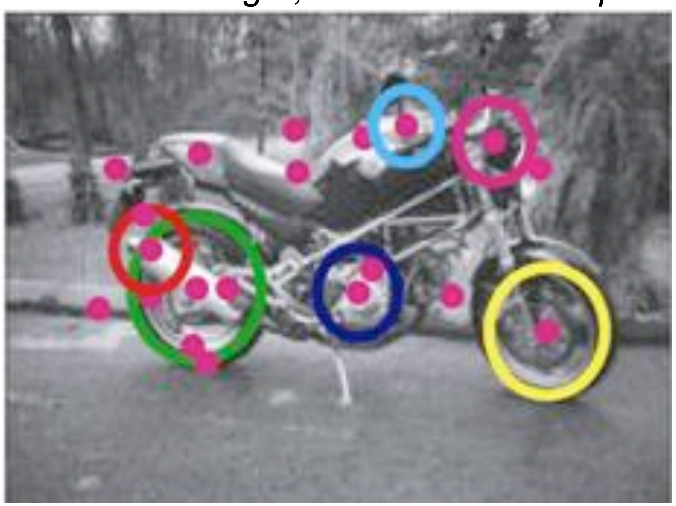
Pictorial Structures

Fischler & Elschlager, IEEE Trans. Comp. 1973



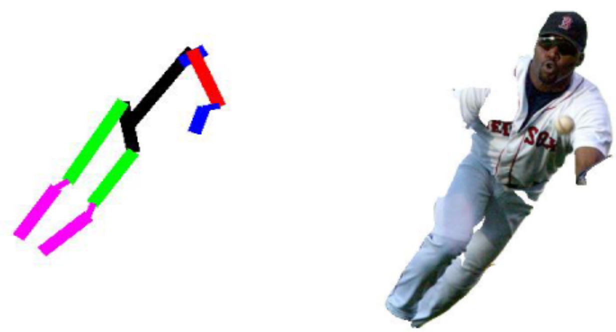
SVM Detectors

Heisele, Poggio, et. al., NIPS 2001



Constellation Model

Fergus, Perona, & Zisserman, CVPR 2003

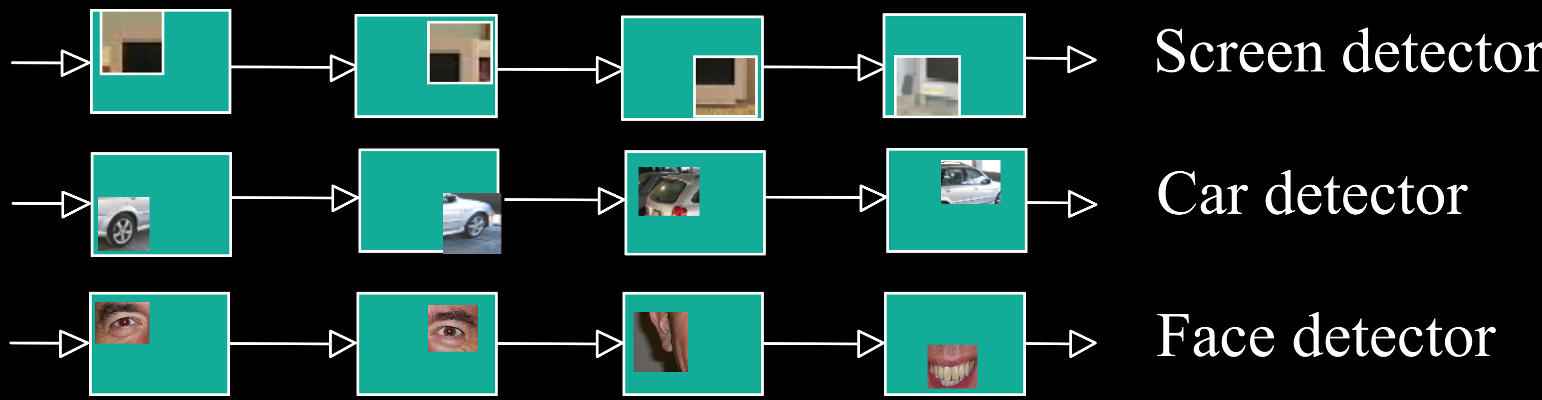


Model-Guided Segmentation

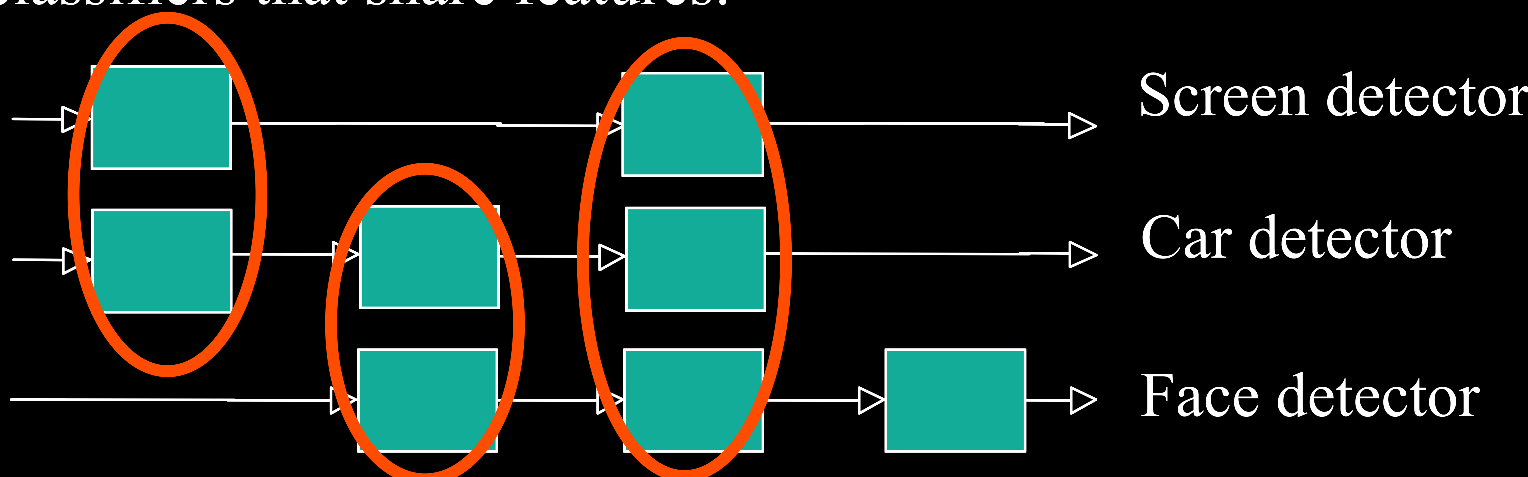
Mori, Ren, Efros, & Malik, CVPR 2004

Additive models and boosting
(more details on Wednesday)

- Independent binary classifiers:

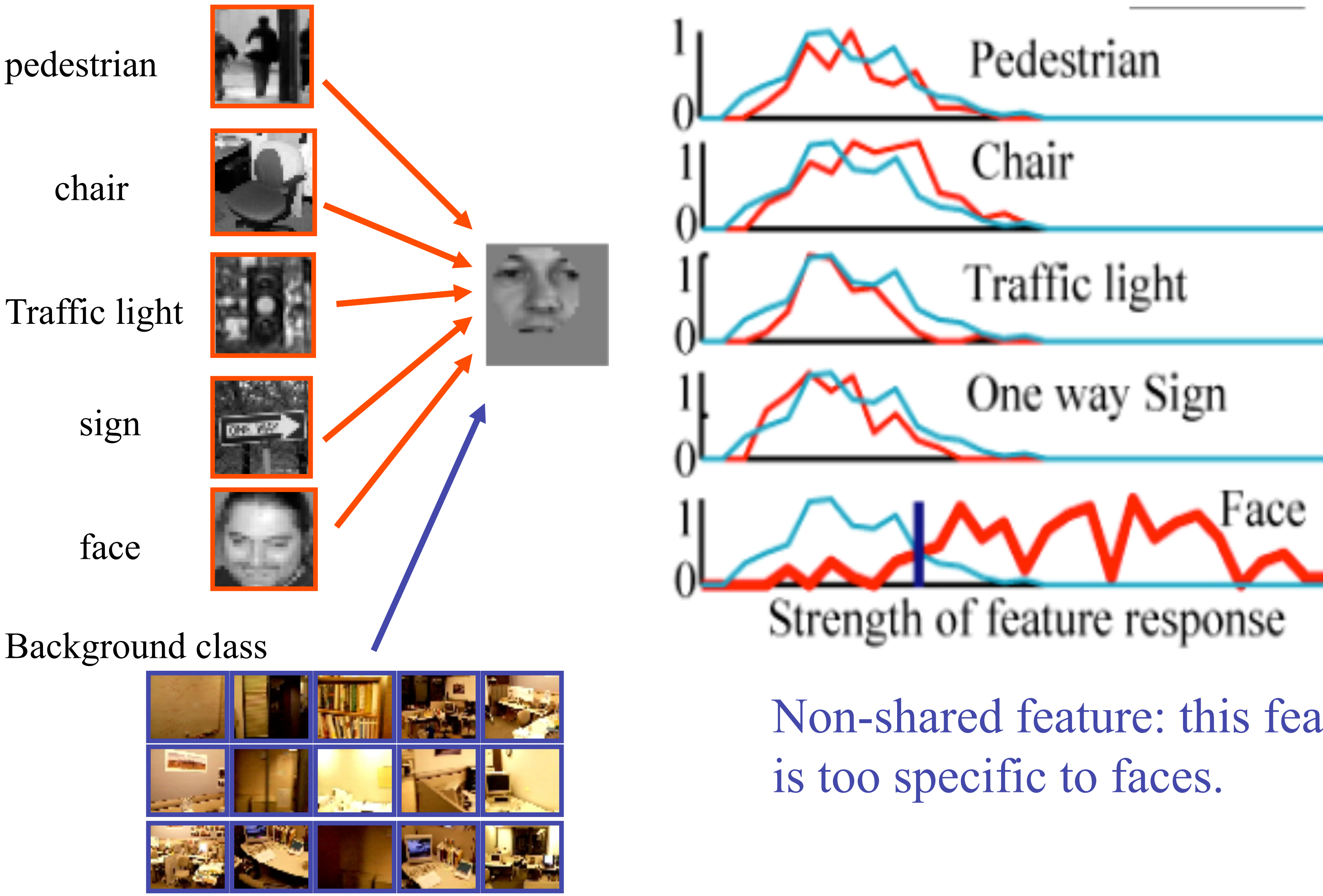


- Binary classifiers that share features:



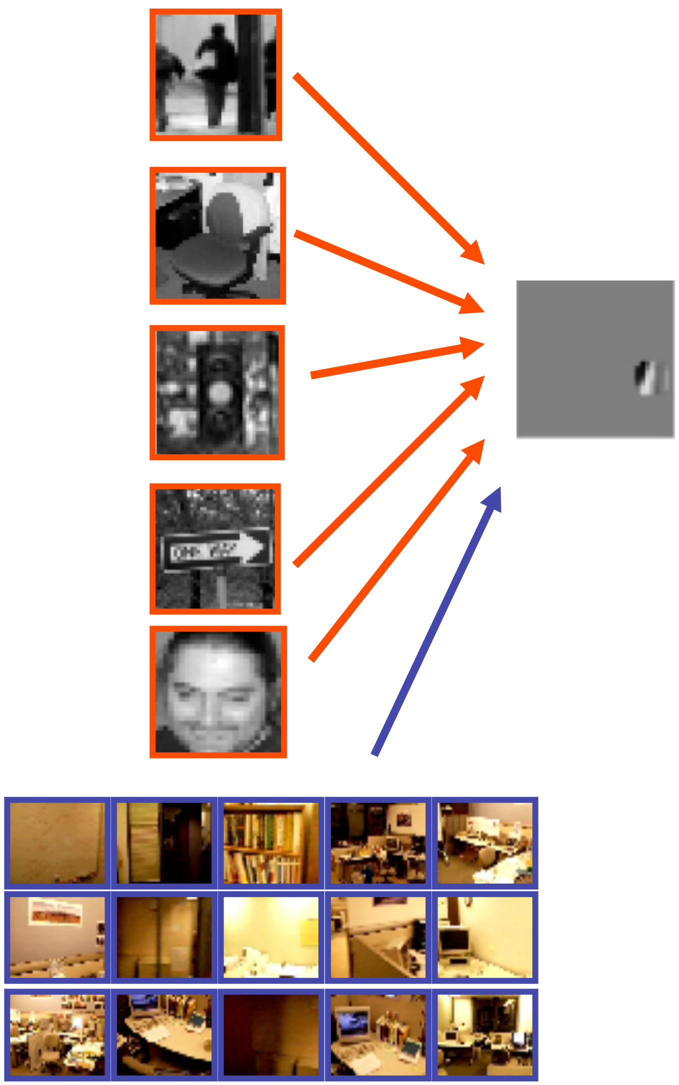
Torralba, Murphy, Freeman. CVPR 2004. PAMI 2007

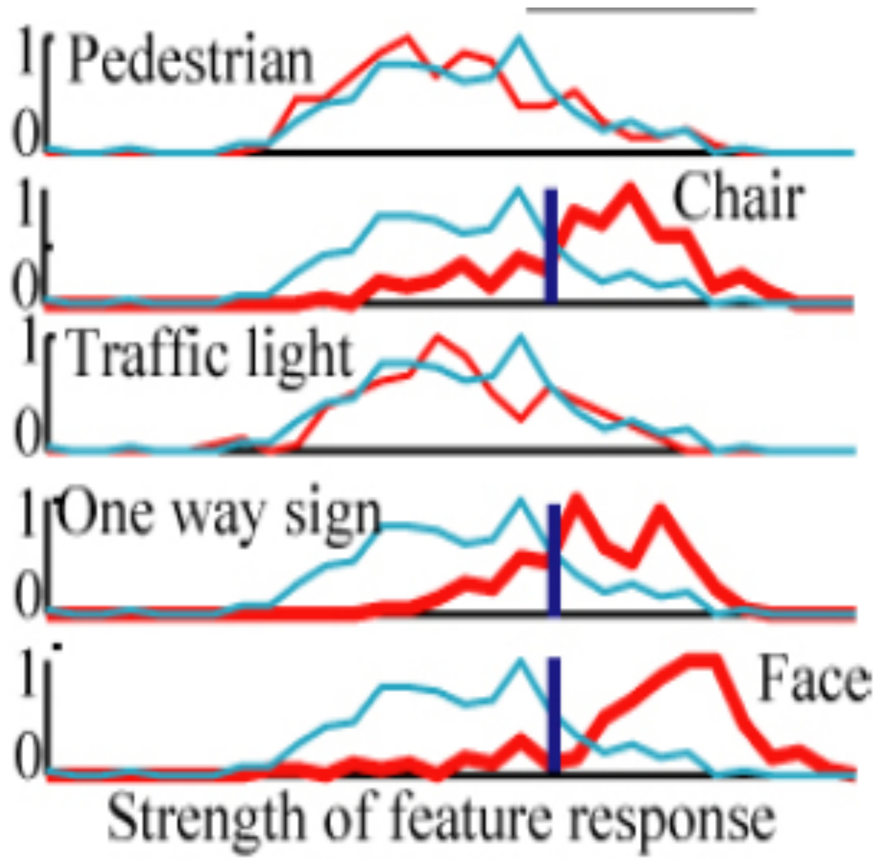
Specific feature



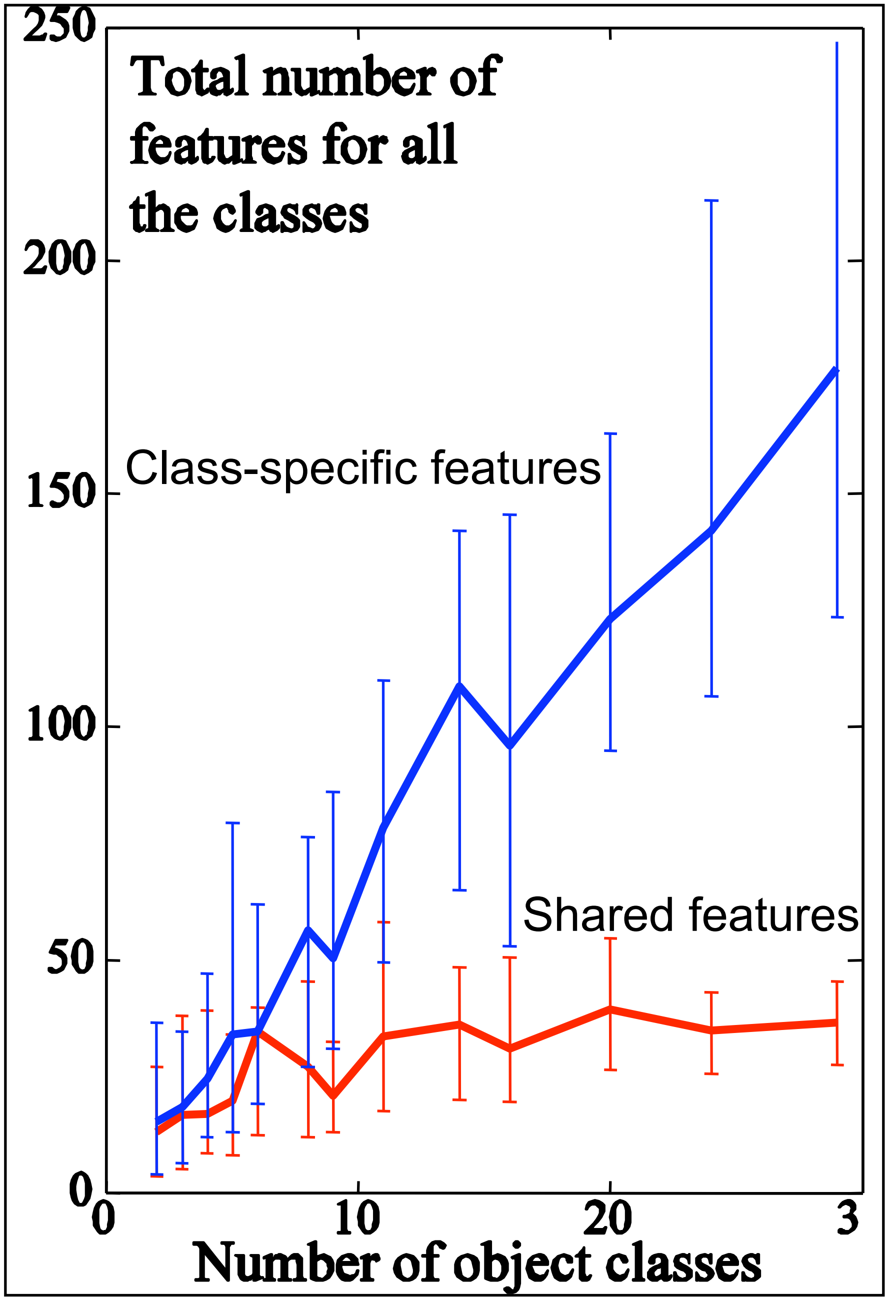
Non-shared feature: this feature is too specific to faces.

Shared feature





shared feature



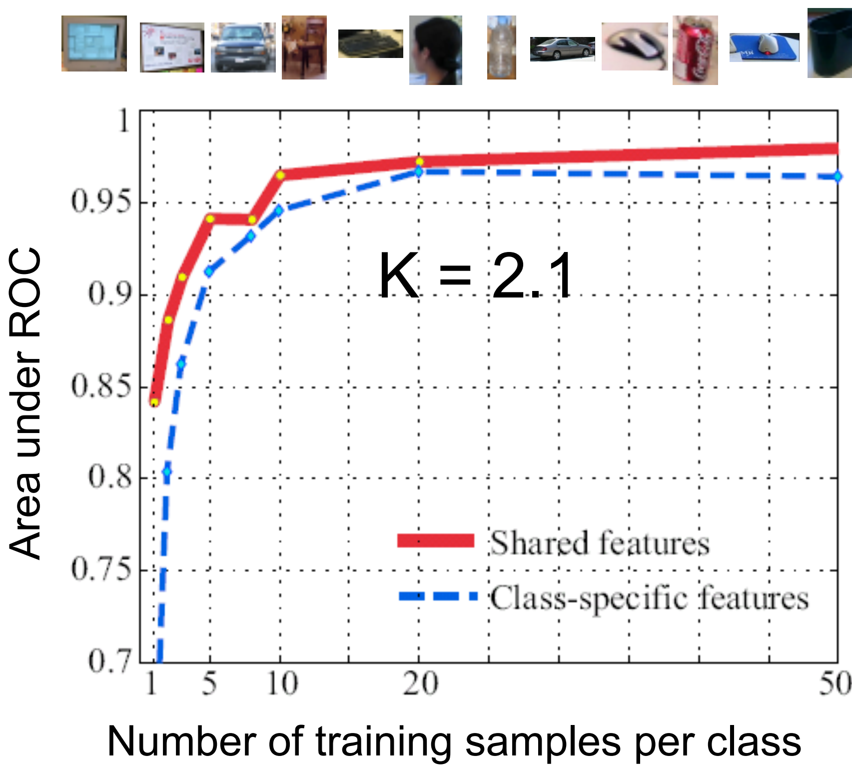
50 training samples/class
29 object classes
2000 entries in the dictionary

Results averaged on 20 runs
Error bars = 80% interval

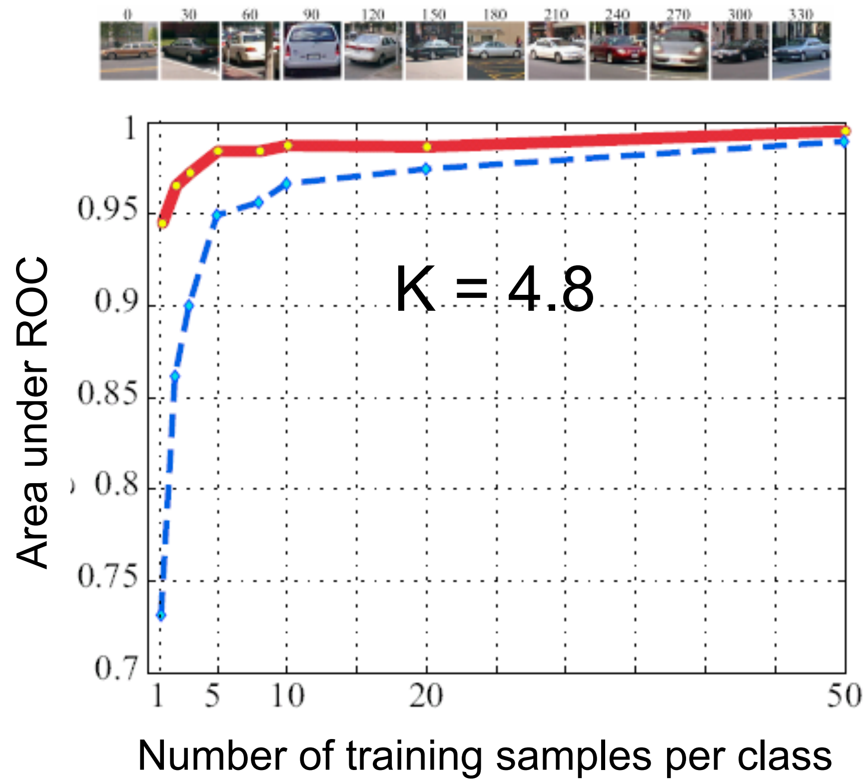
Torralba, Murphy, Freeman. CVPR 2004. PAMI 2007

Generalization as a function of object similarities

12 unrelated object classes

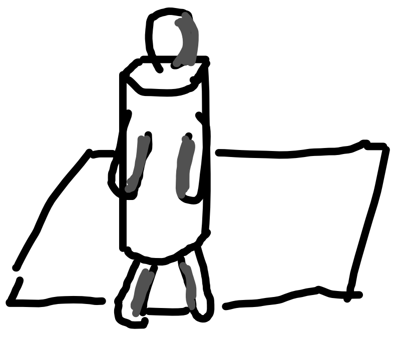


12 viewpoints

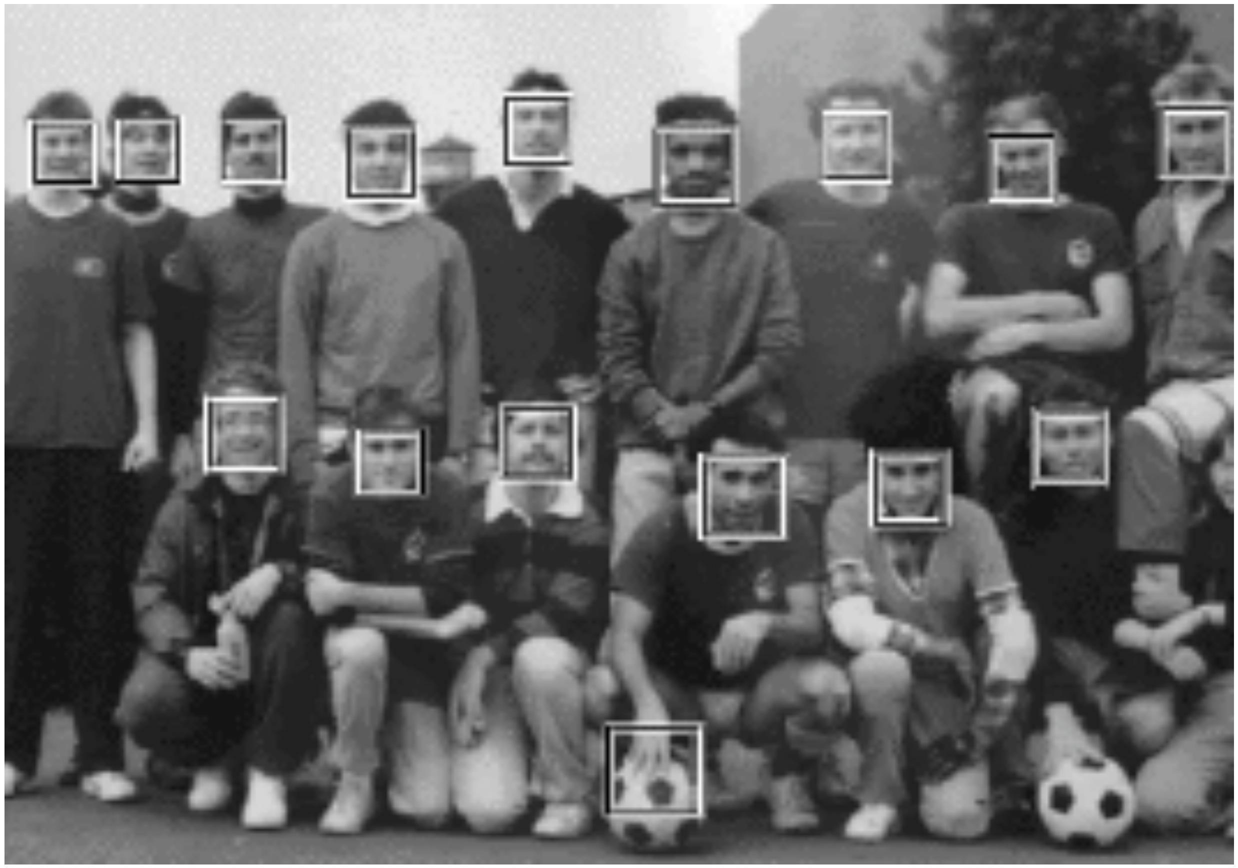


Torralba, Murphy, Freeman. CVPR 2004. PAMI 2007

3D object models



2D frontal face detection



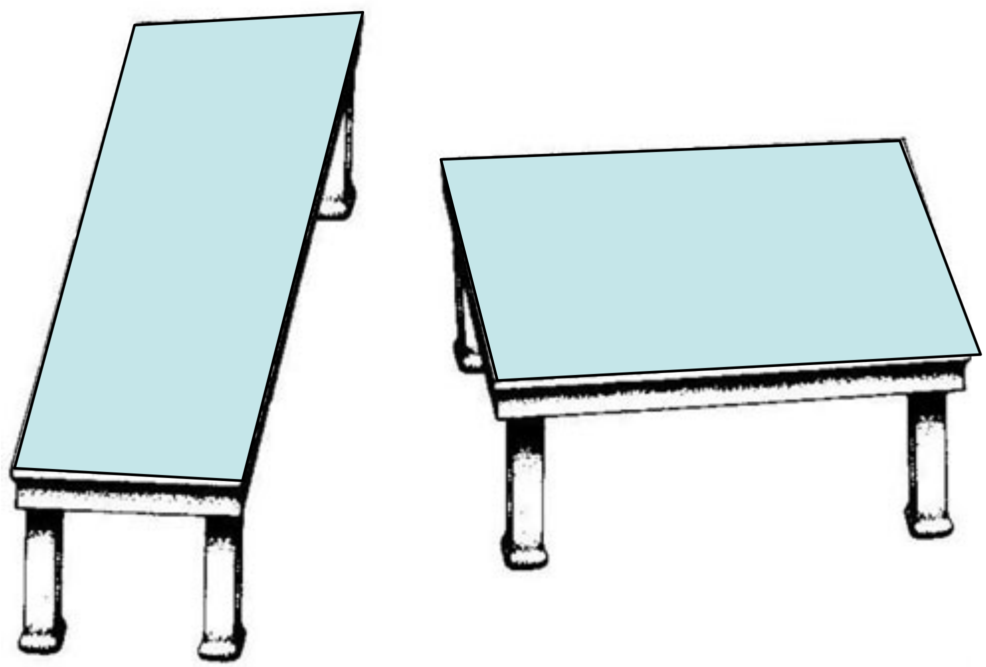
Amazing how far they have gotten with so little…

People have the bad taste of not being rotationally symmetric

Examples of un-collaborative subjects

Objects are not flat

3D drives perception of important object attributes



by Roger Shepard (”Turning the Tables”)

Depth processing is automatic, and we can not shut it down…

Class experiment

Class experiment

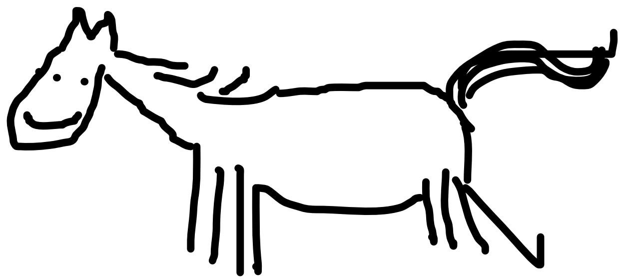
Experiment 1: draw a horse (the entire body, not just the head) in a white piece of paper.

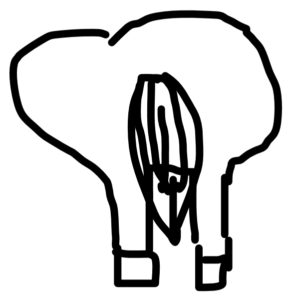
Do not look at your neighbor! You already know how a horse looks like… no need to cheat.

Class experiment

Experiment 2: draw a horse (the entire body, not just the head) but this time chose a viewpoint as weird as possible.

Anonymous participant



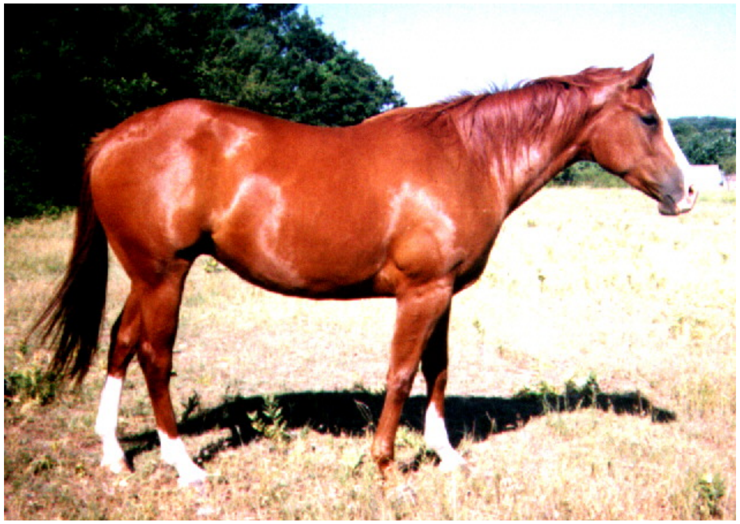


3D object categorization

Wait: object categorization in humans is **not** invariant to 3D pose

3D object categorization

Despite we can categorize all three pictures as being views of a horse, the three pictures do not look as being equally typical views of horses. And they do not seem to be recognizable with the same easiness.

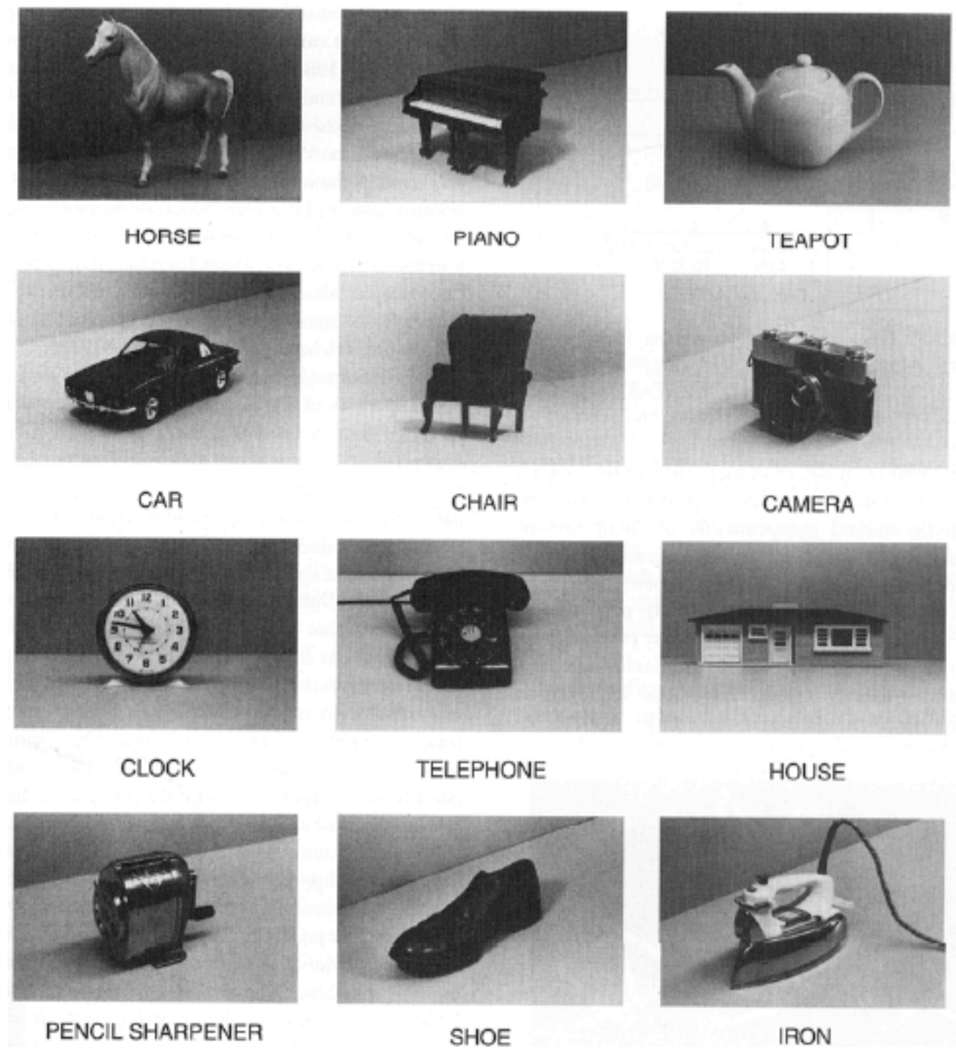


Canonical Perspective

Examples of canonical perspective:

In a recognition task, reaction time correlated with the ratings.

Canonical views are recognized faster at the entry level.

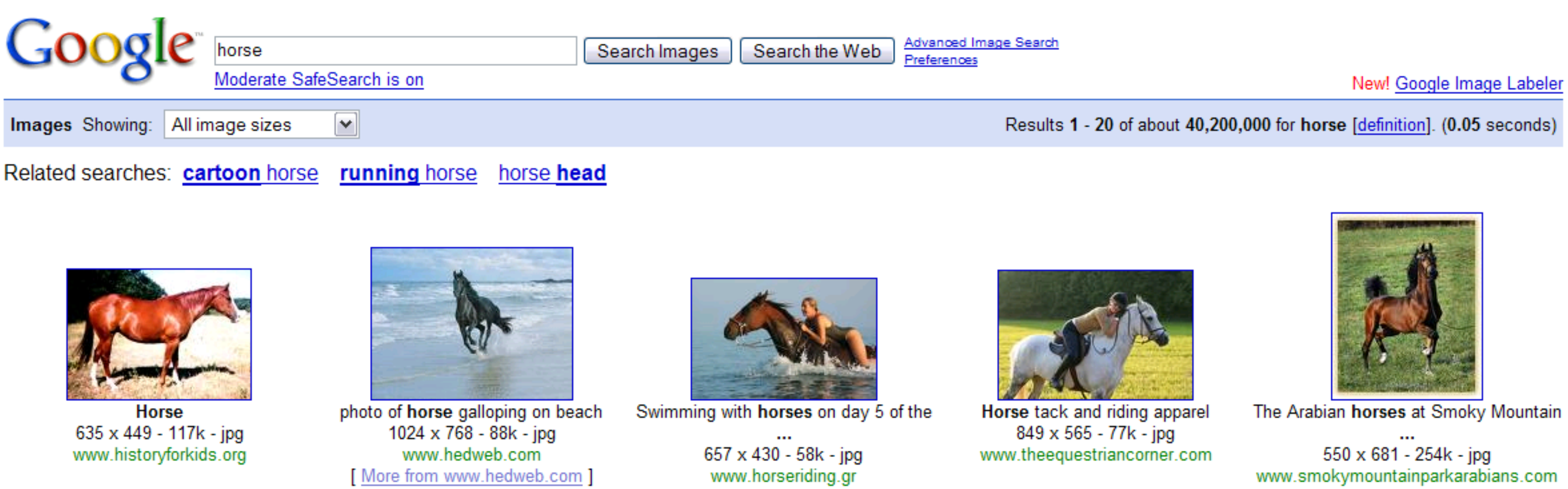


From *Vision Science*, Palmer

Canonical Viewpoint

Frequency hypothesis: easiness of recognition is related to the number of times we have see the objects from each viewpoint.

For a computer, using its Google memory, a horse looks like:

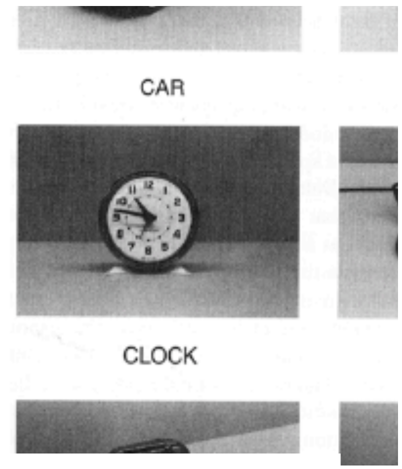


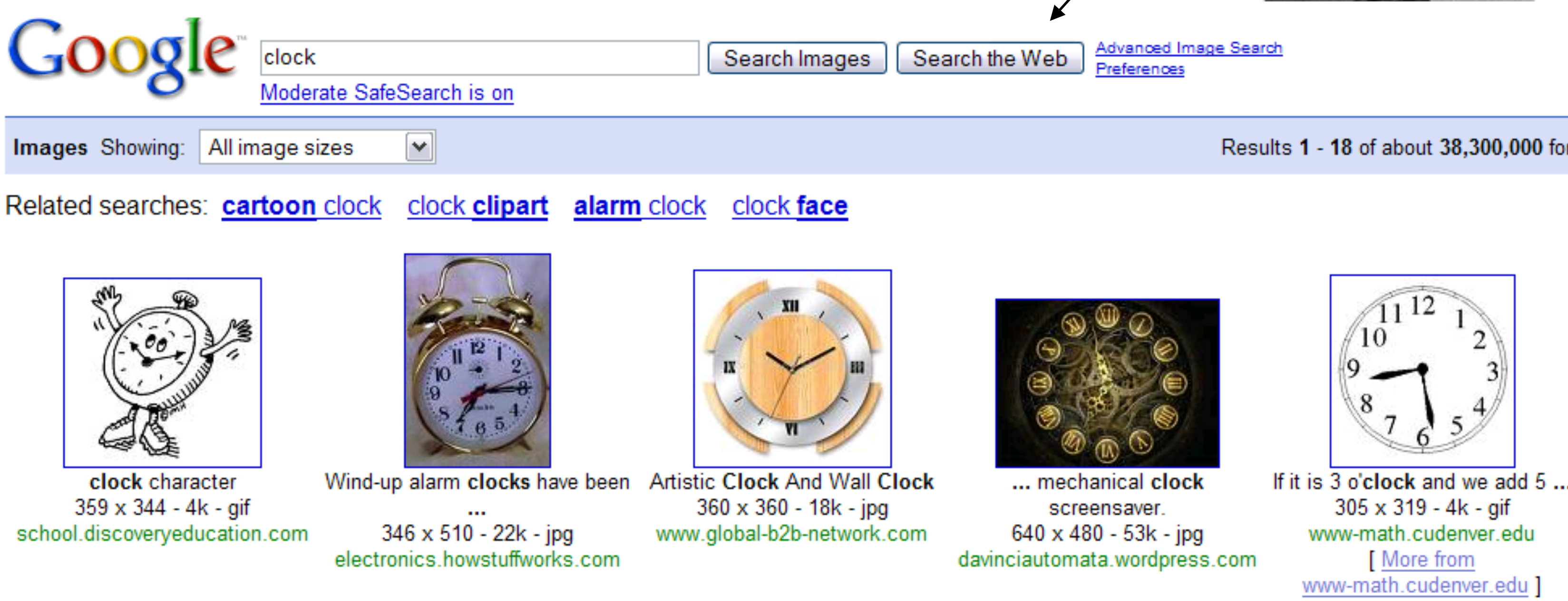
It is not a uniform sampling on viewpoints
(some artificial datasets might contain non natural statistics)

Canonical Viewpoint

Maximal information hypothesis:

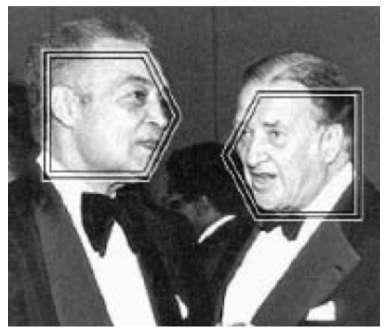
Clocks are preferred as purely frontal





Solution to deal with 3D variations: "do not deal with it"

"not"-Dealing with rotations and pose:



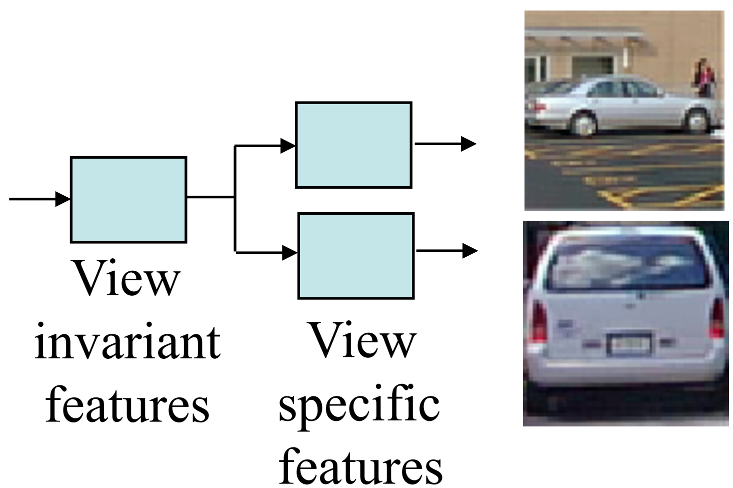
Train a different model for each view.

(b) For cars, classifiers are trained on 8 viewpoints

The combined detector is invariant to pose variations without an explicit 3D model.

Shared features for Multi-view object detection

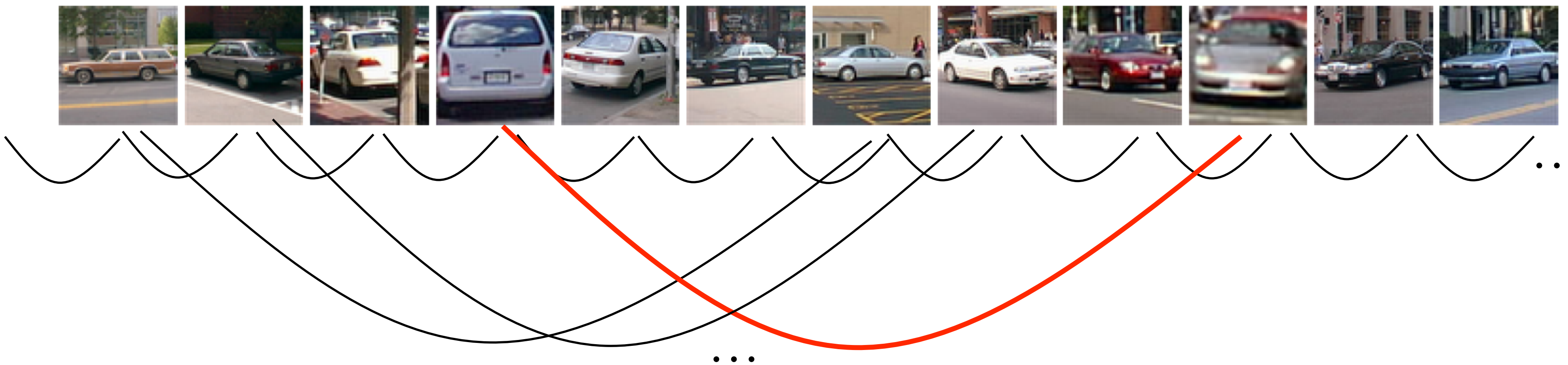
Training does not require having different views of the same object.



Torralba, Murphy, Freeman. PAMI 07

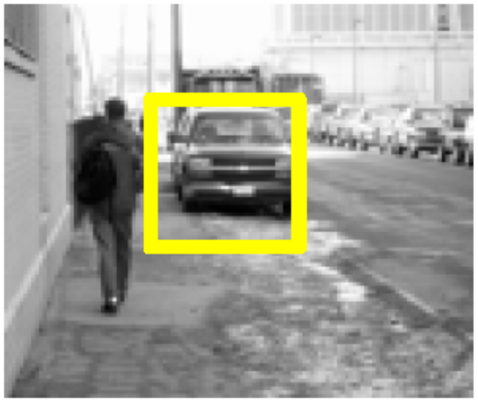
Shared features for Multi-view object detection

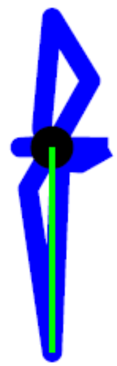
Sharing is not a tree. Depends also on 3D symmetries.

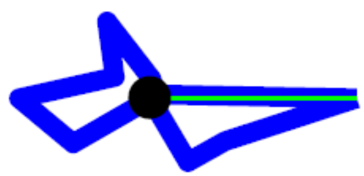


Torralba, Murphy, Freeman. PAMI 07

Multi-view object detection







Strong learner H response for car as function of assumed view angle

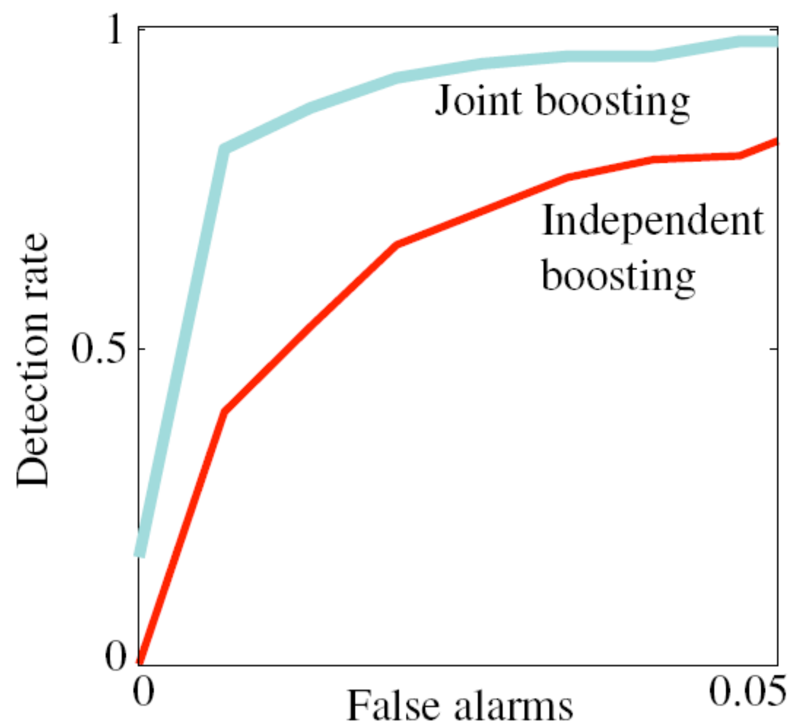


Fig. 19. ROC for view invariant car detection. The graph compares the ROC for the multiview classifier trained using joint boosting for 12 views and using independent boosting for each view. In both cases, the classifier is trained with 20 samples per view and only 70 features (stumps) are used.

Torralba, Murphy, Freeman. PAMI 07

Voting schemes

Towards Multi-View Object Class Detection
Alexander Thomas
Vittorio Ferrari
Bastian Leibe
Tinne Tuytelaars
Bernt Schiele
Luc Van Gool

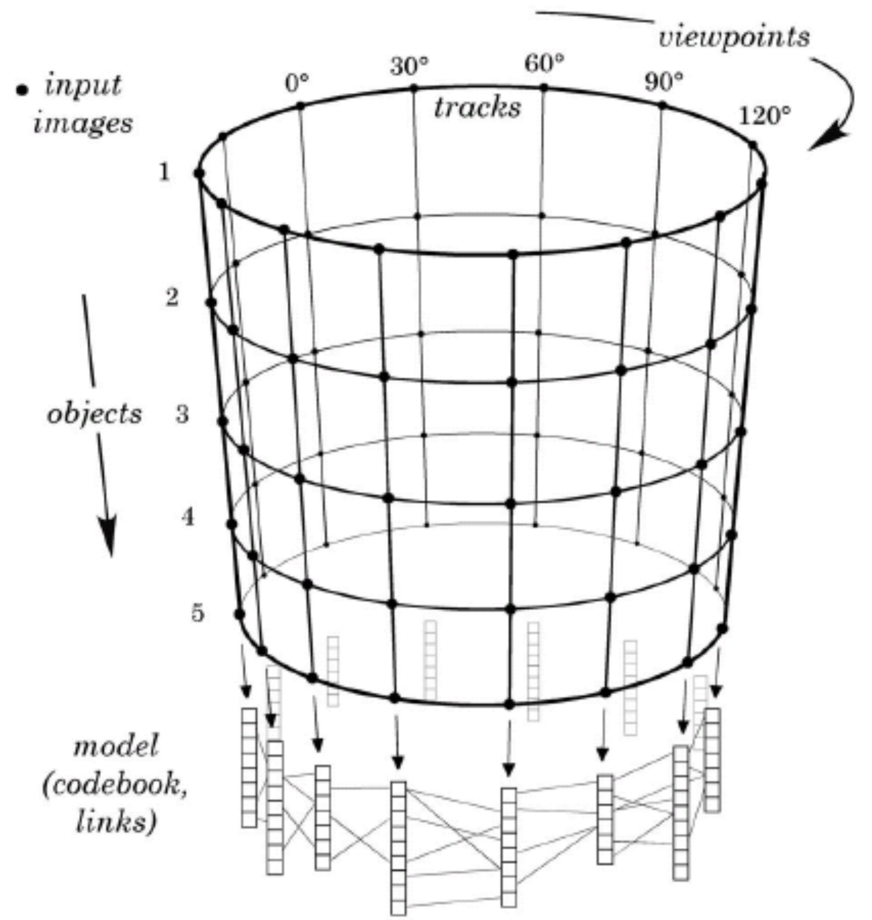
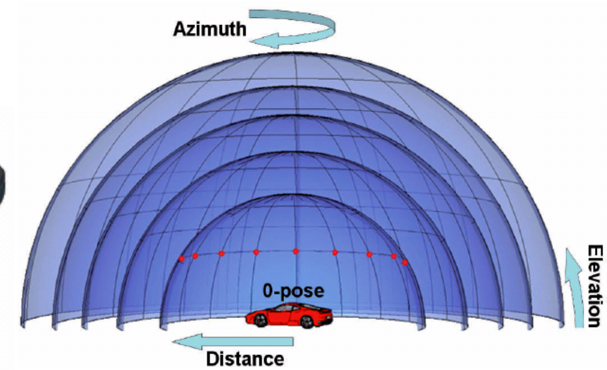


Figure 2. *Visual representation of our multi-view model. Only viewpoints lying on a circle around the object are shown. However, the proposed method supports the general case of viewpoints distributed over the whole viewing sphere.*

Viewpoint-Independent Object Class Detection using 3D Feature Maps

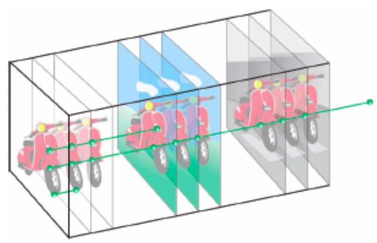
Training dataset: synthetic objects

Figure 1. Examples for 3D models of our two-class training database.



Discretization of the camera parameters azimuth, elevation and distance during training.

Features



Voting scheme and detection

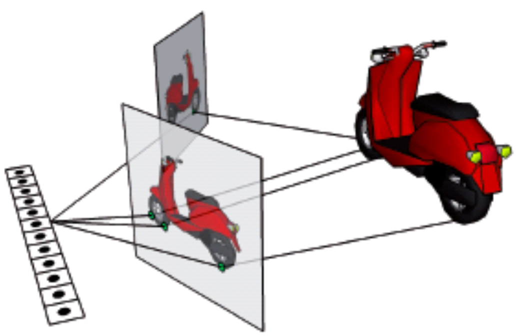
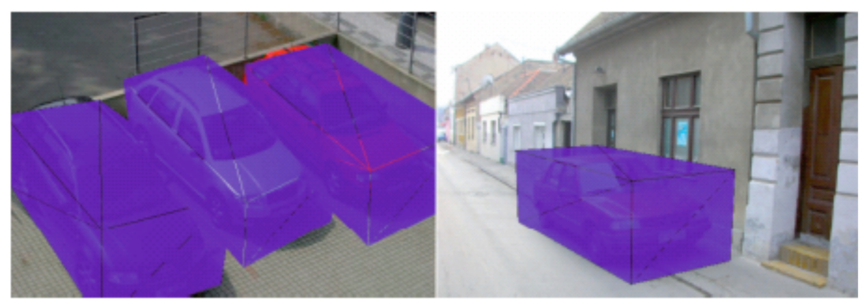


Figure 4. Each codebook entry stores the mean descriptor and the 3D positions of all the similar features which form a cluster.

Each cluster casts votes for the voting bins of the discrete poses contained in its internal list.



Liebelt, Schmid, Schertler. CVPR 2008

Stages of processing

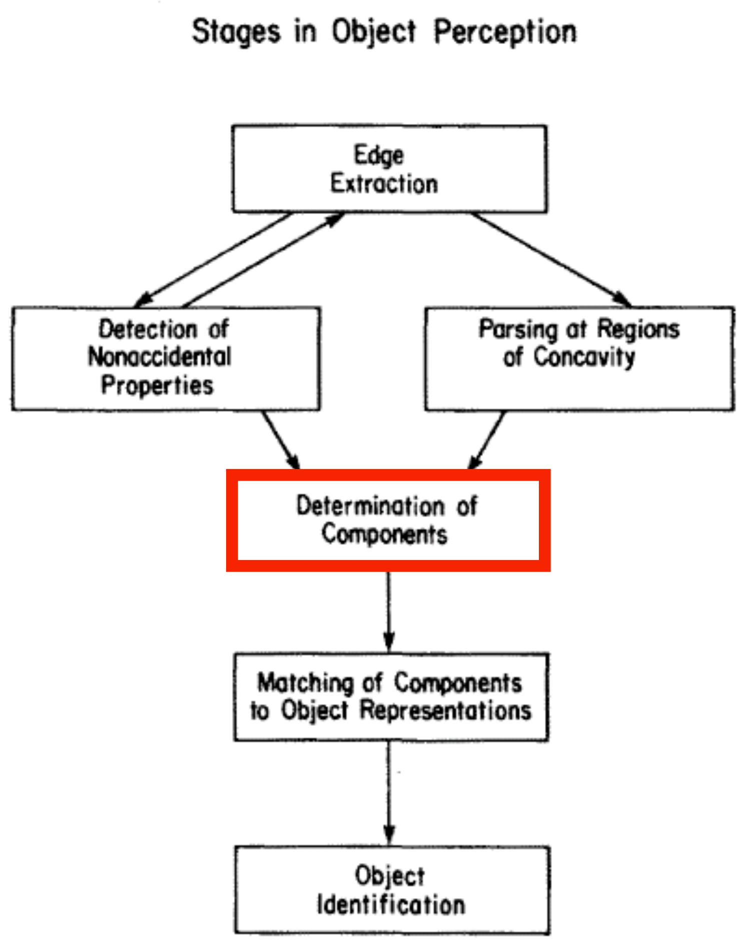


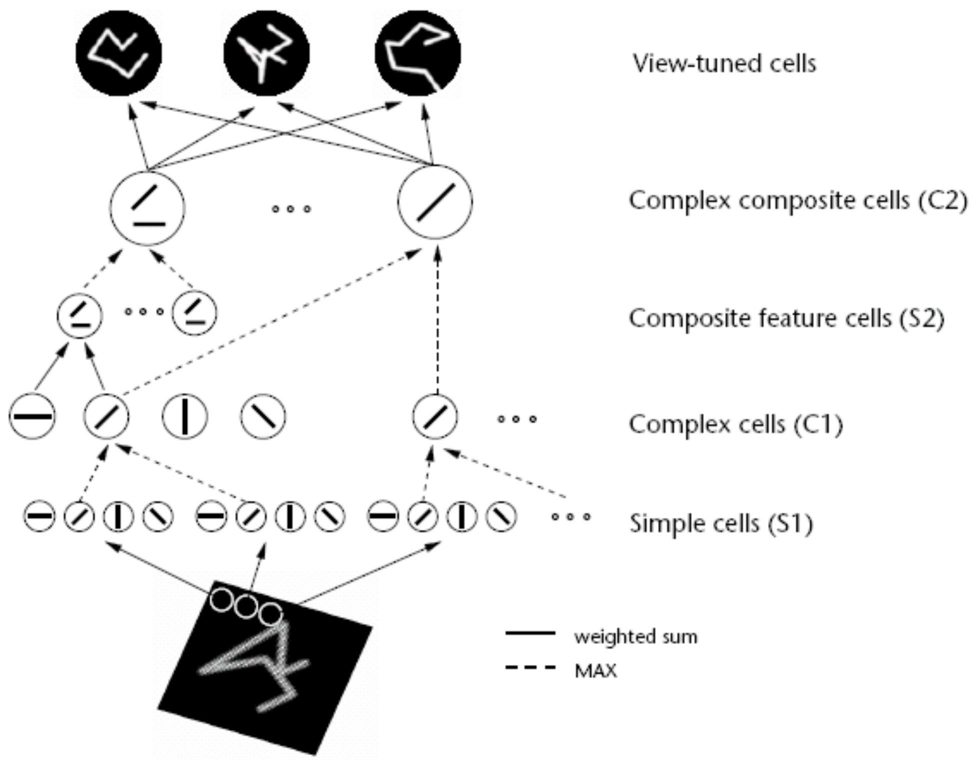
Figure 2. Presumed processing stages in object recognition.

"Parsing is performed, primarily at concave regions, simultaneously with a detection of nonaccidental properties."

Models of object recognition

I. Biederman, “Recognition-by-components: A theory of human image understanding,” *Psychological Review*, 1987.

M. Riesenhuber and T. Poggio, “Hierarchical models of object recognition in cortex,” *Nature Neuroscience* 1999.

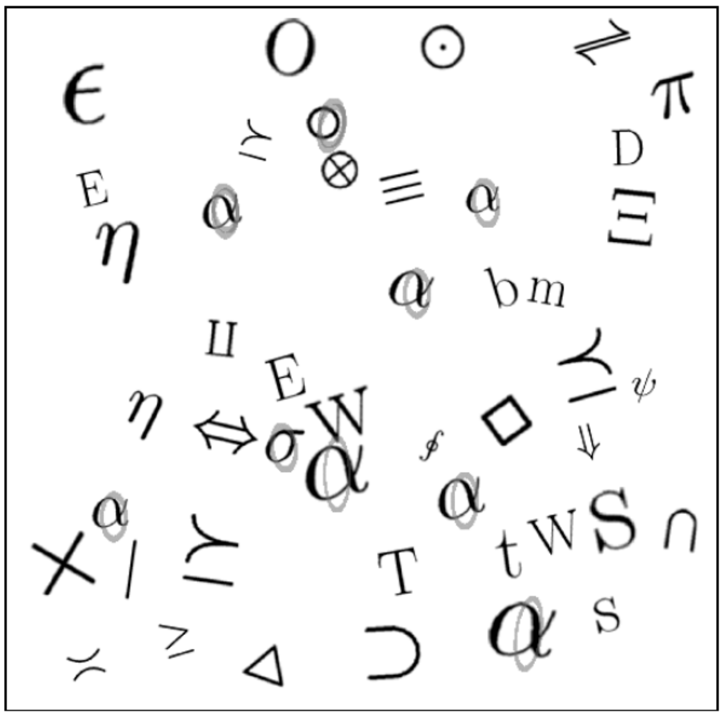


T. Serre, L. Wolf and T. Poggio. “Object recognition with features inspired by visual cortex”. CVPR 2005

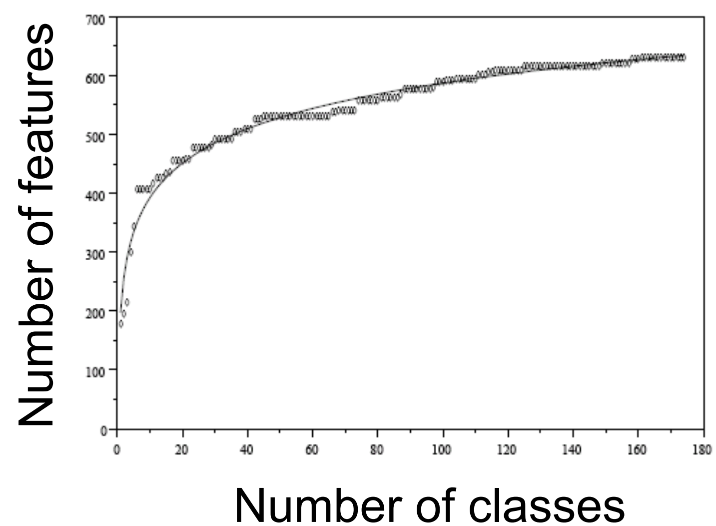
Reusable Parts

Krempp, Geman, & Amit "Sequential Learning of Reusable Parts for Object Detection". TR 2002

Goal: Look for a vocabulary of edges that reduces the number of features.



Examples of reused parts

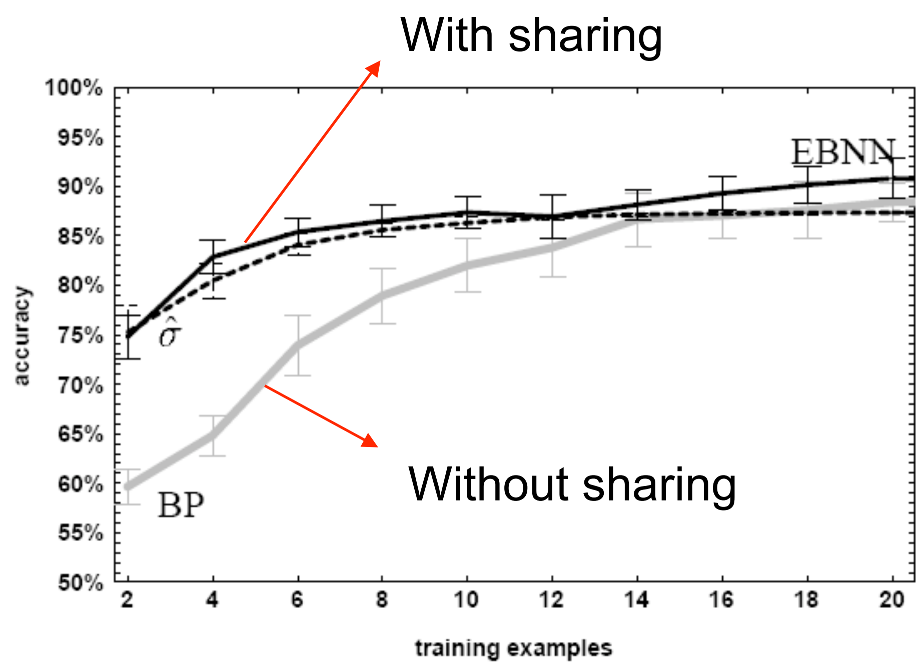


Sharing invariances

S. Thrun. Is Learning the n-th Thing Any Easier Than Learning The First? NIPS 1996

Knowledge is transferred between tasks via a learned model of the invariances of the domain: object recognition is invariant to rotation, translation, scaling, lighting, … These invariances are common to all object recognition tasks.

Toy world

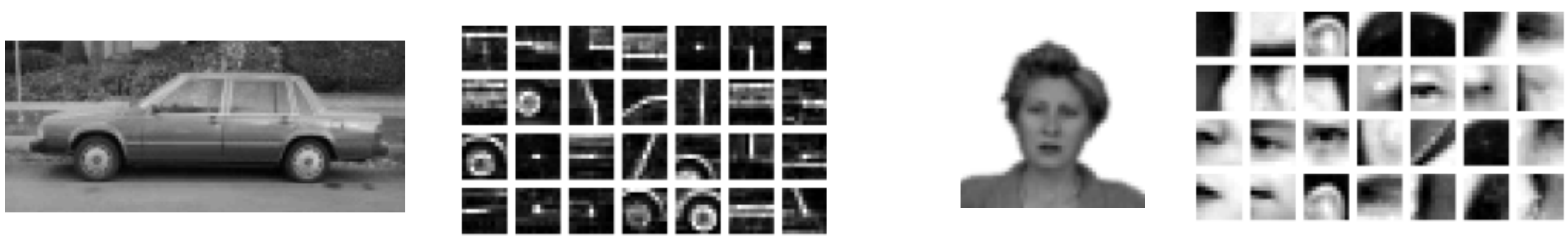


Some symptoms of one-vs-all multiclass approaches

Part-based object representation (looking for meaningful parts):

- A. Agarwal and D. Roth

- M. Weber, M. Welling and P. Perona



…

These studies try to recover parts that are meaningful. But is this the right thing to do? The derived parts may be too specific, and they are not likely to be useful in a general system.

Sharing patches

- Bart and Ullman, 2004

For a new class, use only features similar to features that where good for other classes:

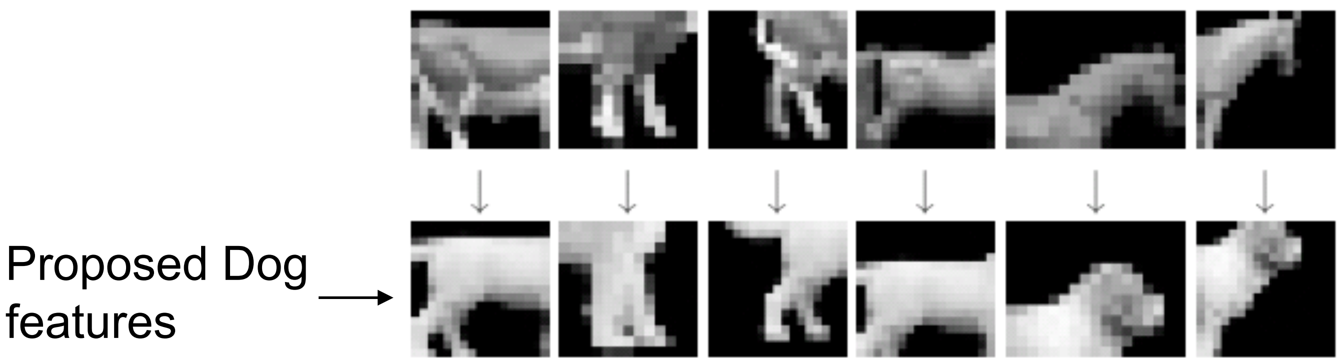
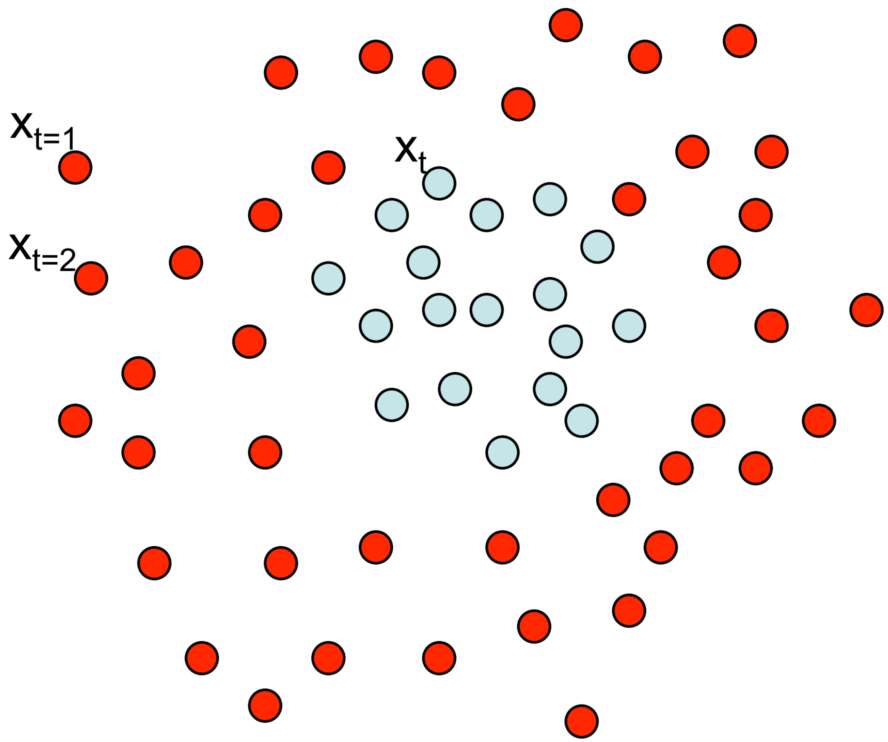


Figure 1. Feature adaptation. (a) Top row: features extracted from multiple images of cows (first three) and horses (last three), as described in section 3.1. Bottom row: features adapted to the dogs class by the proposed cross-generalization algorithm (section 3.2), using a single dog image.

Boosting

- It is a sequential procedure:



Each data point has

a class label:

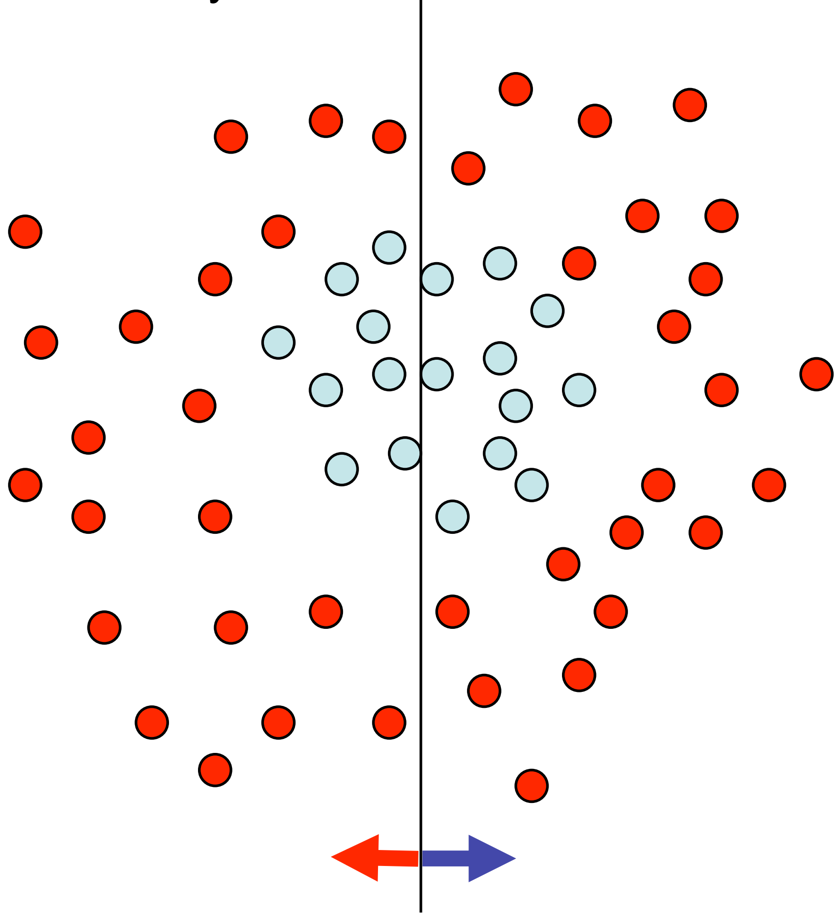
$$y_t = \begin{cases} +1 & (\bullet) \\ -1 & (\circ) \end{cases}$$

and a weight:

$w_t = 1$

Toy example

Weak learners from the family of lines



Each data point has

a class label:

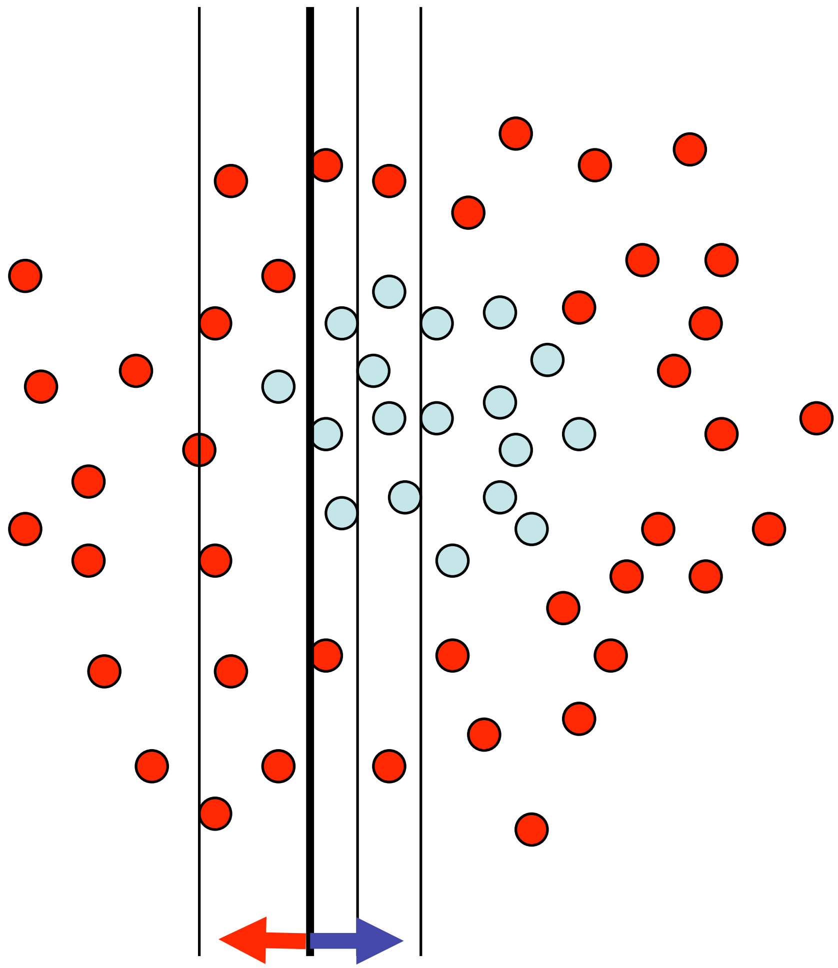
$$y_t = \begin{cases} +1 \; (\bullet) \\ -1 \; (\circ) \end{cases}$$

and a weight:

$w_t = 1$

h => p(error) = 0.5 it is at chance

Toy example



Each data point has

a class label:

$$y_t = \begin{cases} +1 \ (\bullet) \\ -1 \ (\circ) \end{cases}$$

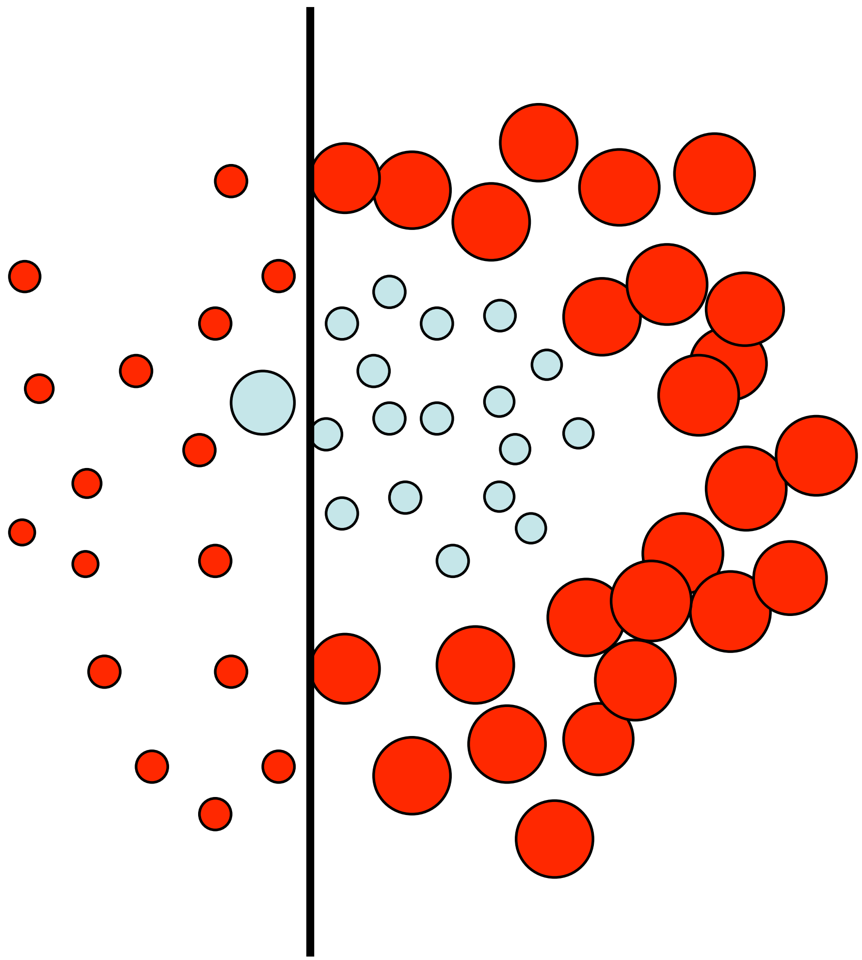
and a weight:

$w_t = 1$

This one seems to be the best

This is a ‘**weak classifier**’: It performs slightly better than chance.

Toy example



Each data point has

a class label:

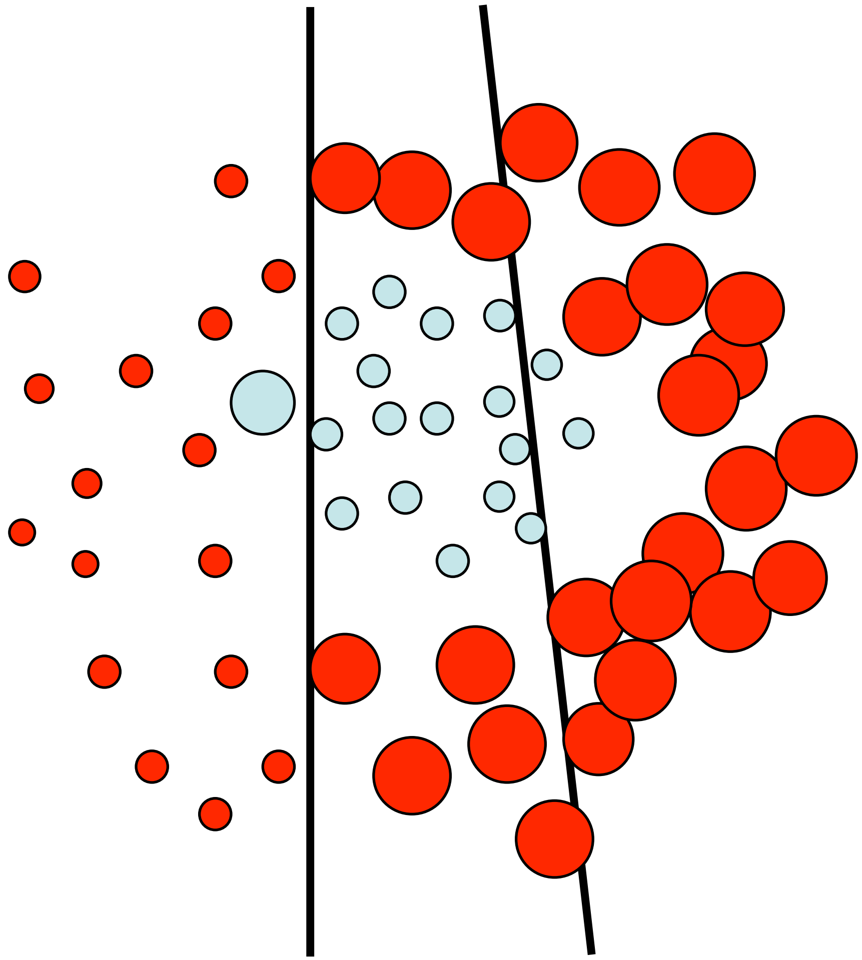
$$y_t = \begin{cases} +1 \\ -1 \end{cases}$$

We update the weights:

$$w_t \leftarrow w_t \exp\{-y_t H_t\}$$

We set a new problem for which the previous weak classifier performs at chance again

Toy example



Each data point has

a class label:

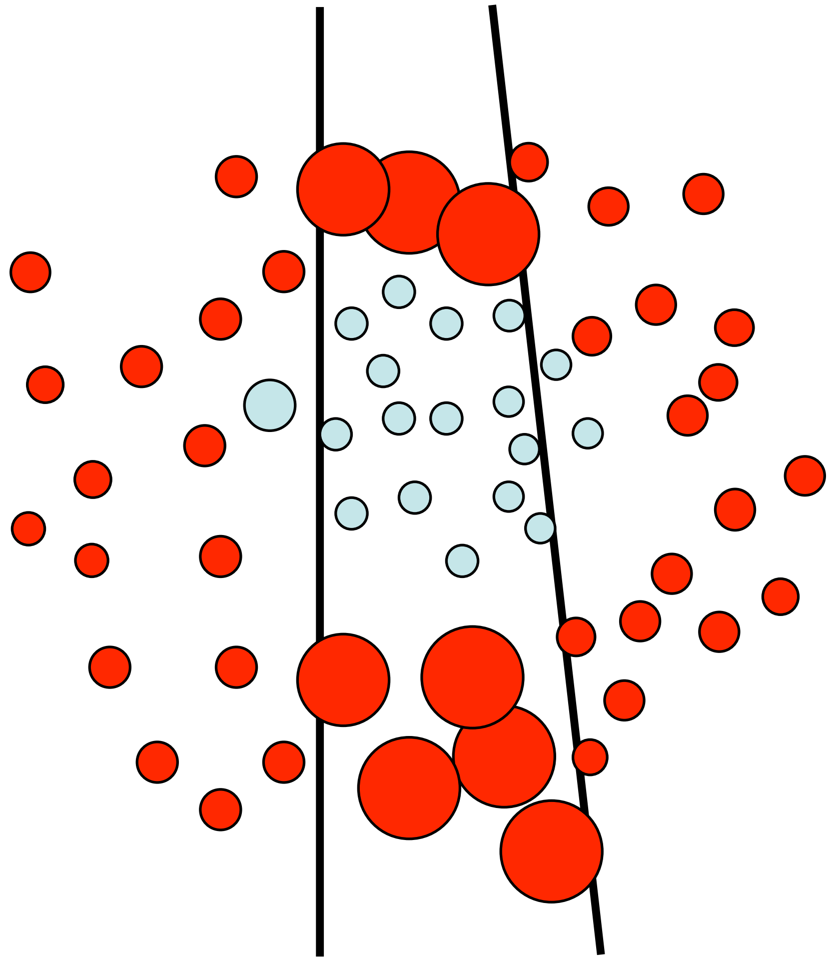
$$y_t = \begin{cases} +1 \; (\;) \\ -1 \; (\;) \end{cases}$$

We update the weights:

$$w_t \leftarrow w_t \exp\{-y_t H_t\}$$

We set a new problem for which the previous weak classifier performs at chance again

Toy example



Each data point has

a class label:

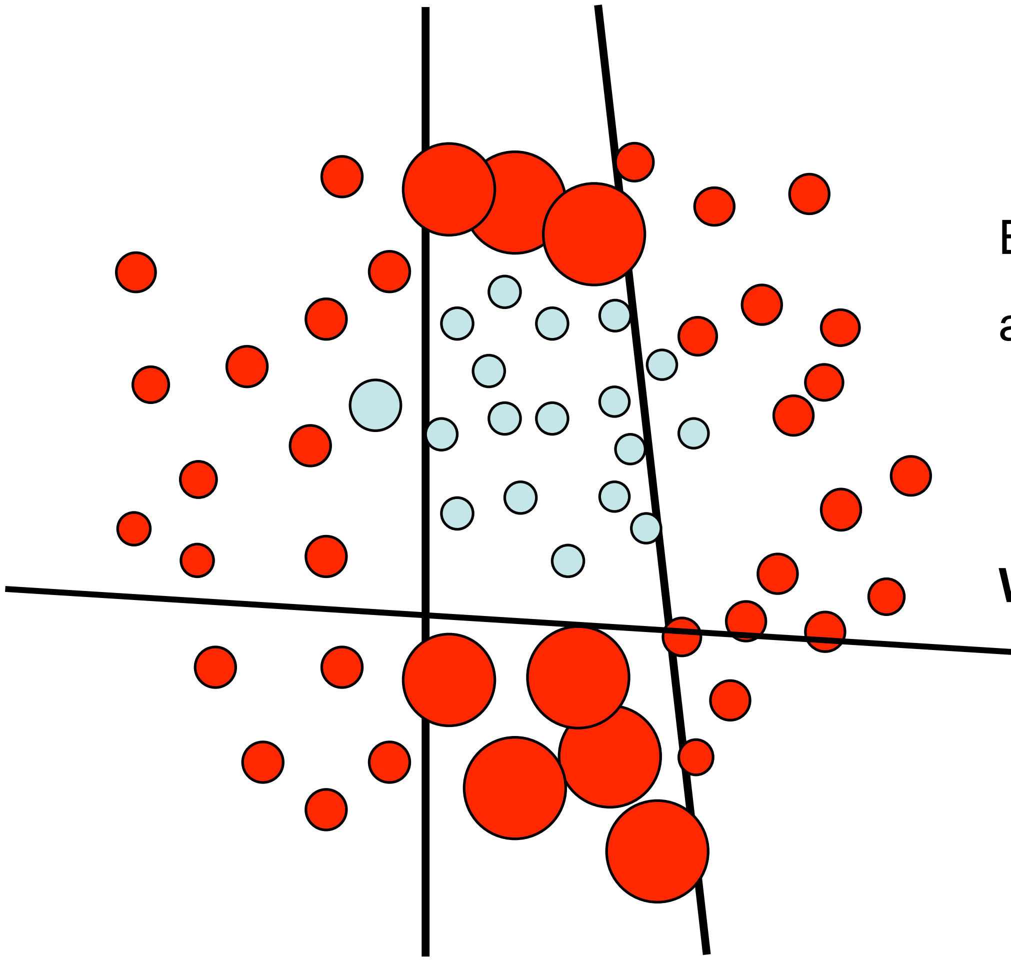
$$y_t = \begin{cases} +1 \ (\bullet) \\ -1 \ (\circ) \end{cases}$$

We update the weights:

$$w_t \leftarrow w_t \exp\{-y_t H_t\}$$

We set a new problem for which the previous weak classifier performs at chance again

Toy example



Each data point has

a class label:

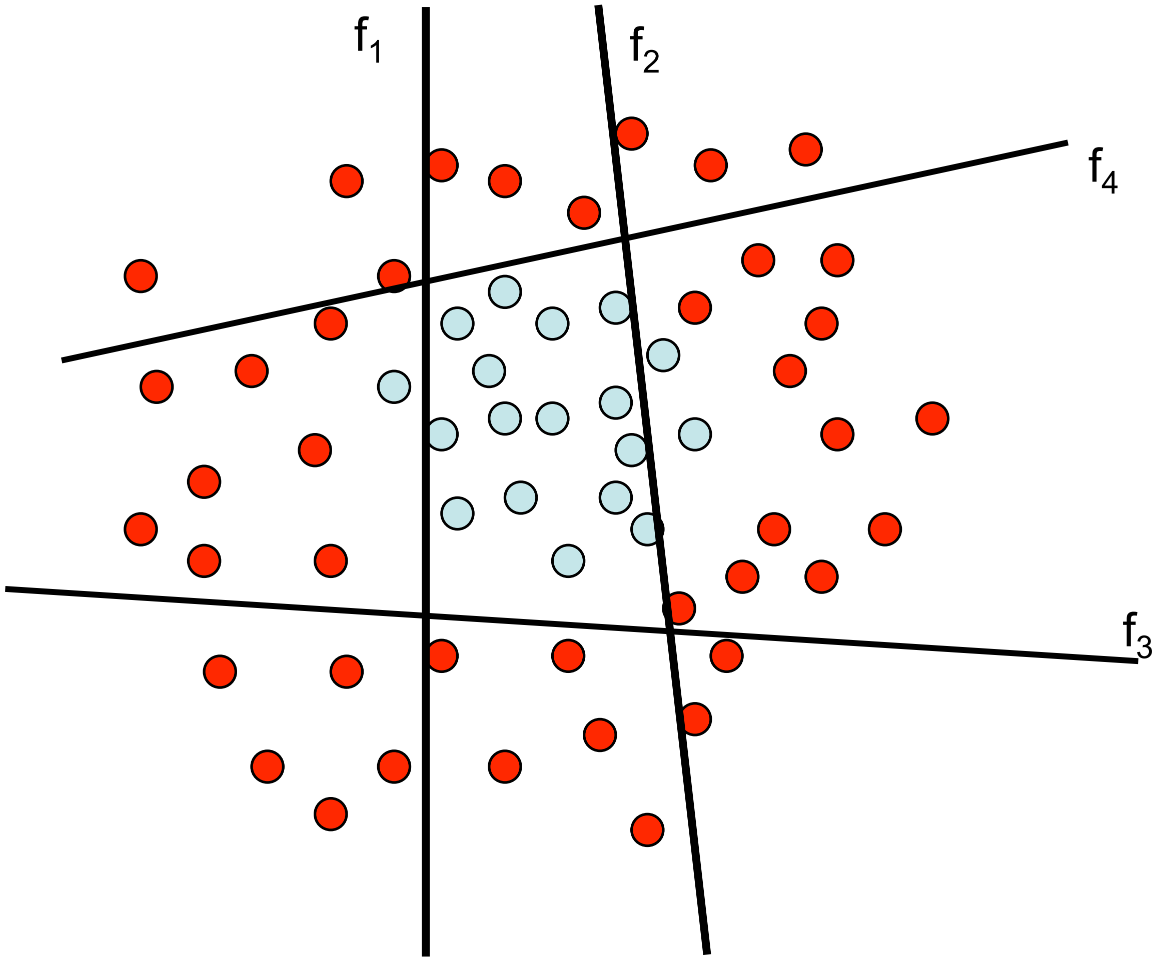
$$y_t = \begin{cases} +1 & (\bullet) \\ -1 & (\circ) \end{cases}$$

We update the weights:

$w_t \leftarrow w_t \exp\{-y_t H_t\}$

We set a new problem for which the previous weak classifier performs at chance again

Toy example



The strong (non- linear) classifier is built as the combination of all the weak (linear) classifiers.