



Visual Recognition Overview

EECS 598-08 Fall 2014

Foundations of Computer Vision

Instructor: Jason Corso (jjcorso)

web.eecs.umich.edu/~jjcorso/t/598F14

Readings: FP 15.1, 18.1, SZ 14

Date: 11/12/14

Plan

- Introduction to visual recognition
- Take home: recognition of single 3D objects



Credit S Savarese.

Classification:

Does this image contain a building? [yes/no]



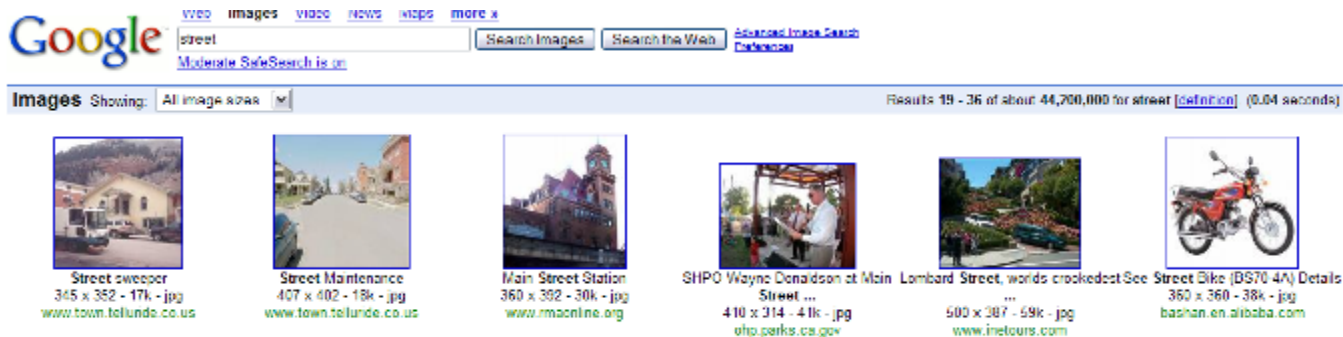
Yes!

Classification:

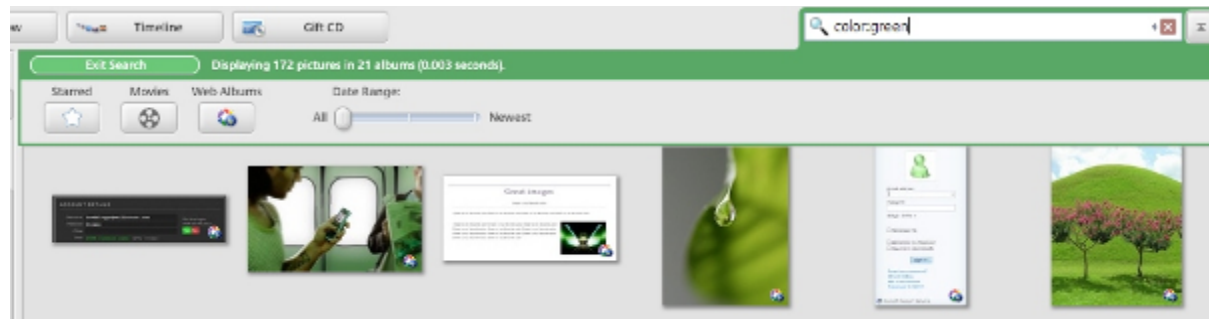
Is this a beach?



Application: Image Search



Organizing photo collections



Detection:

Does this image contain a car? [where?]



car

Detection:

Which object does this image contain? [where?]



Building



clock



person



car

Applications of Detection



Assistive technologies



Surveillance



Security



Assistive driving

Applications of computer vision



- Detecting faces
- Computational photography



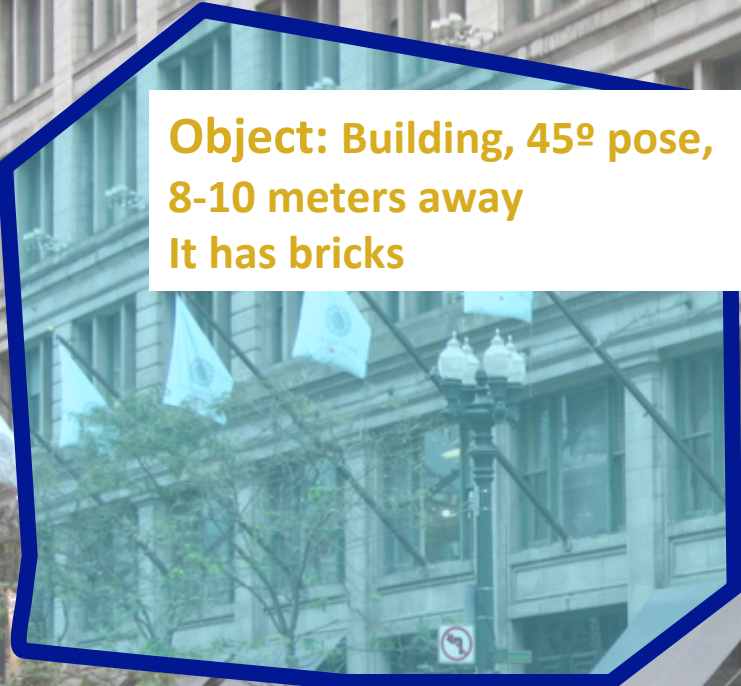
[Face priority AE] When a bright part of the face is too bright

Semantic Segmentation:

Accurate localization and recognition jointly



Semantic Segmentation: Estimating object semantic & geometric attributes



**Object: Building, 45° pose,
8-10 meters away
It has bricks**



**Object: Person, back;
1-2 meters away**



Object: Police car, side view, 4-5 m away

Categorization vs Single instance recognition

Which building is this? *Marshall Field* building in Chicago



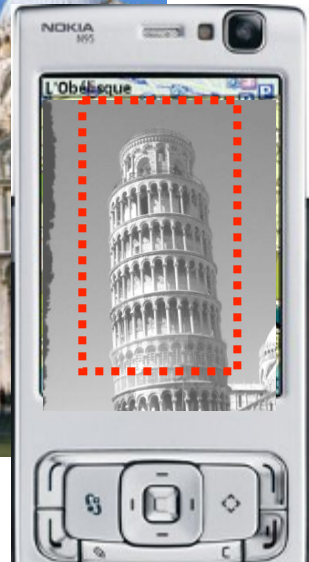
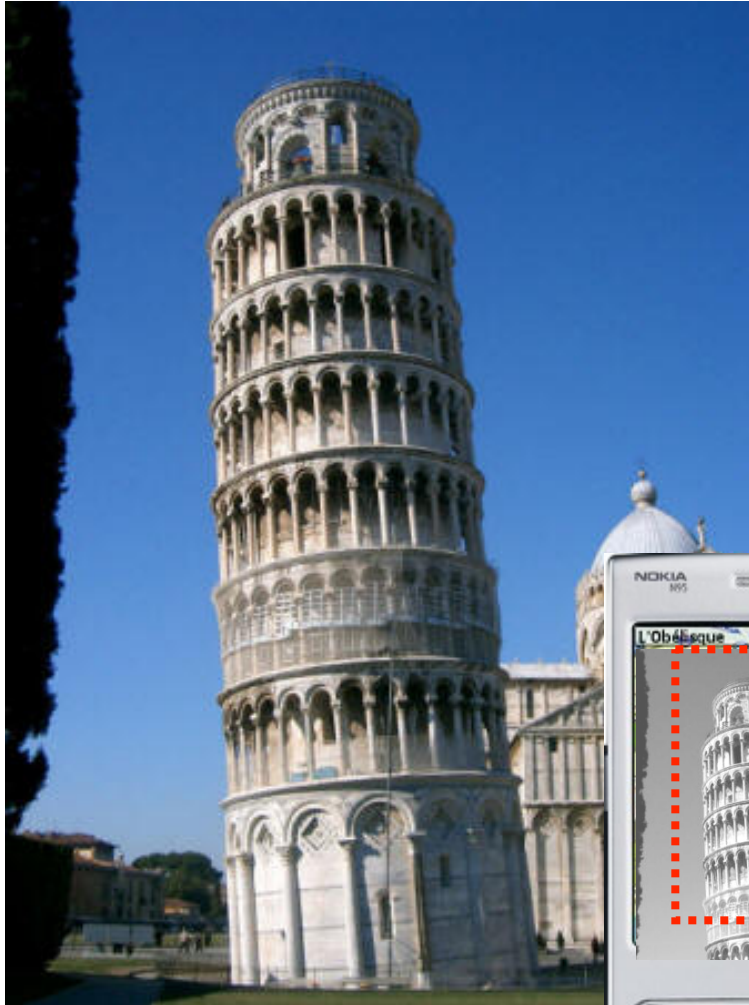
Categorization vs Single instance recognition

Where is the crunchy nut?



Applications of computer vision

- Recognizing landmarks in mobile platforms



+ GPS

LookTel – Real-Time Object Recognition on a Mobile Device



<http://www.youtube.com/watch?v=EXkSHh9GRbo>

- Robotics
- Navigation
- Interaction
- Manipulation

Object: car
Object: door; almost frontal view



Activity or Event recognition

What are these people doing?



Visual Recognition

- Design algorithms that are capable to
 - Classify images or videos
 - Detect and localize objects
 - Estimate semantic and geometrical attributes
 - Classify human activities and events

Why is this challenging?

How many object categories are there?

~10,000 to 30,000



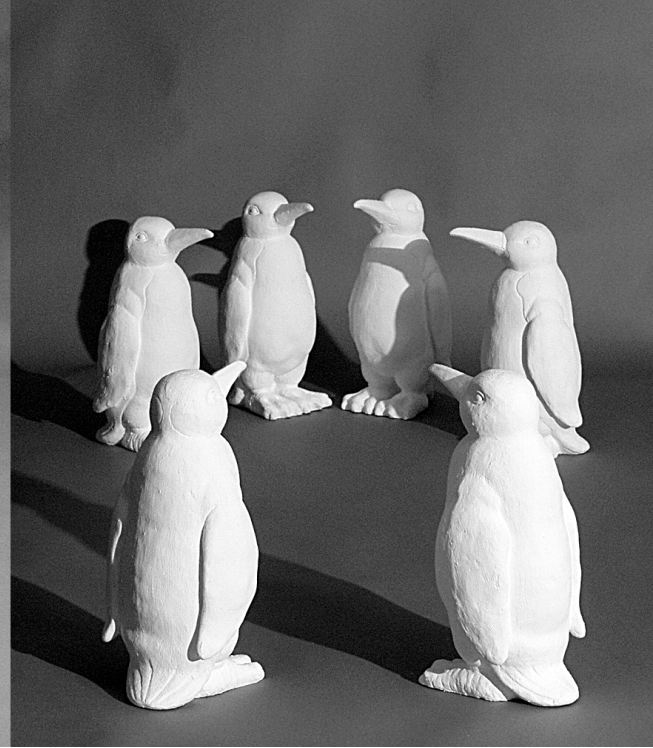
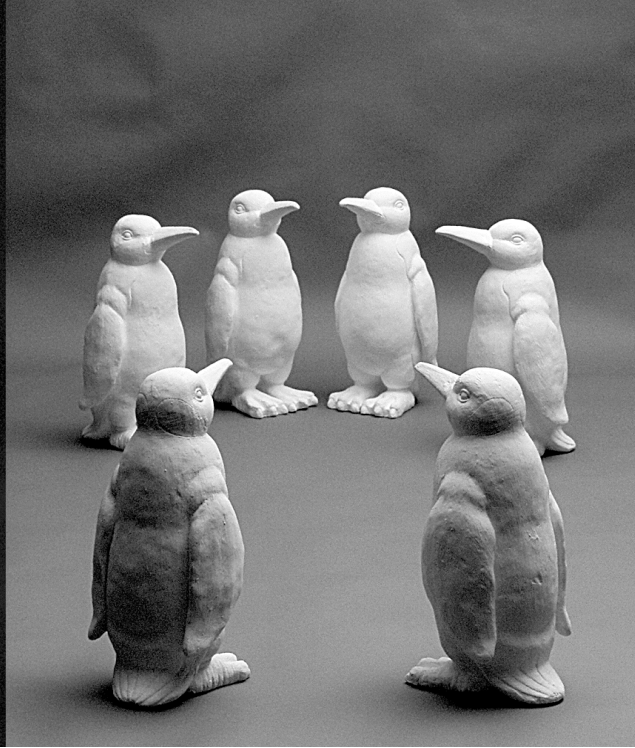
Challenges: viewpoint variation



Michelangelo 1475-1564

slide credit: Fei-Fei, Fergus & Torralba

Challenges: illumination



Challenges: scale



Challenges: deformation



Challenges: occlusion



Challenges: background clutter



Kilmeny Niland. 1995

Challenges: intra-class variation



Some early works on object categorization



- Turk and Pentland, 1991
- Belhumeur, Hespanha, & Kriegman, 1997
- Schneiderman & Kanade 2004
- Viola and Jones, 2000



- Amit and Geman, 1999
- LeCun et al. 1998
- Belongie and Malik, 2002



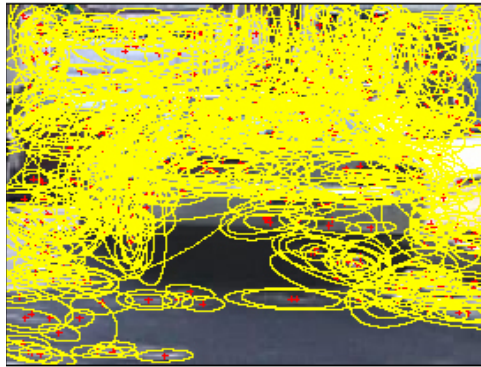
- Schneiderman & Kanade, 2004
- Argawal and Roth, 2002
- Poggio et al. 1993

Basic Problems in Object Recognition

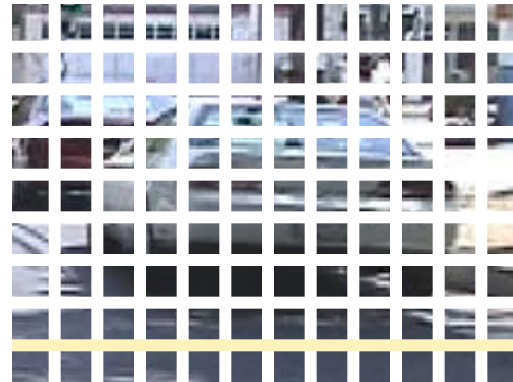
- Representation
 - How to represent an object category; which classification scheme?
- Learning
 - How to learn the classifier, given training data
- Recognition
 - How the classifier is to be used on novel data

Representation

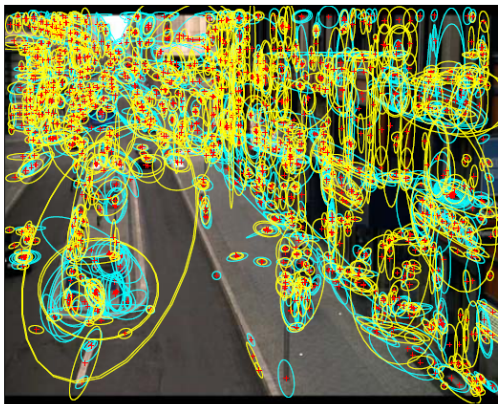
- Building blocks: Sampling strategies



Interest operators



Dense, uniformly



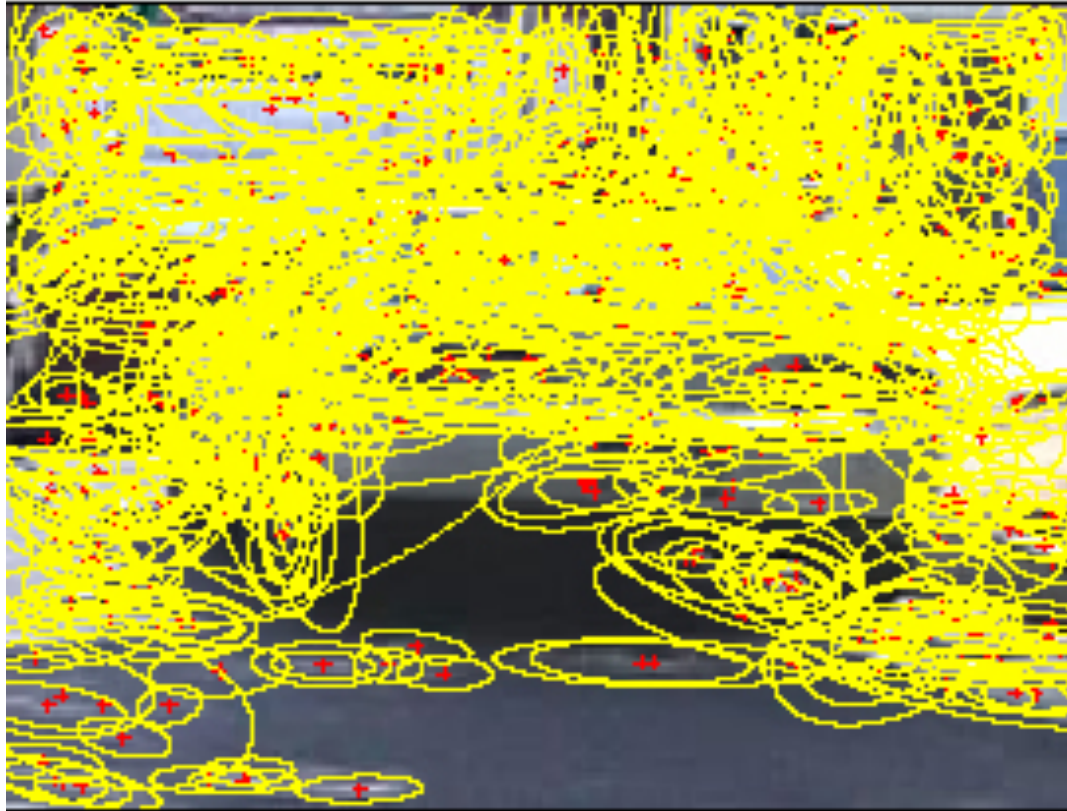
Multiple interest operators



Randomly

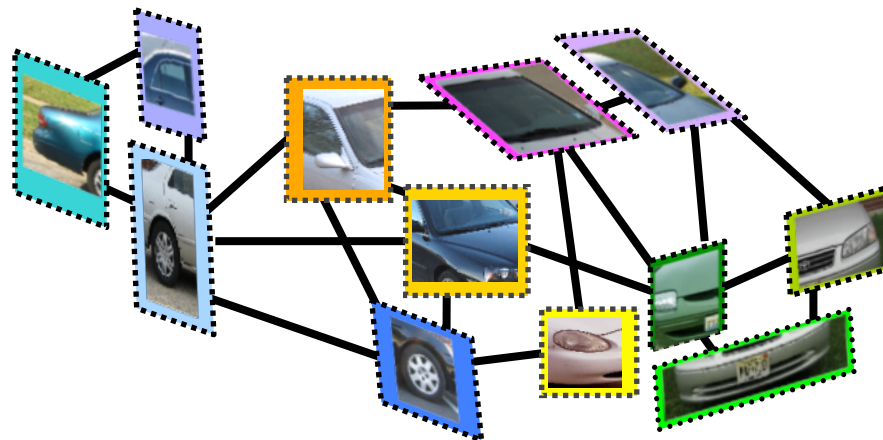
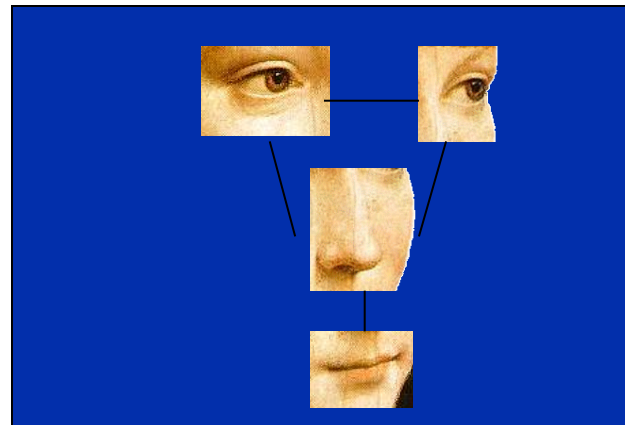
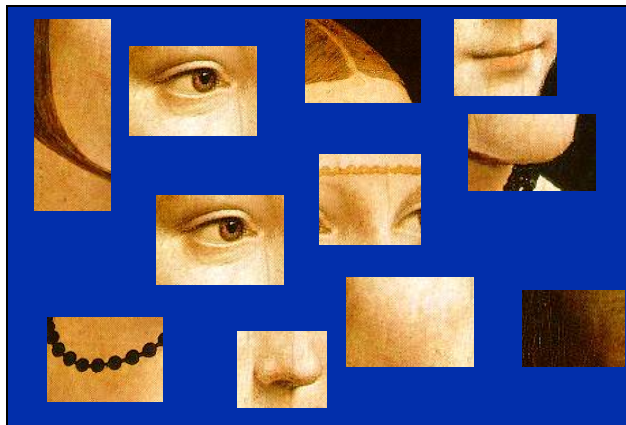
Representation

- Building blocks: Choice of descriptors [SIFT, HOG, codewords....]



Representation

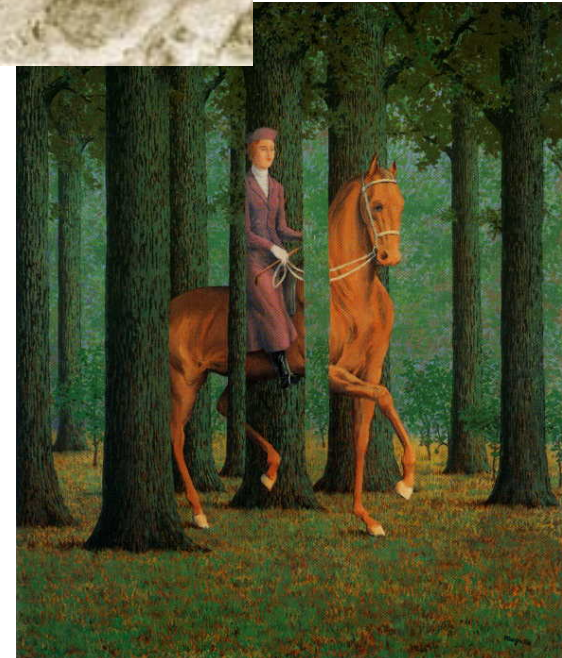
- Appearance only or location and appearance



Representation

– Invariances

- View point
- Illumination
- Occlusion
- Scale
- Deformation
- Clutter
- etc.



Representation

- To handle intra-class variability, it is convenient to describe object categories using probabilistic models
- Object models: Generative vs Discriminative vs hybrid

Object categorization: the statistical viewpoint



$$p(\textit{zebra} \mid \textit{image})$$

vs.

$$p(\textit{no zebra} \mid \textit{image})$$

- Bayes rule: $P(A|B) = \frac{P(B|A) P(A)}{P(B)}$

$$\frac{p(\textit{zebra} \mid \textit{image})}{p(\textit{no zebra} \mid \textit{image})}$$

Object categorization: the statistical viewpoint



$$p(\textit{zebra} | \textit{image})$$

vs.

$$p(\textit{no zebra} | \textit{image})$$

- Bayes rule: $P(A|B) = \frac{P(B|A) P(A)}{P(B)}$

$$\underbrace{\frac{p(\textit{zebra} | \textit{image})}{p(\textit{no zebra} | \textit{image})}}_{\text{posterior ratio}} = \underbrace{\frac{p(\textit{image} | \textit{zebra})}{p(\textit{image} | \textit{no zebra})}}_{\text{likelihood ratio}} \cdot \underbrace{\frac{p(\textit{zebra})}{p(\textit{no zebra})}}_{\text{prior ratio}}$$

Object categorization: the statistical viewpoint

- Discriminative methods model posterior
- Generative methods model likelihood and prior

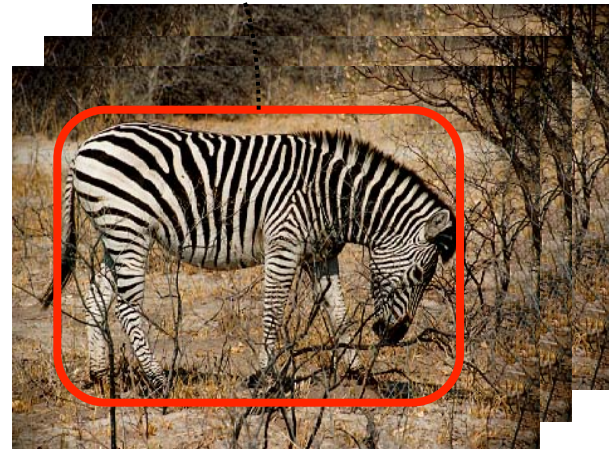
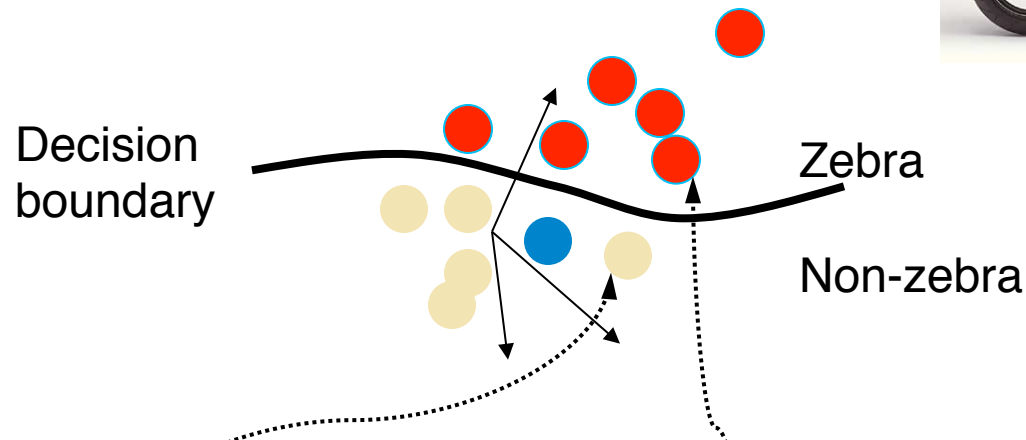
- Bayes rule:

$$\underbrace{\frac{p(\text{zebra} | \text{image})}{p(\text{no zebra} | \text{image})}}_{\text{posterior ratio}} = \underbrace{\frac{p(\text{image} | \text{zebra})}{p(\text{image} | \text{no zebra})}}_{\text{likelihood ratio}} \cdot \underbrace{\frac{p(\text{zebra})}{p(\text{no zebra})}}_{\text{prior ratio}}$$

Discriminative models

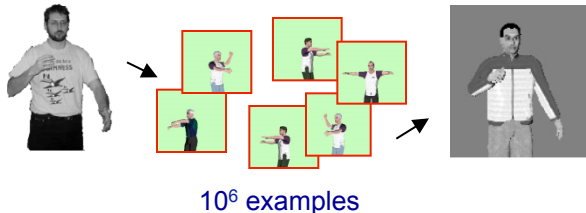
- Modeling the posterior ratio:

$$\frac{p(\text{zebra} | \text{image})}{p(\text{no zebra} | \text{image})}$$



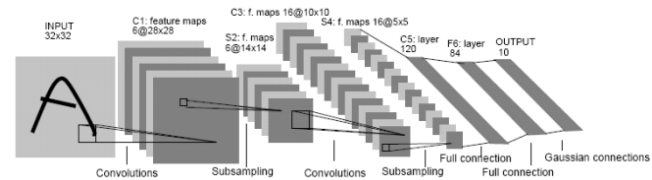
Discriminative models

Nearest neighbor



Shakhnarovich, Viola, Darrell 2003
Berg, Berg, Malik 2005...

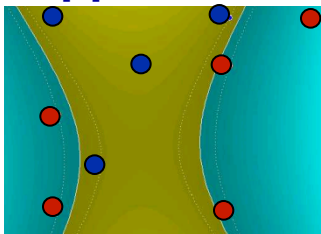
Neural networks



LeCun, Bottou, Bengio, Haffner 1998
Rowley, Baluja, Kanade 1998

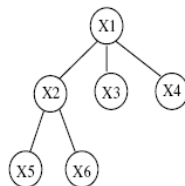
...

Support Vector Machines



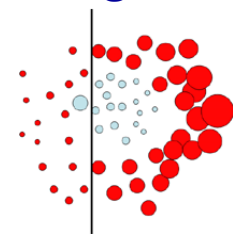
Guyon, Vapnik, Heisele,
Serre, Poggio...

Latent SVM Structural SVM



Felzenszwalb 00
Ramanan 03...

Boosting

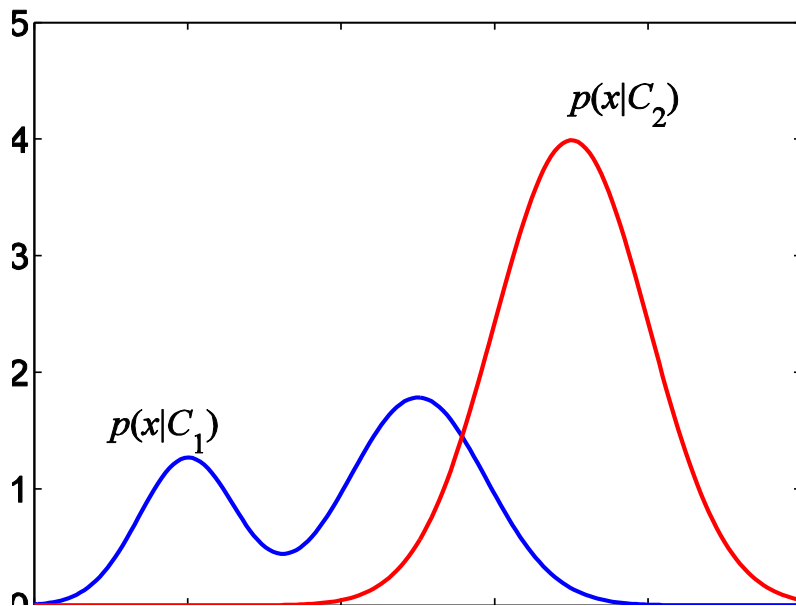


Viola, Jones 2001,
Torralba et al. 2004,
Opelt et al. 2006,...

Generative models

- Modeling the likelihood ratio:

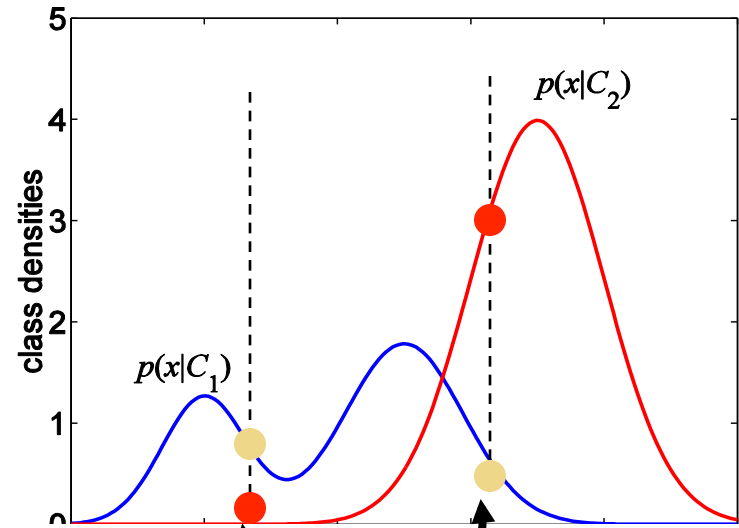
$$\frac{p(\text{image} \mid \text{zebra})}{p(\text{image} \mid \text{no zebra})}$$



Generative models



$p(\text{image} \mid \text{zebra})$	$p(\text{image} \mid \text{no zebra})$
High	Low
Low	High



Generative models

- Naïve Bayes classifier
 - Csurka Bray, Dance & Fan, 2004
- Hierarchical Bayesian topic models (e.g. pLSA and LDA)
 - Object categorization: Sivic et al. 2005, Sudderth et al. 2005
 - Natural scene categorization: Fei-Fei et al. 2005
- 2D Part based models
 - Constellation models: Weber et al 2000; Fergus et al 200
 - Star models: ISM (Leibe et al 05)
- 3D part based models:
 - multi-aspects: Sun, et al, 2009

Basic Problems in Object Recognition

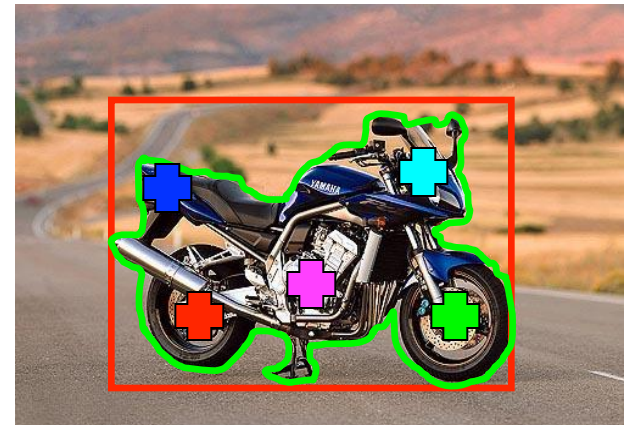
- Representation
 - How to represent an object category; which classification scheme?
- Learning
 - How to learn the classifier, given training data?
- Recognition
 - How the classifier is to be used on novel data?

Learning

- Learning parameters: What are you maximizing?
Likelihood (Gen.) or performances on train/validation set
(Disc.)

Learning

- Learning parameters: What are you maximizing?
Likelihood (Gen.) or performances on train/validation set (Disc.)
- Level of supervision
 - Manual segmentation; bounding box; image labels; noisy labels
- Batch/incremental
- Priors



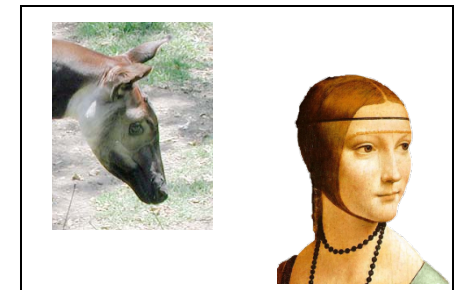
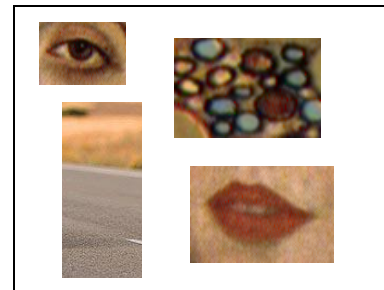
Learning

- Learning parameters: What are you maximizing? Likelihood (Gen.) or performances on train/validation set (Disc.)
- Level of supervision
 - Manual segmentation; bounding box; image labels; noisy labels

- Batch/incremental

- Priors

- Training images:
 - Issue of overfitting
 - Negative images for discriminative methods

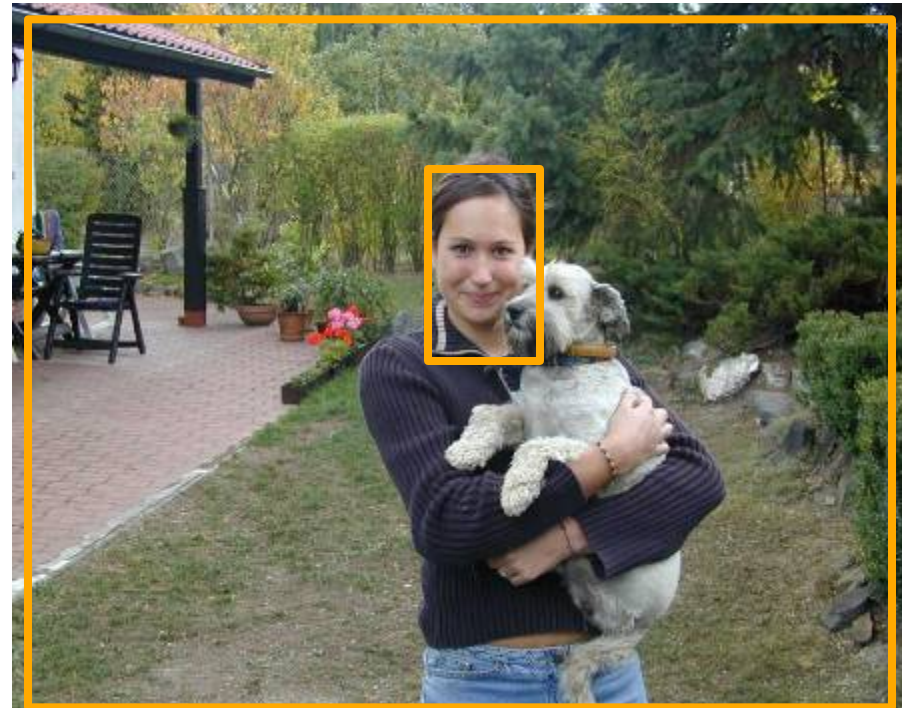


Basic Problems in Object Recognition

- Representation
 - How to represent an object category; which classification scheme?
- Learning
 - How to learn the classifier, given training data?
- Recognition
 - How the classifier is to be used on novel data?

Recognition

- Recognition task: classification, detection, etc..



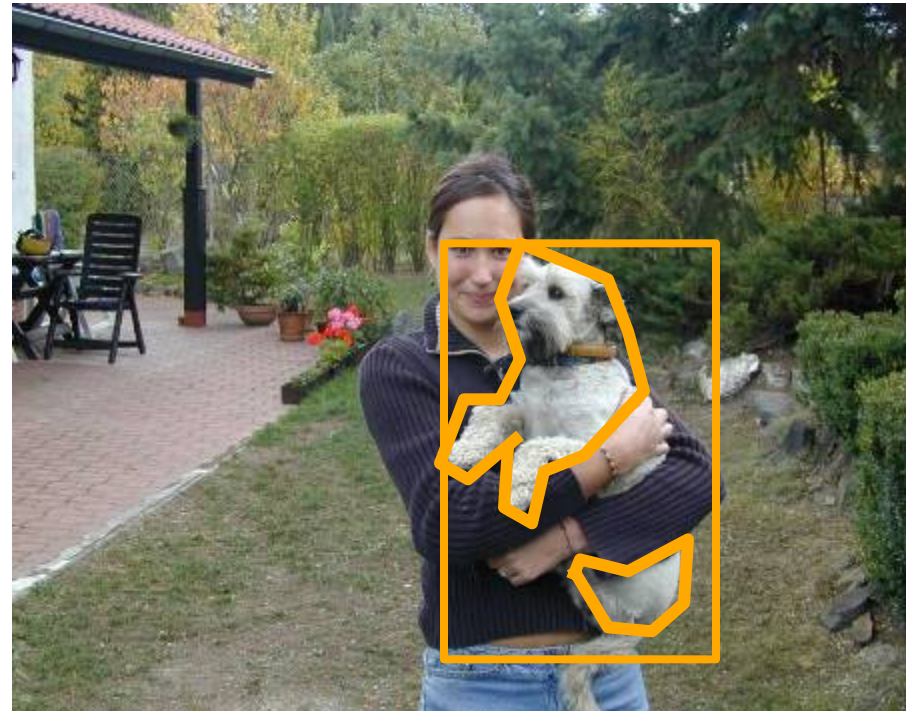
Recognition

- Recognition task
- Search strategy: Sliding Windows [Viola, Jones 2001](#),
 - Simple
 - Computational complexity (x, y, S, θ, N of classes)
 - BSW by Lampert et al 08
 - Also, Alexe, et al 10



Recognition

- Recognition task
- Search strategy: Sliding Windows Viola, Jones 2001,
 - Simple
 - Computational complexity (x, y, S, θ, N of classes)
 - BSW by Lampert et al 08
 - Also, Alexe, et al 10
 - Localization
 - Objects are not boxes



Recognition

– Recognition task

– Search strategy: Sliding Windows [Viola, Jones 2001](#),

- Simple
- Computational complexity (x, y, S, θ, N of classes)

- BSW by [Lampert et al 08](#)

- Also, [Alexe, et al 10](#)

- Localization

- Objects are not boxes
 - Prone to false positive

Non max suppression:

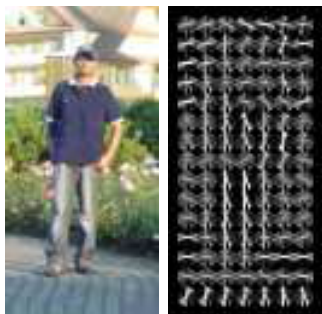
[Canny '86](#)

.....

[Desai et al , 2009](#)



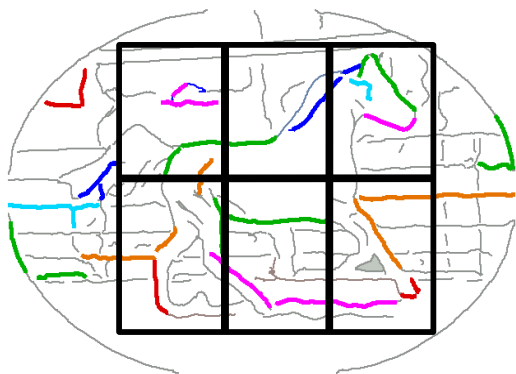
Successful methods using sliding windows



- Subdivide scanning window
- In each cell compute histogram of gradients orientation.

Code available: <http://pascal.inrialpes.fr/soft/olt/>

[Dalal & Triggs, CVPR 2005]



- Subdivide scanning window
- In each cell compute histogram of codewords of adjacent segments

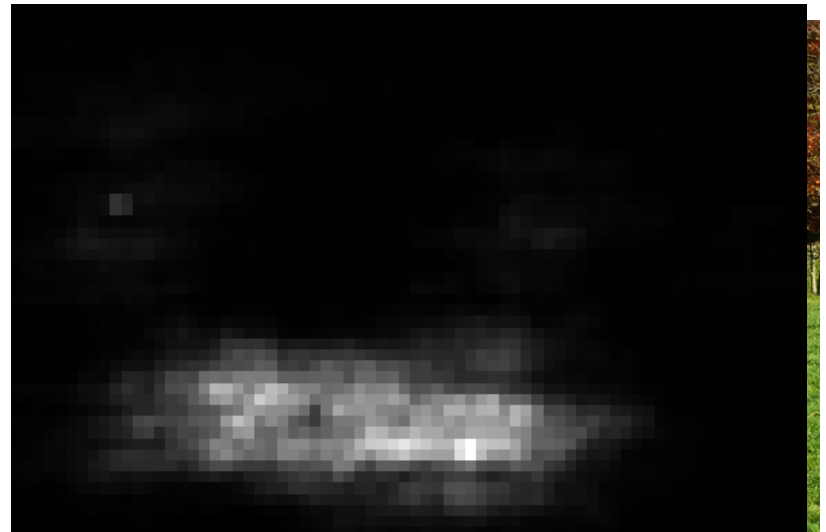
Code available: <http://www.vision.ee.ethz.ch/~calvin>

[Ferrari & al, PAMI 2008]

Recognition

- Recognition task
- Search strategy : Probabilistic “heat maps”

- Fergus et al 03
- Leibe et al 04



Recognition

- Recognition task
- Search strategy :
 - Hypothesis generation + verification

Recognition

- Recognition task
- Search strategy
- Attributes

- Savarese, 2007
- Sun et al 2009
- Liebelt et al., '08, 10
- Farhadi et al 09

Category: car
Azimuth = 225°
Zenith = 30°

- It has metal
- it is glossy
- has wheels

- Farhadi et al 09
- Lampert et al 09
- Wang & Forsyth 09



Recognition

- Recognition task
- Search strategy
- Attributes
- Context

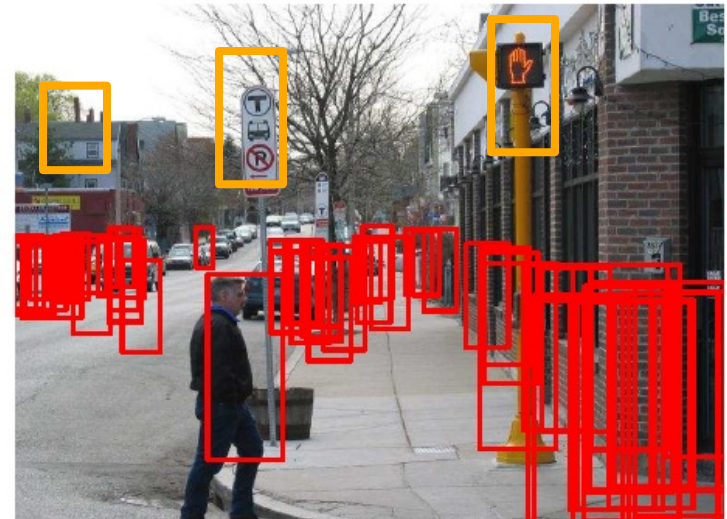


Semantic:

- Torralba et al 03
- Rabinovich et al 07
- Gupta & Davis 08
- Heitz & Koller 08
- L-J Li et al 08
- Bang & Fei-Fei 10

Geometric

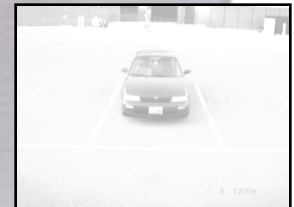
- Hoiem, et al 06
- Gould et al 09
- Bao, Sun, Savarese 10



Recognition of 3D objects



Single 3D object recognition

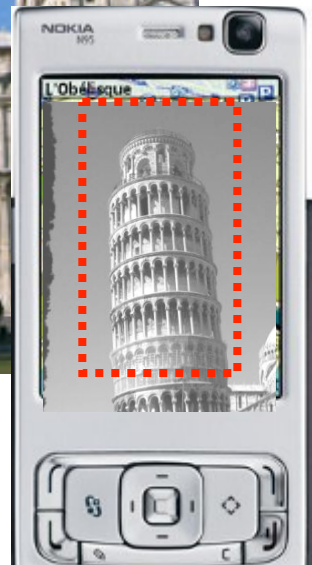


- Ballard, '81
- Grimson & L.-Perez, '87
- Lowe, '87

- Edelman et al. '91
- Ullman & Barsi, '91
- Rothwell '92
- Linderberg, '94
- Murase & Nayar '94

- Zhang et al '95
- Schmid & Mohr, '96
- Schiele & Crowley, '96
- Lowe, '99
- Jacob & Barsi, '99
- Mahamud and Herbert, 00

- Rothganger et al., '04
- Ferrari et al., '05
- Moreels and Perona, 05
- Brown & Lowe '05
- Snavely et al '06
- Yin & Collins, '07



+ GPS

Where is the crunchy nut?



Usual Challenges:

Variability due to:

- View point
- Illumination
- Occlusions

Recognition of single 3D objects

-Representation

-Features

-2D/3D Geometrical
constraints

-Model learning

-Recognition

-Hypothesis generation

-Validation

- Rothganger et al. '04, '06
- Brown et al, '05
- Lowe '99, '04
- Ferrari et al. '04, '06
- Lazebnick et al '04

Representation

Interest points -- or Regions (group of interest points)

- Detection

- Difference of Gaussian (DOG) [Lowe '99]
- Harris-Laplacian [Mikolajczyk & Schmid '01]
- Kadir-Brady [Kadir et al. '01]
- Laplacian [Gårding & Lindeberg, '96]

- Adaptation [invariants]

- Scale, rotation
- Affine

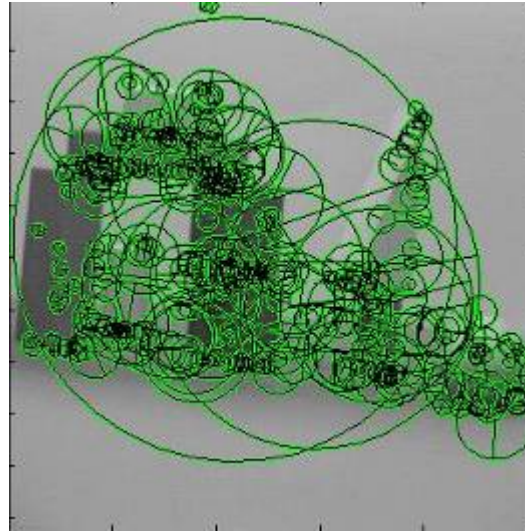
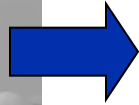
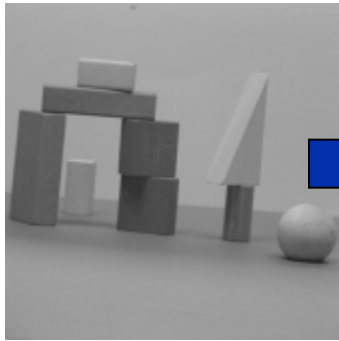
- Description

- SIFT
- Color histograms

Geometrical constraints

- 2D spatial layout of keypoints
- Tracks of keypoints (regions) across views
- 3D locations and/or surface normals

Difference of Gaussian (DOG): used in Lowe 99, Brown et al '05



Courtesy of D. Lowe

Harris-Laplace: used in Rothganger et al. '06



Courtesy of Rothganger et al.

Representation

Interest points -- or Regions (group of interest points)

- Detection

- Difference of Gaussian (DOG) [Lowe '99]
- Harris-Laplacian [Mikolajczyk & Schmid '01]
- Kadir-Brady [Kadir et al. '01]
- Laplacian [Gårding & Lindeberg, '96]



- x,y
- Scale
- Orientation
- Affine structure

- Adaptation [invariants]

- Scale, rotation
- Affine

- Description

- SIFT
- Color histograms

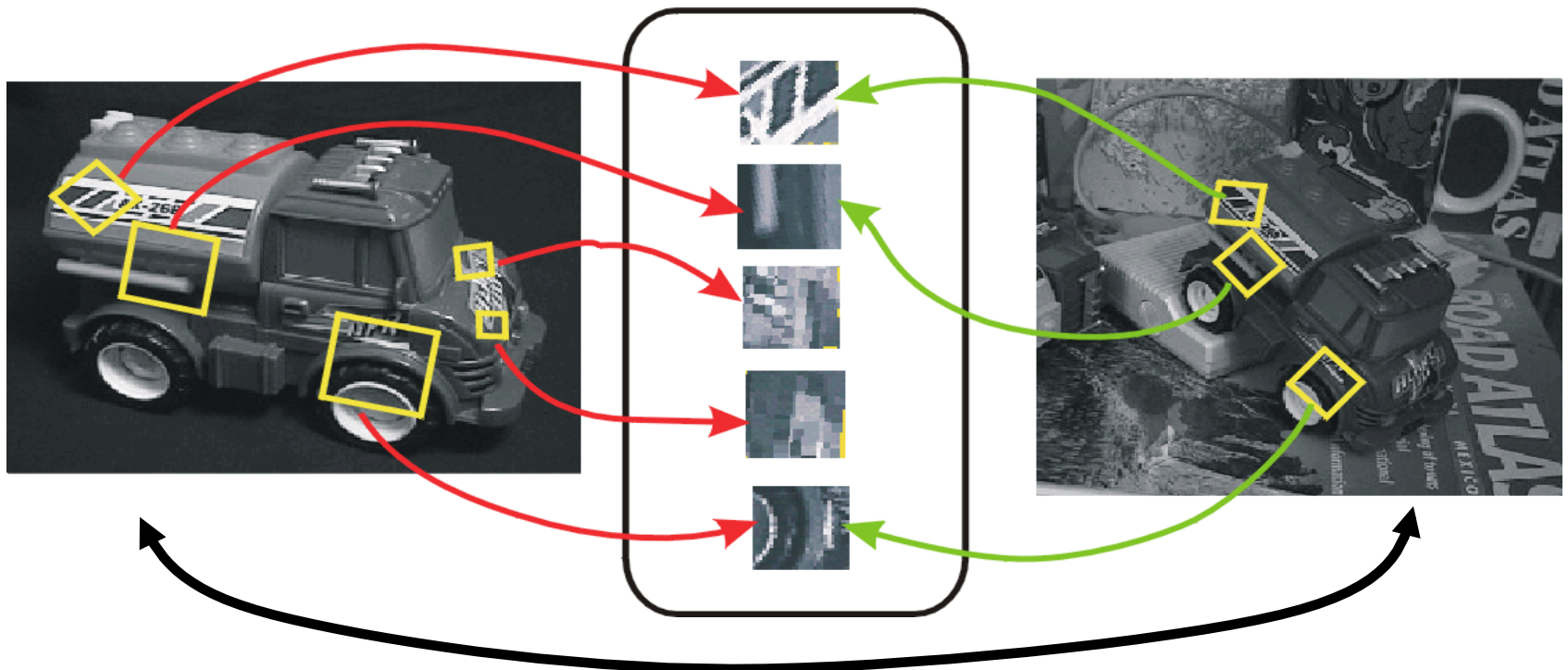
Geometrical constraints

- 2D spatial layout of keypoints
- Tracks of keypoints (regions) across views
- 3D locations and/or surface normals

Scale & orientation adaptation

[used in Lowe '99]

- keypoints are transformed in order to be invariant to translation, rotation, scale transformations



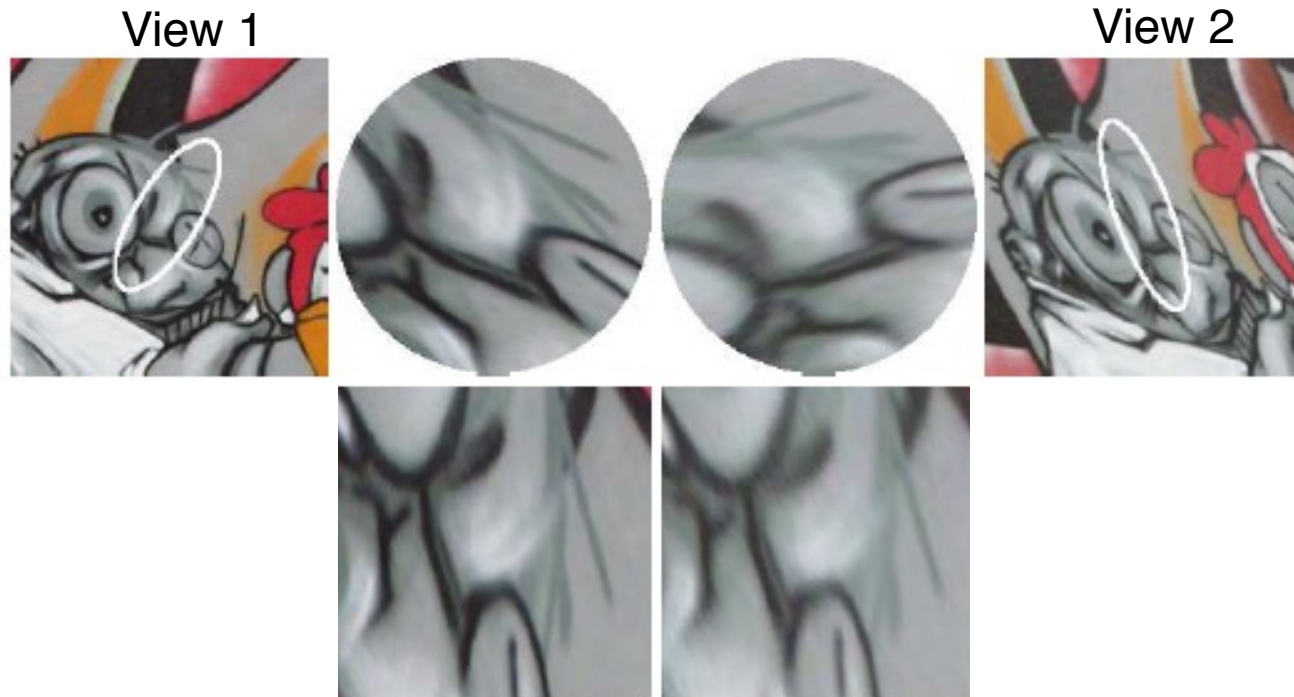
Courtesy of D. Lowe

Change of scale, pose, illumination...

Scale & orientation adaptation

[used in Rothganger et al. '03, '06]

1. Define elliptical region **using second moment matrix**
2. Use main canonical orientation to remove orientation ambiguity
3. Map ellipsis onto unit square



Courtesy of Rothganger et al

Representation

Interest points -- or Regions (group of interest points)

- Detection

- Difference of Gaussian (DOG) [Lowe '99]
- Harris-Laplacian [Mikolajczyk & Schmid '01]
- Kadir-Brady [Kadir et al. '01]
- Laplacian [Gårding & Lindeberg, '96]

- Adaptation [invariants]

- Scale, rotation
- Affine

- Description

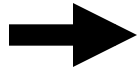
- SIFT
- Color histograms

Object representation

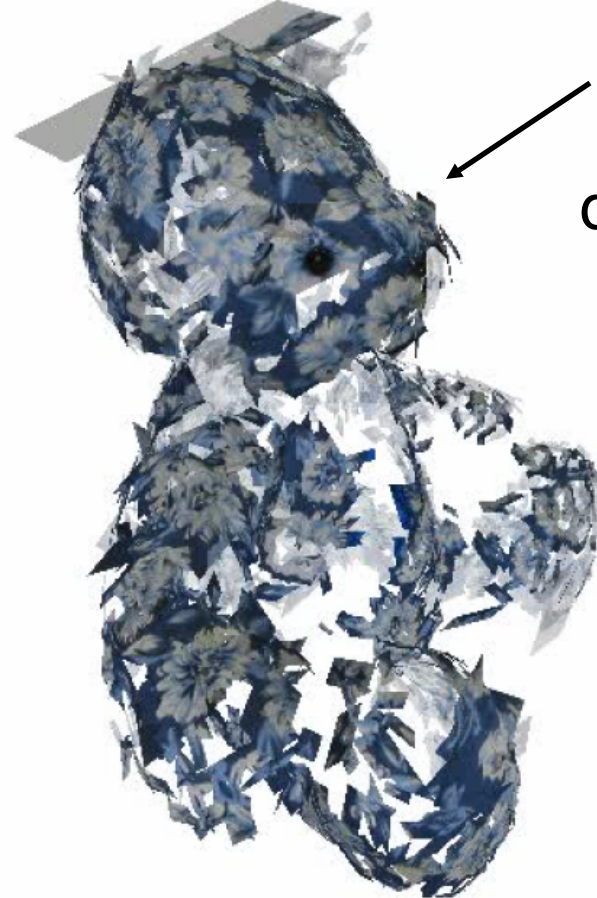
- 3D locations and/or surface normals
- 2D spatial layout of keypoints [collections of views]
- Tracks of keypoints (regions) across views

Object representation: 2D or 3D location of key points

[Lowe '99]



Rothganger et al. '06



$x, y, z +$
 $h, v +$
descriptor

Courtesy of Rothganger et al

Basic scheme

-Representation

- Features

- 2D/3D Geometrical constraints

-Model learning

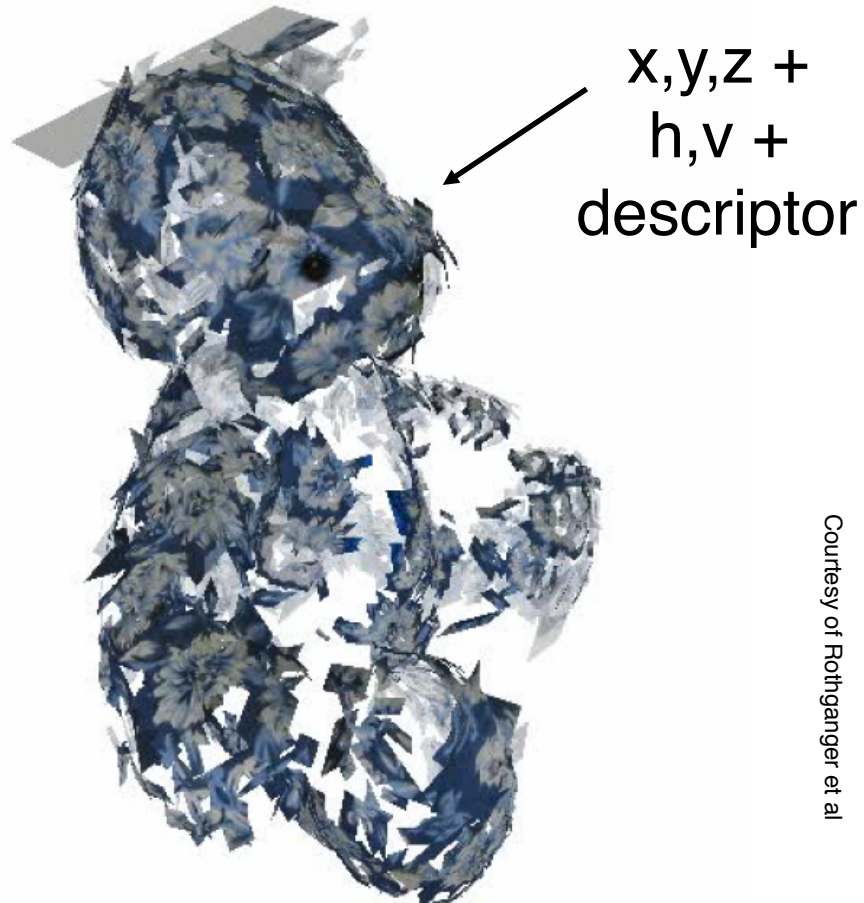
-Recognition

- hypothesis generation

- validation

Model learning

Rothganger et al. '03 '06



Courtesy of Rothganger et al

Model learning

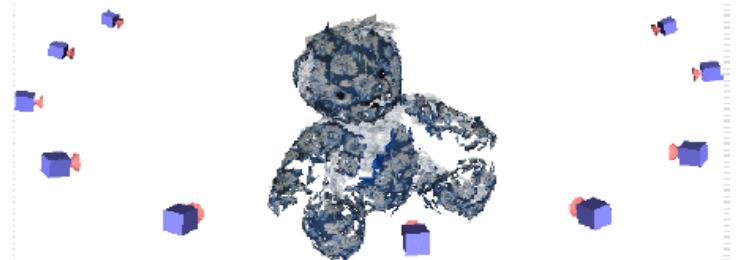
Rothganger et al. '03 '06

Build a 3D model:

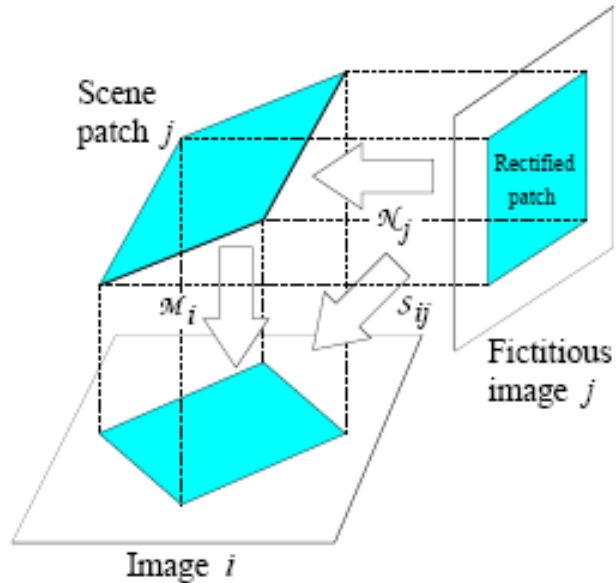
- N images of object from N different views
- Extract key points from each view
- Match key points between 2 views
- Use affine structure from motion to compute 3D location and orientation + camera locations from 2 views
- Find connected components
- Use bundle adjustment to refine the model
- Upgrade model to Euclidean assuming zero skew and square pixels



$$E = \sum_{j=1}^n \sum_{i \in I_j} |\mathcal{S}_{ij} - \mathcal{M}_i \mathcal{N}_j|^2,$$

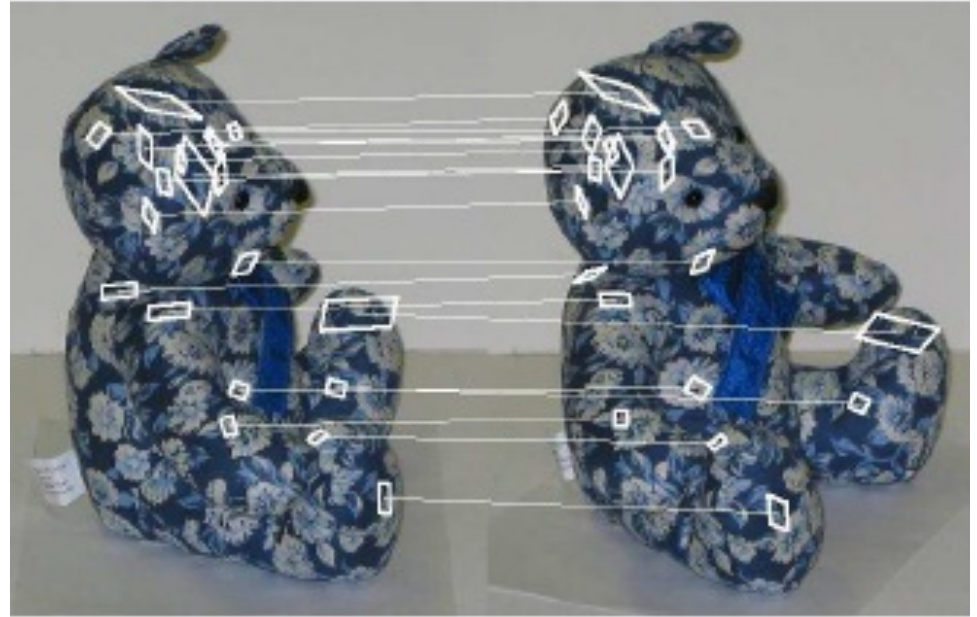


RANSAC Rothganger et al. '03 '06



$$\hat{S} \stackrel{\text{def}}{=} \begin{bmatrix} S_{11} & \dots & S_{1n} \\ \vdots & \ddots & \vdots \\ S_{m1} & \dots & S_{mn} \end{bmatrix} = \begin{bmatrix} \mathcal{M}_1 \\ \vdots \\ \mathcal{M}_m \end{bmatrix} [\mathcal{N}_1 \dots \mathcal{N}_n],$$

$$\mathcal{N}_j = \begin{bmatrix} H_j & V_j & C_j \\ 0 & 0 & 1 \end{bmatrix}$$



Courtesy of Rothganger et al

Algorithm:

[Affine factorization
Tomasi & Kanade '92]

Sample set = set of matches between views

1. Select a random sample of minimum required size [2 matches]
2. Compute a putative model from these
3. Compute the set of inliers to this model from whole sample space
4. Continue until model with the most inliers over all samples is found

Learnt models

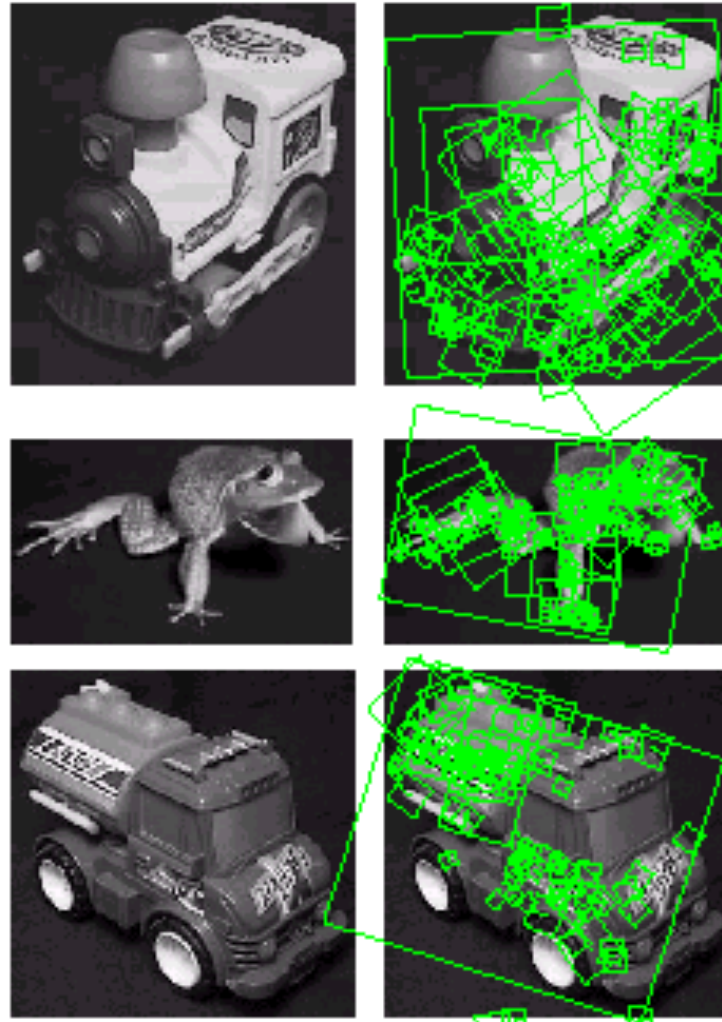
Rothganger et al. '03 '06



Courtesy of Rothganger et al

Learnt models

[Lowe '99]



Basic scheme

-Representation

- Features

- 2D/3D Geometrical constraints

-Model learning

-Recognition [object instance from object model]

- hypothesis generation

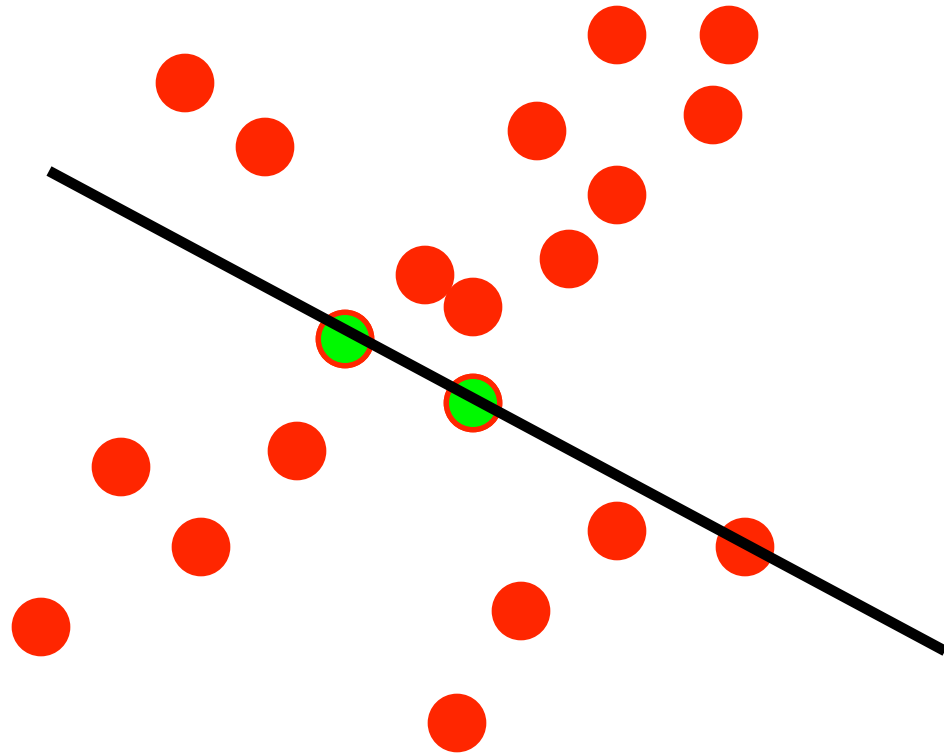
- Model verification

Recognition

Goal: given a query image I , identify object model in the image I (match learned model to I)

- Generate hypothesis
- Verify hypothesis
- Select hypothesis with lowest fitting error
- Generate recognition results

Line fitting with outliers

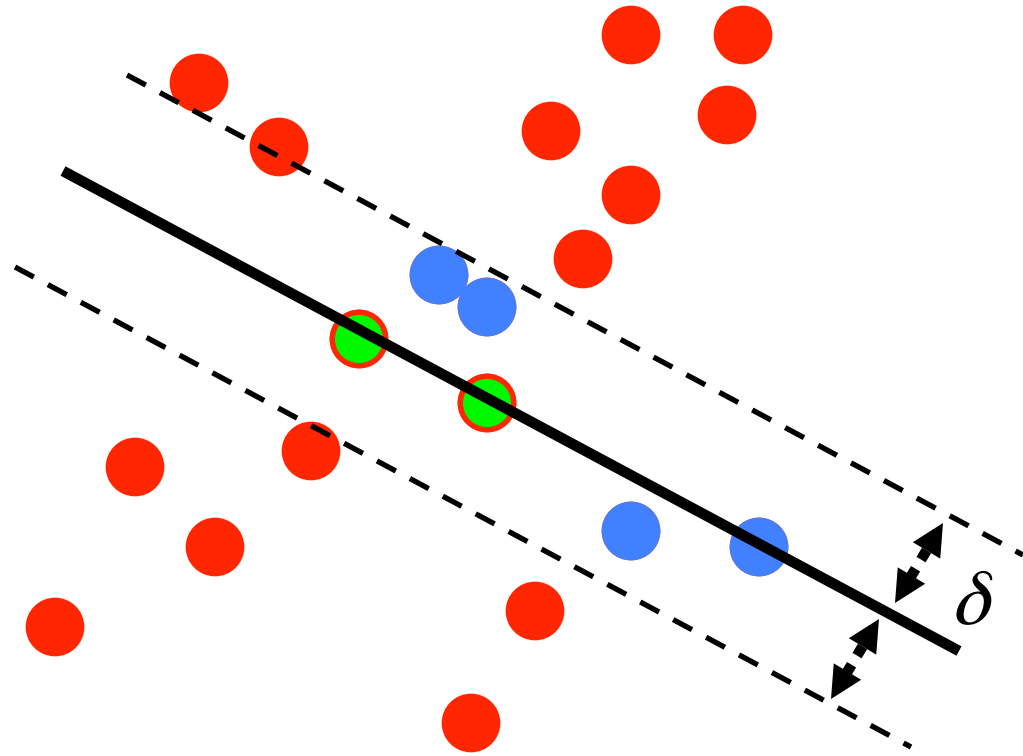


Sample set = set of points in 2D

Algorithm:

1. Select random sample of minimum required size to fit model [?] = [2]
 2. Compute a putative model from sample set
 3. Compute the set of inliers to this model from whole data set
- Repeat 1-3 until model with the most inliers over all samples is found

Line fitting with outliers



Sample set = set of points in 2D

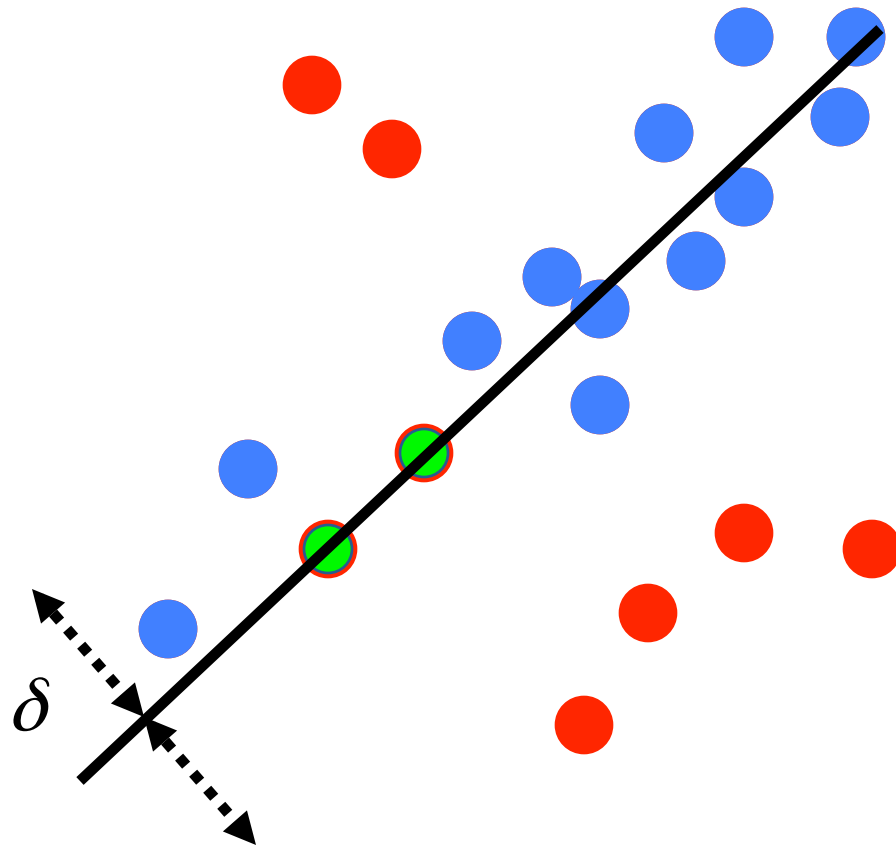
$$|\mathcal{O}| = 14$$

Algorithm:

1. Select random sample of minimum required size to fit model [?] = [2]
2. Compute a putative model from sample set
3. Compute the set of inliers to this model from whole data set

Repeat 1-3 until model with the most inliers over all samples is found

Line fitting with outliers

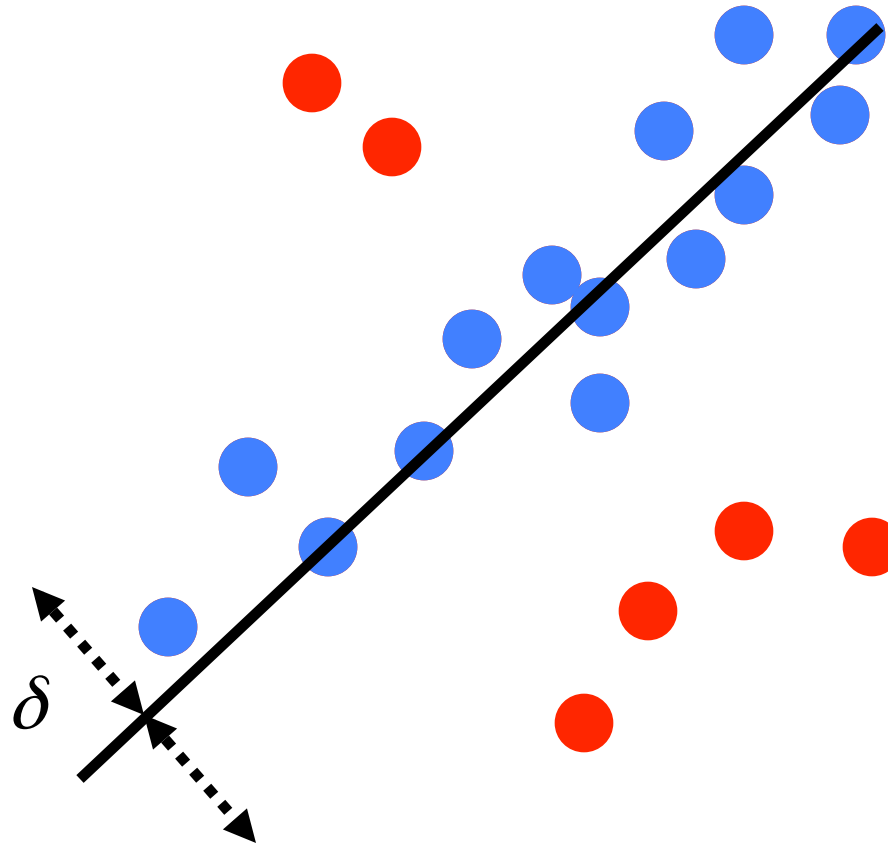


$$|\mathcal{O}| = 6$$

Algorithm:

1. Select random sample of minimum required size to fit model [?]
 2. Compute a putative model from sample set
 3. Compute the set of inliers to this model from whole data set
- Repeat 1-3 until model with the most inliers over all samples is found

Line fitting with outliers



$$\pi : I \rightarrow \{P, O\}$$

such that:

$$f(P, \beta) < \delta$$

$$\min_{\pi} |O|$$

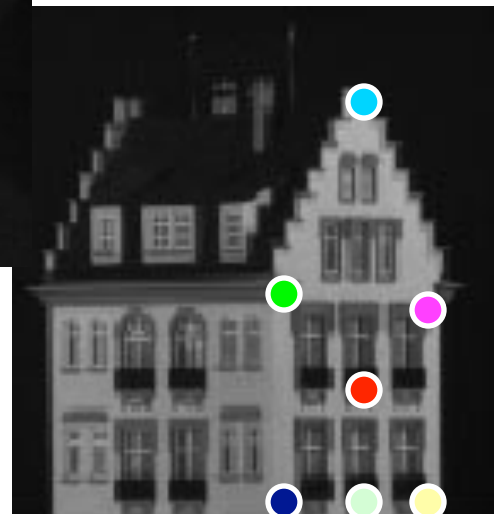
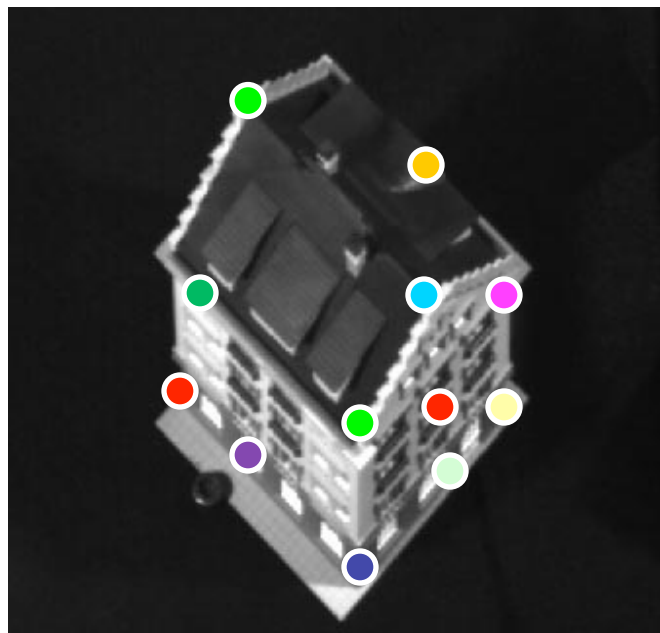
Model parameters

$$f(P, \beta) = \left\| \beta - (P^T P)^{-1} P^T \right\|$$

Recognition

Goal: given a query image I , identify object model in the image I (match learned model to I)

query



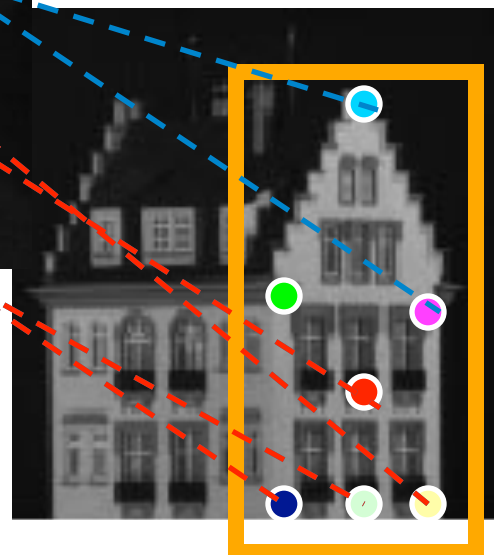
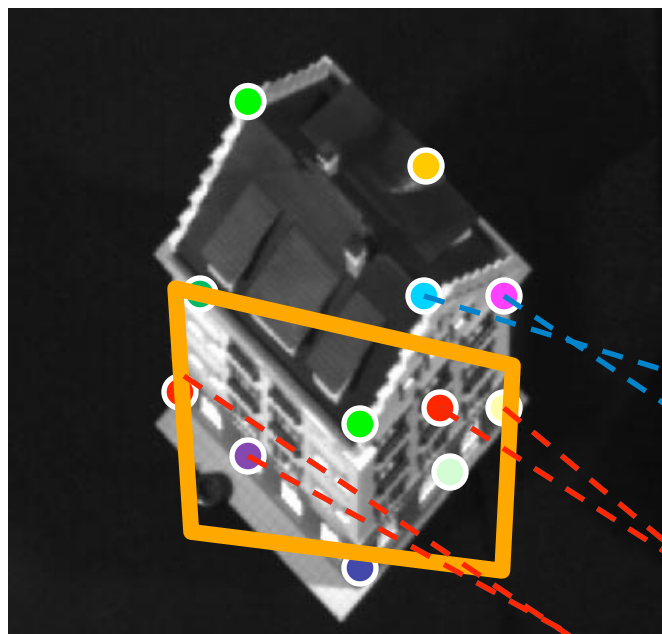
model

Recognition

Goal: given a query image I , identify object model in the image I (match learned model to I)

- Generate hypothesis
- Verify hypothesis
- Select hypothesis with lowest fitting error
- Generate recognition results

query



model

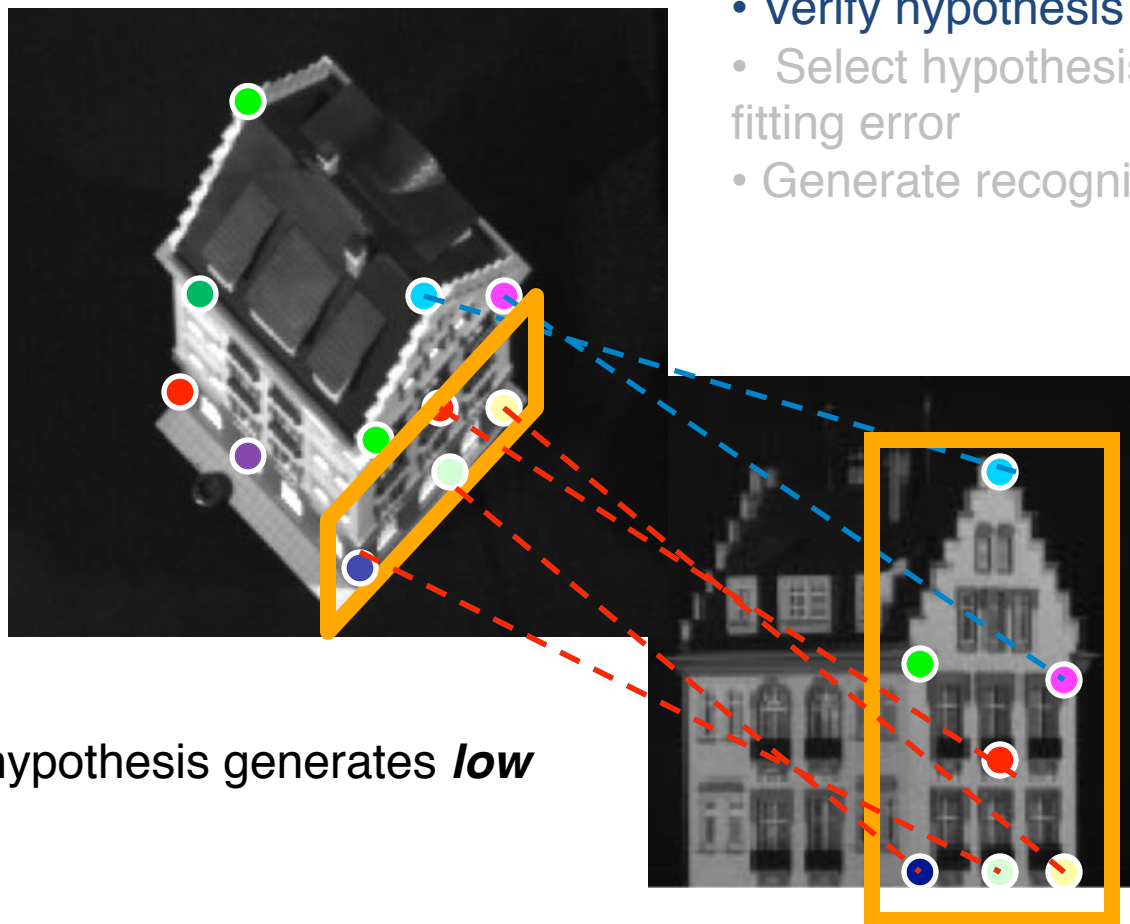
Verification: The hypothesis generates *high* fitting error

Recognition

Goal: given a query image I , identify object model in the image I (match learned model to I)

- Generate hypothesis
- Verify hypothesis
- Select hypothesis with lowest fitting error
- Generate recognition results

query



Verification: The hypothesis generates *low* fitting error

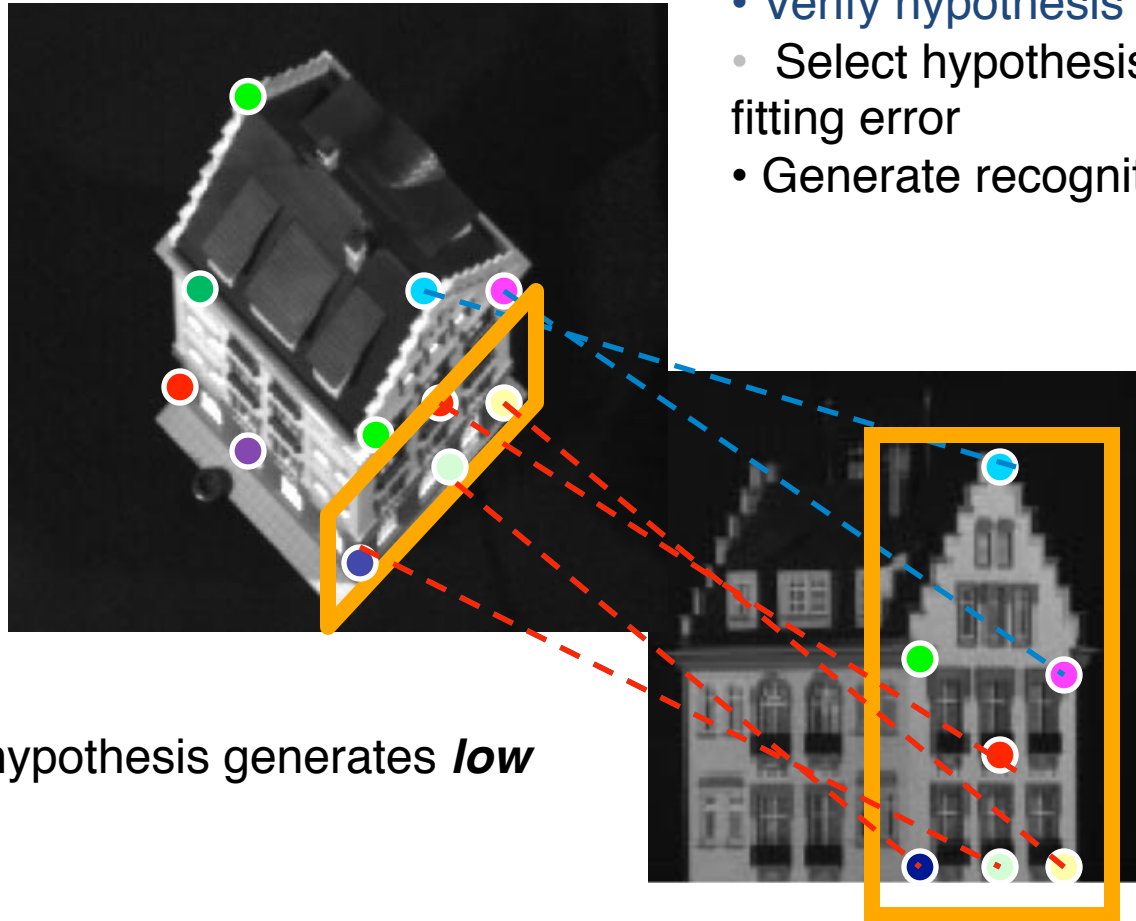
model

Recognition

Goal: given a query image I , identify object model in the image I (match learned model to I)

- Generate hypothesis
- Verify hypothesis
- Select hypothesis with lowest fitting error
- Generate recognition results

query



Verification: The hypothesis generates *low* fitting error

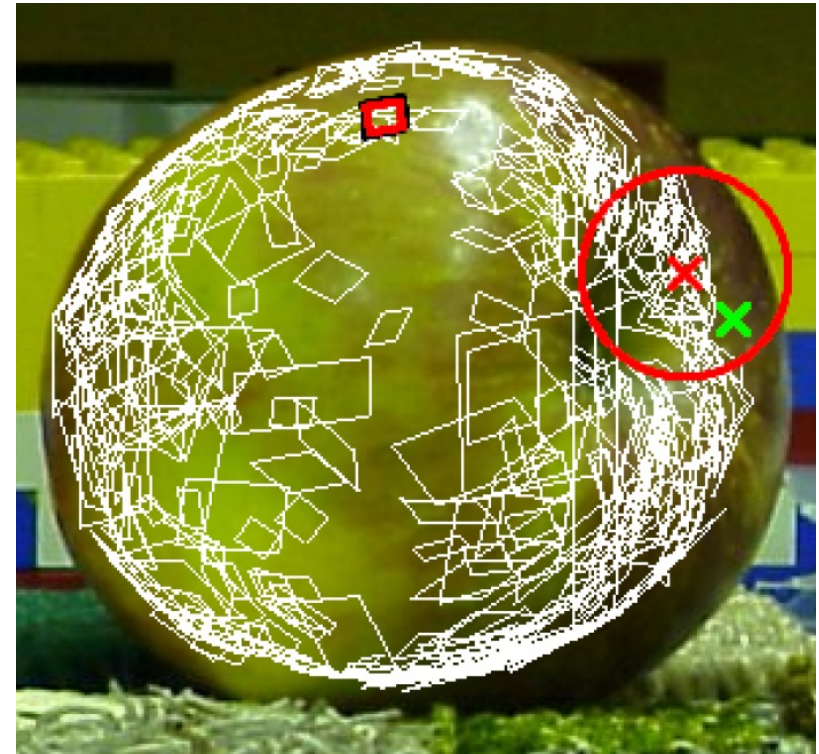
model

Recognition

hypothesis generation & model verification

[Rothganger et al. '03 '06]

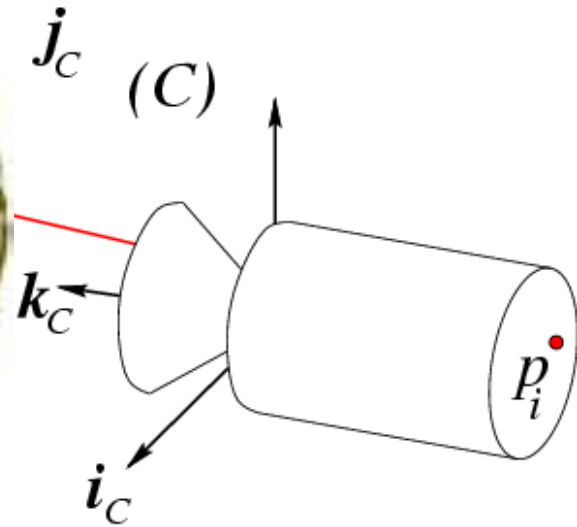
- Find (appearance based) matches between model keypoints and test image
- Use RANSAC to find a set of matches consistent with a candidate camera pose:
 - For every 2 pairs of matches
 - Compute camera
 - Use camera to project other matched 3D model patches into test image
 - Verification test



Courtesy of Rothganger et al

Recognition

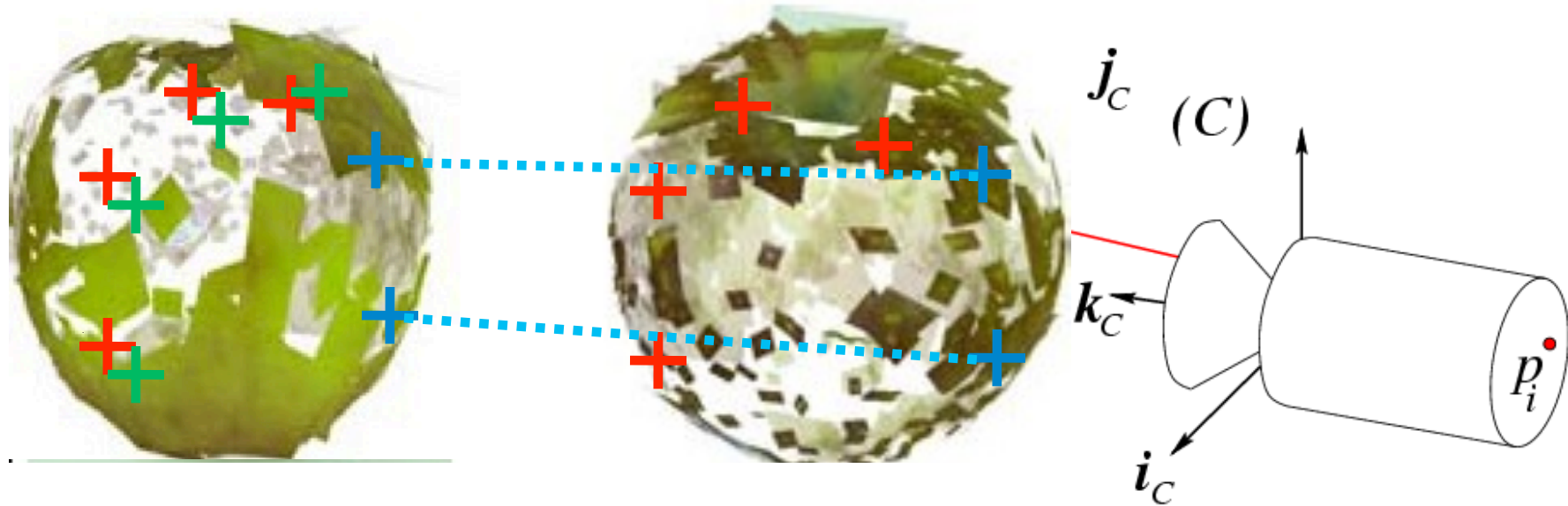
[Rothganger et al. '03 '06]



1. Find matches between model and test image features

Recognition

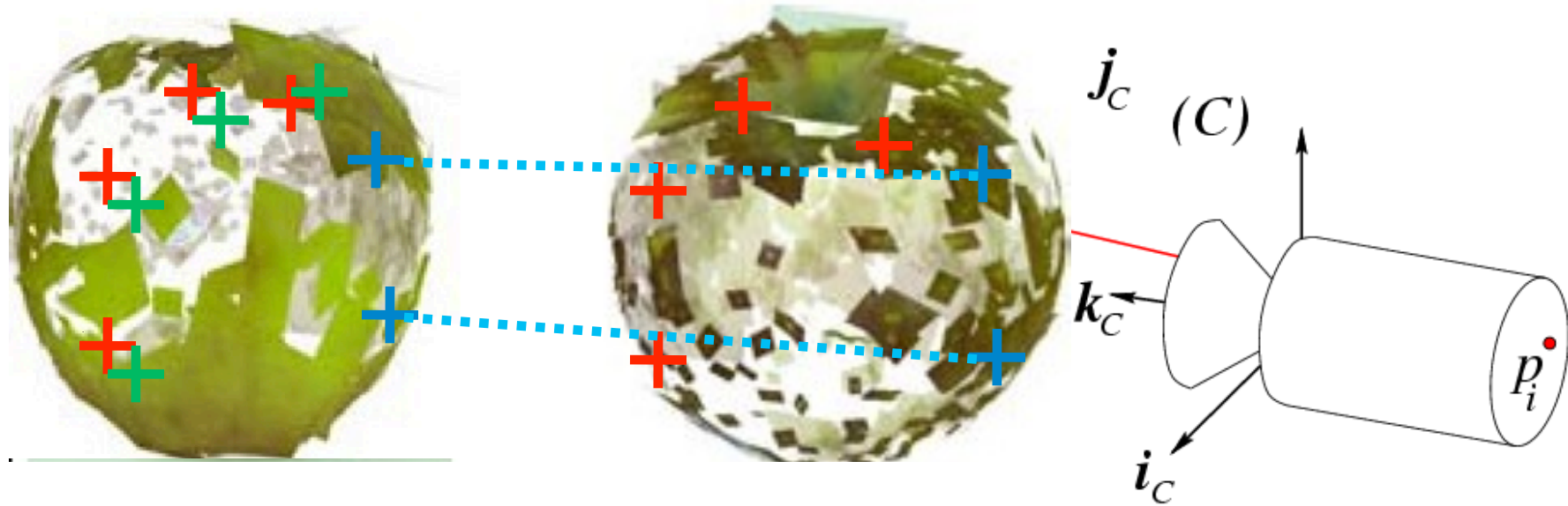
[Rothganger et al. '03 '06]



1. Find matches between model and test image features
2. Generate hypothesis:
 - Compute transformation M from N matches ($N=2$; affine camera; affine key points)
3. Model verification
 - Use M to project other matched 3D model features into test image
 - Compute residual = $D(\text{projections, measurements})$

Recognition

[Rothganger et al. '03 '06]



Goal:

Estimate (fit) the best M in presence of outliers

Object to recognize



Initial matches based on appearance



Matches verified with geometrical constraints



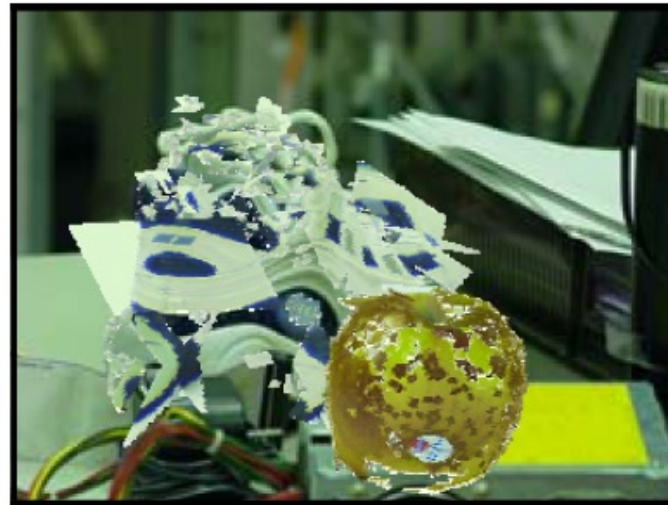
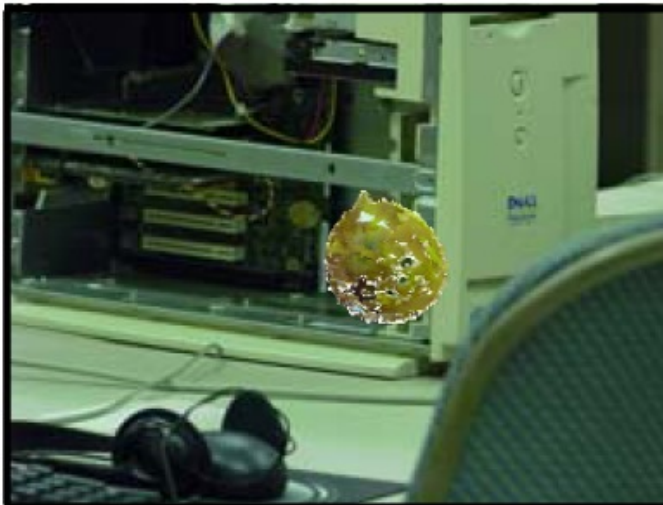
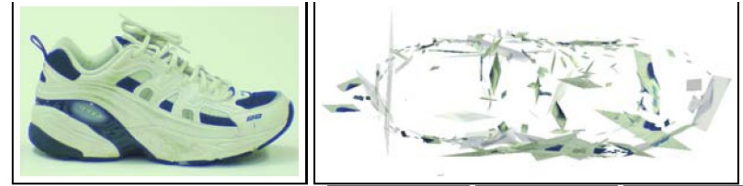
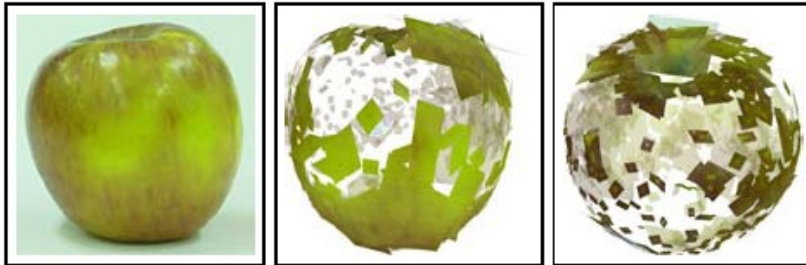
Recovered pose



Courtesy of Rothganger et al

3D Object Recognition results

Rothganger et al. '03 '06



Courtesy of Rothganger et al

- Handle severe clutter

3D Object Recognition results

Lowe. '99, '04



- Handle severe occlusions
- Fast!

Courtesy of D. Lowe

3D Object Recognition results

[Ferrari et al '04]

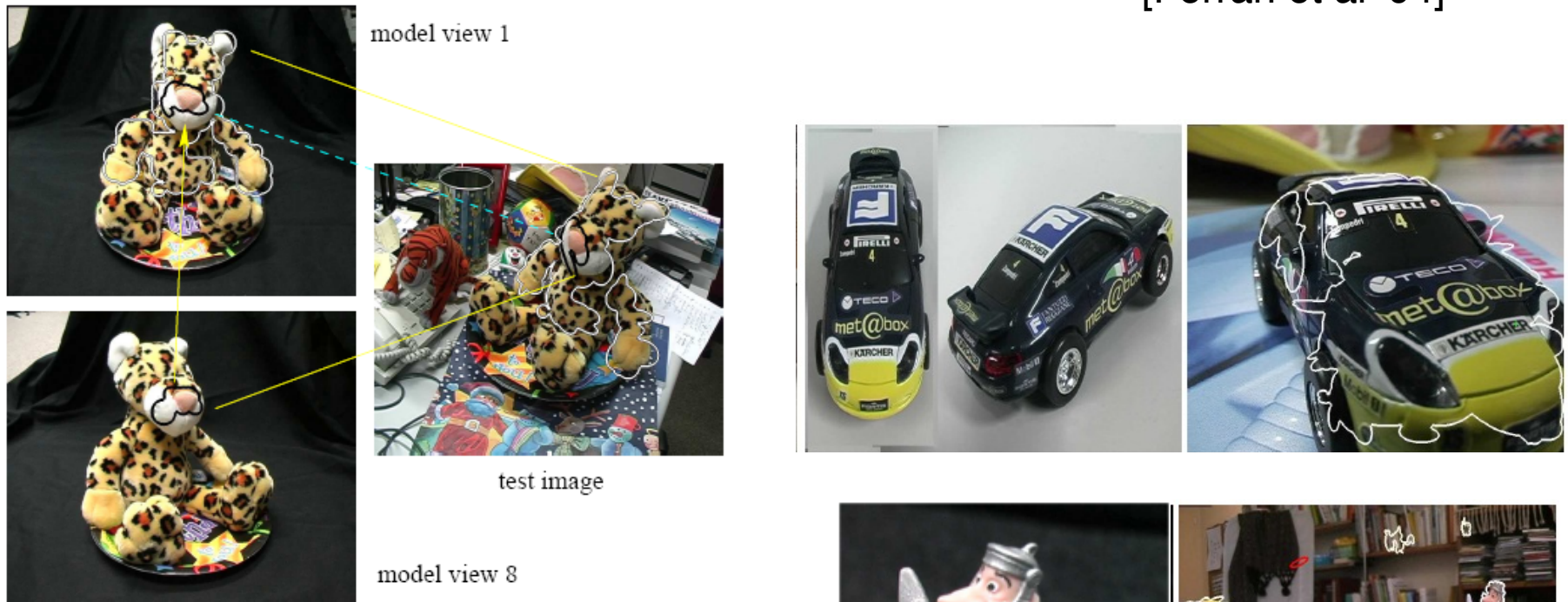


Figure 17: *Two compatible (and correct) GAMs. The nose GAM (black) is initially matched from model view 8, and is transferred to model view 1. Note how the other GAM (white) is very large and covers the head, arms and chest. A GAM can extend over multiple facets when the combination of viewpoints and surface orientations make the affine transformations of the region matches vary smoothly even across facet edges. In these cases, the resulting GAMs are larger and therefore more reliable and relevant.*

Courtesy of Ferrari et al

3D Object Recognition results

Edward Hsiao, Alvaro Collet and Martial Hebert. **Making specific features less discriminative to improve point-based 3D object recognition.** *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, June, 2010.



Next Lecture: AdaBoost and Face Detection

- Readings: FP 17.1; SZ 14.1
- BG: “The Boosting Approach to Machine Learning” by Schapire, MSRI Workshop on Nonlinear Estimation and Classification 2002.
- More Background: “Rapid Object Detection using a Boosted Cascade of Simple Features” Viola and Jones, CVPR 2001.