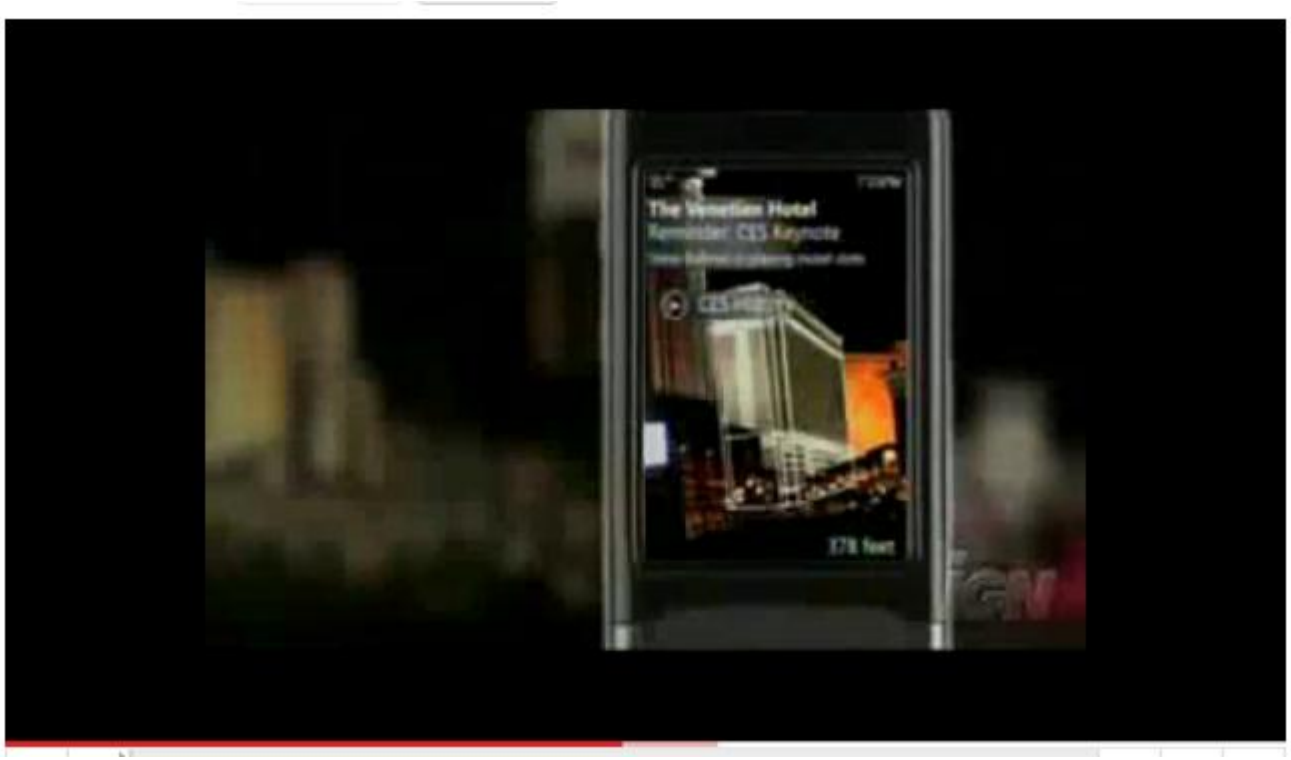
CEE598 - Visual Sensing for Civil Infrastructure Eng. & Mgmt.

Session 18 – Object Recognition I

Mani Golparvar-Fard
Department of Civil and Environmental Engineering
3129D, Newmark Civil Engineering Lab
e-mail: mgolpar@illinois.edu

Object Recognition

- Bill Gates demoing visual recognition gadget @ CES 2008



http://www.youtube.com/watch?v=LwRsvKhWSB0&feature=related

Outline

- Object Recognition

- Introduction
- Recognition of single 3D objects
  - Bag of world models
  - Part based models
  - Models for 3D objects categorization

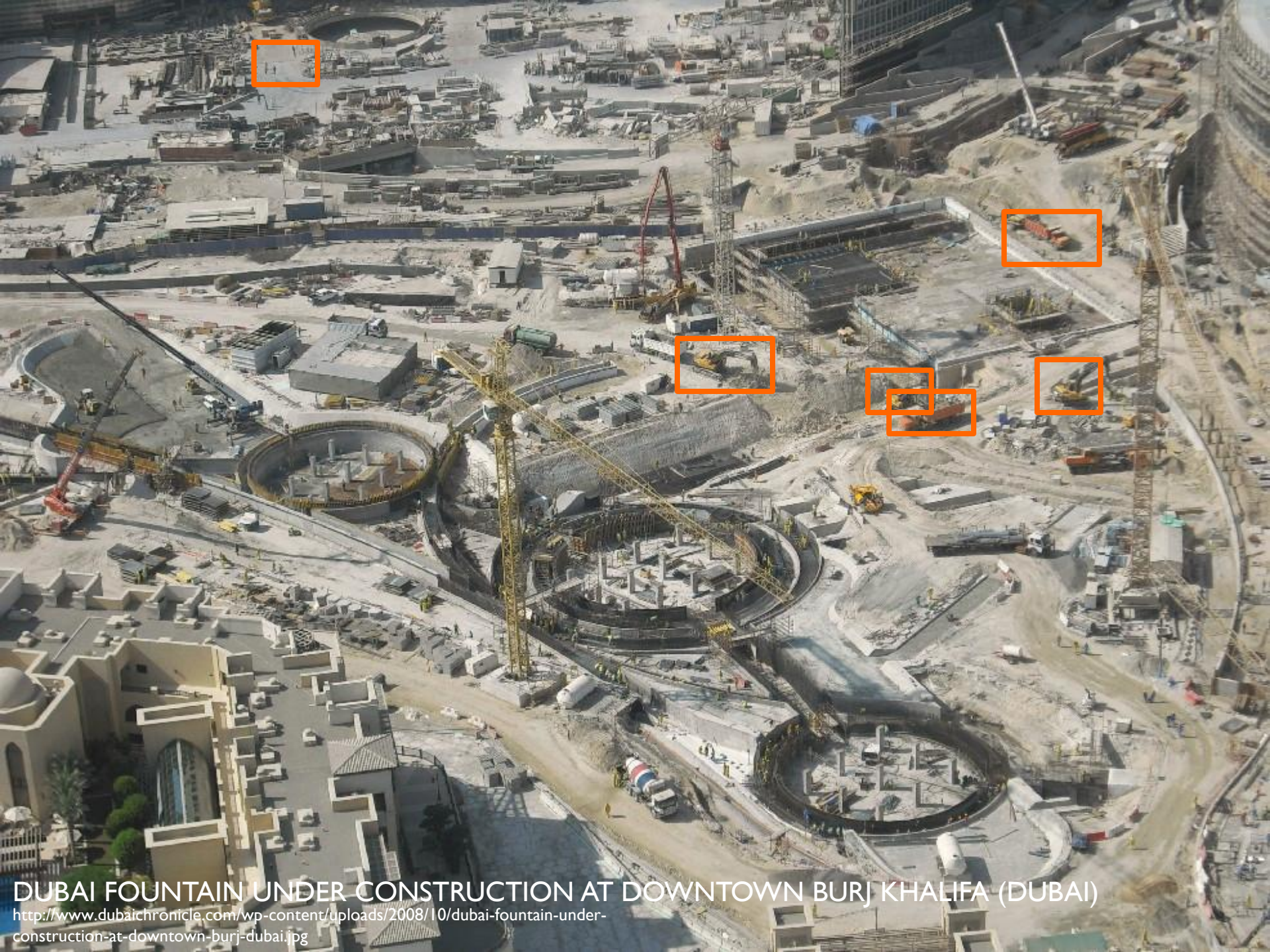
Some slides of this lectures are courtesy of prof S. Savarese, prof S. Lazebnik & S. Seitz, R. Szeliski, M. Pollefeys, K. Hassan-Shafique. S. Thrun

Outline

- Object Recognition

- Introduction
- Recognition of single 3D objects
- Bag of world models
- Part based models
- Models for 3D objects categorization

Part of this segment is based on the tutorial *"**Recognizing and Learning Object Categories: Year 2007**"*, by Prof A. Torralba, R. Fergus and F. Li

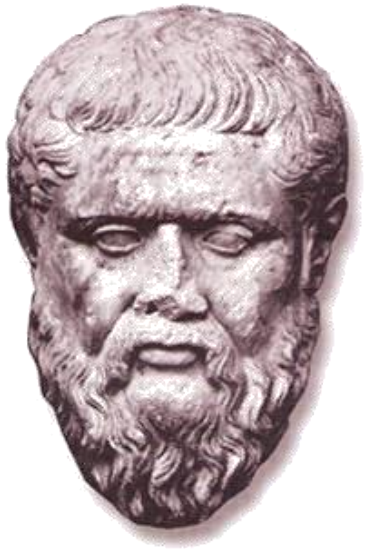


DUBAI FOUNTAIN UNDER CONSTRUCTION AT DOWNTOWN BURJ KHALIFA (DUBAI)

http://www.dubaichronicle.com/wp-content/uploads/2008/10/dubai-fountain-under-construction-at-downtown-burj-dubai.jpg

Bruegel, 1564

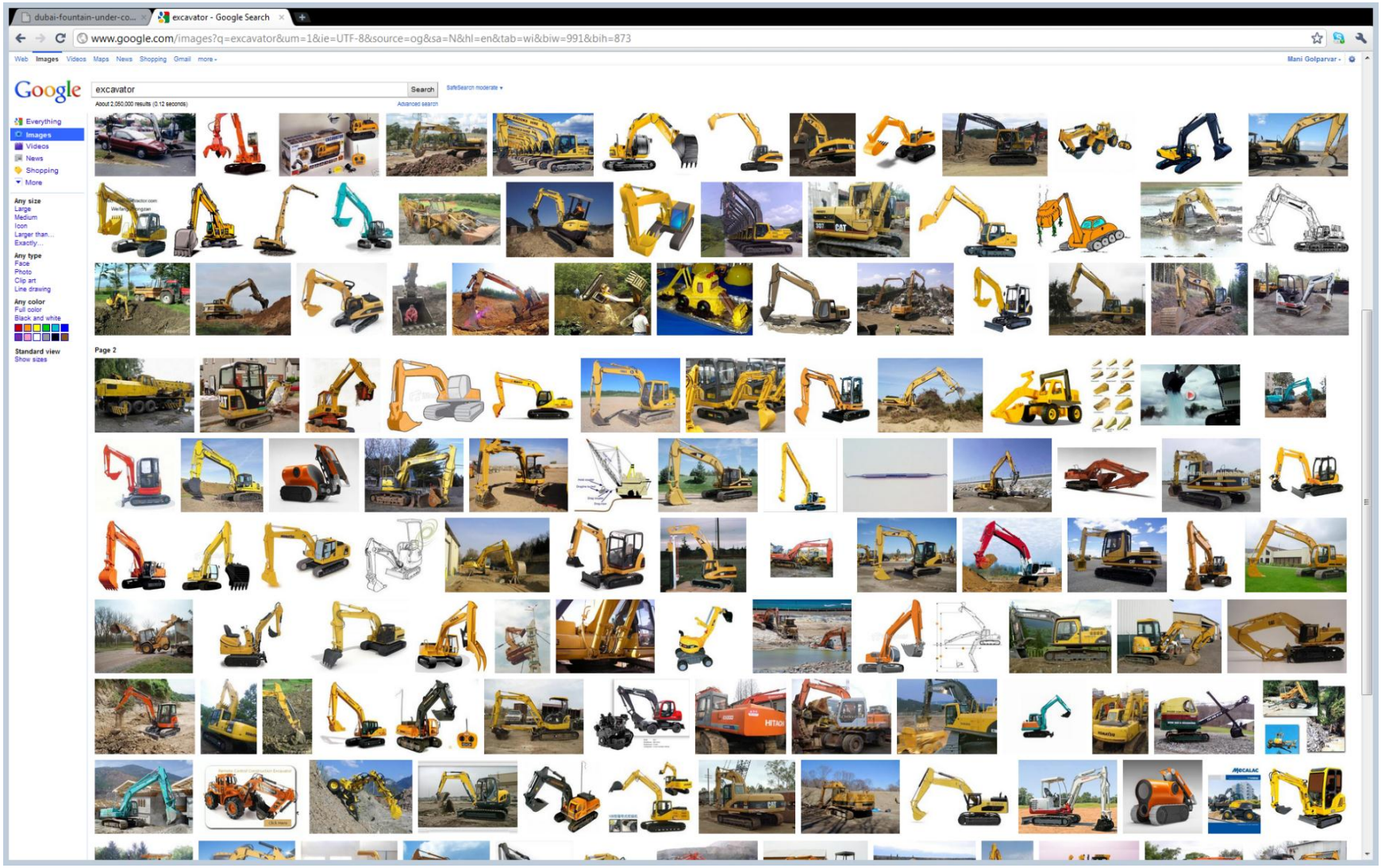
Plato said…



- Ordinary objects are classified together if they `participate' in the same abstract Form, such as the Form of a Human or the Form of Quartz.
- Forms are proper subjects of philosophical investigation, for they have the highest degree of reality.
- Ordinary objects, such as humans, trees, and stones, have a lower degree of reality than the Forms.
- Fictions, shadows, and the like have a still lower degree of reality than ordinary objects and so are not proper subjects of philosophical enquiry.

What is this abstract form?

Example: Excavators



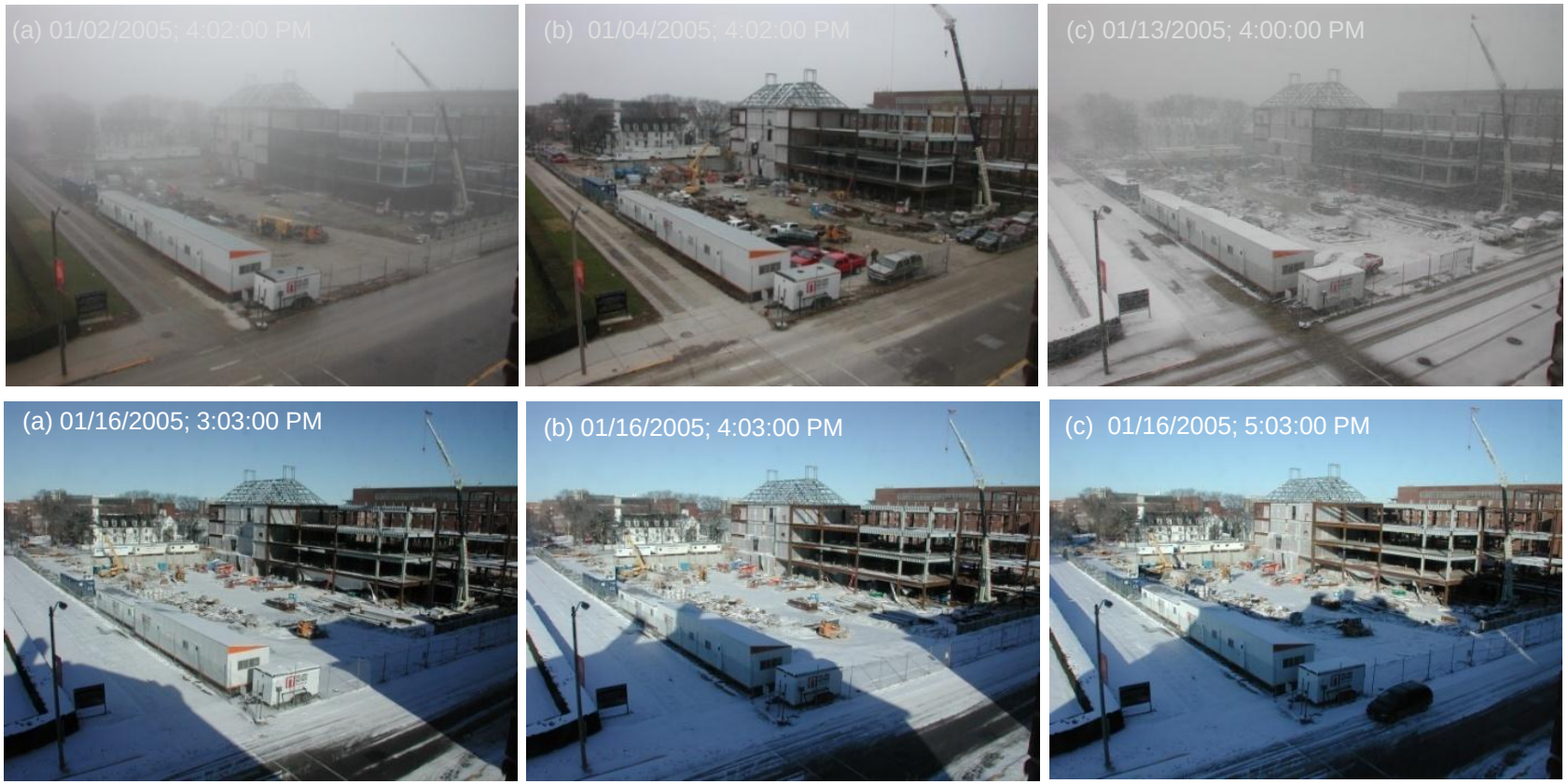

~10,000 to 30,000

Challenges: Viewpoint Variation

Michelangelo 1475-1564

slide credit: Fei-Fei, Fergus & Torralba

Challenges: illumination



Project: Institute of Genomic Biology, Courtesy of College of ACES, UIUC

Sources: Golparvar-Fard et al. (2009)

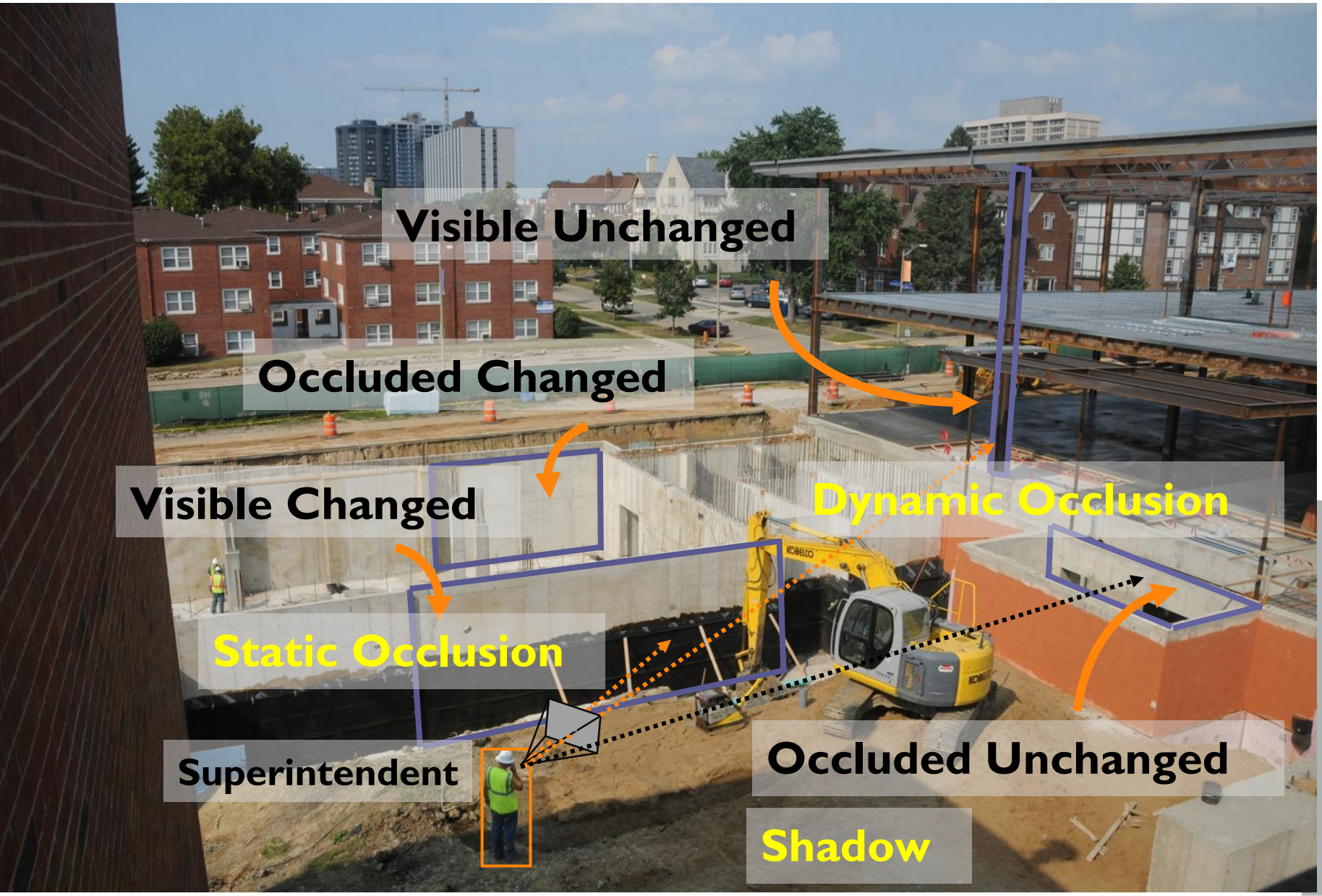
Challenges: scale

Challenges: scale

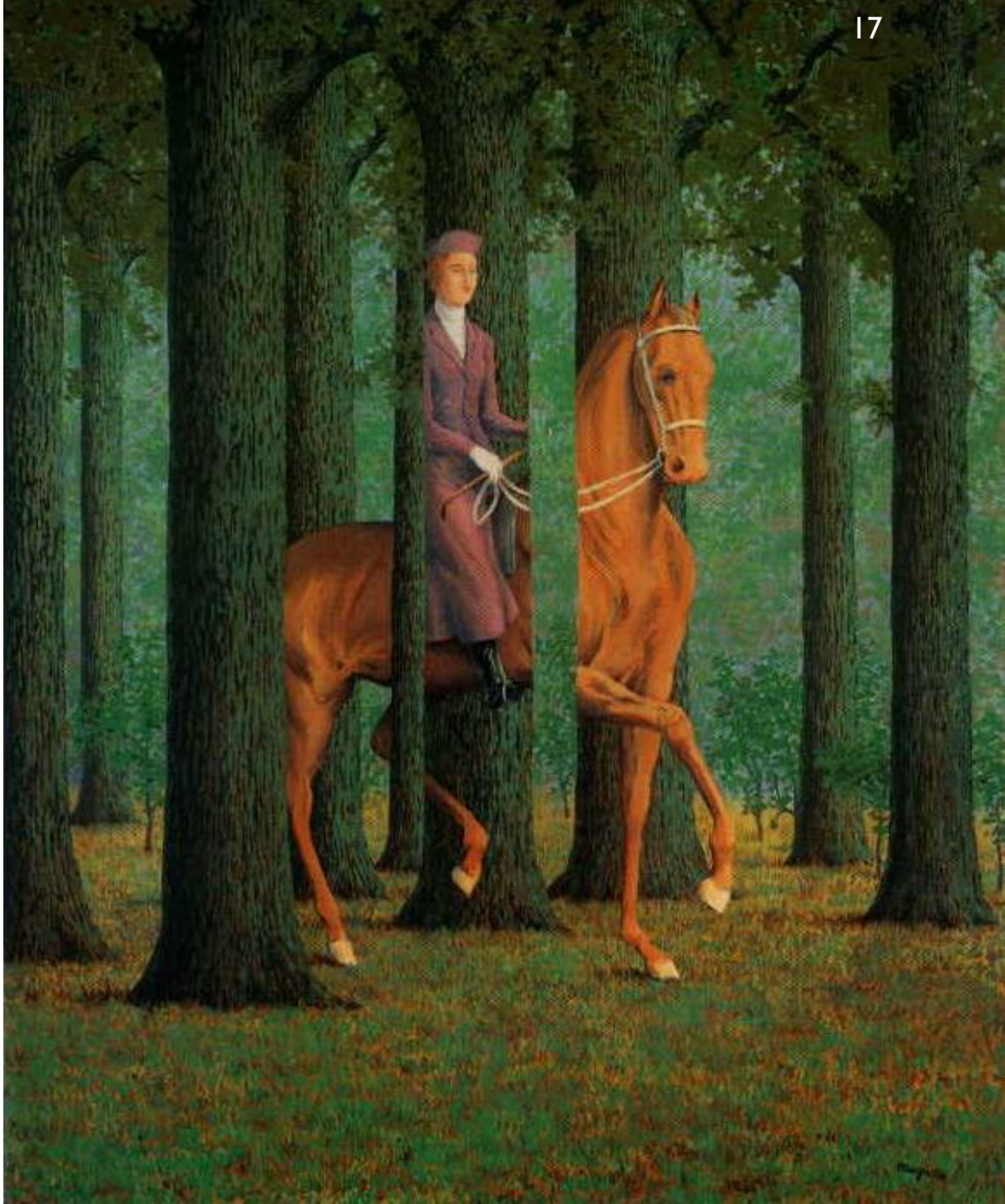
slide credit: Fei-Fei, Fergus & Torralba

Challenges: deformation

Challenges: occlusion



Challenges: occlusion



Magritte, 1957

slide credit: Fei-Fei, Fergus & Torralba

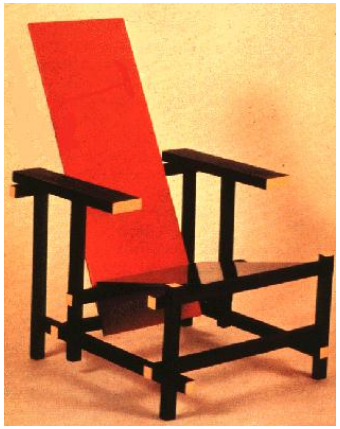
Challenges: background clutter

Challenges: background clutter

Kilmeny Niland. 1995

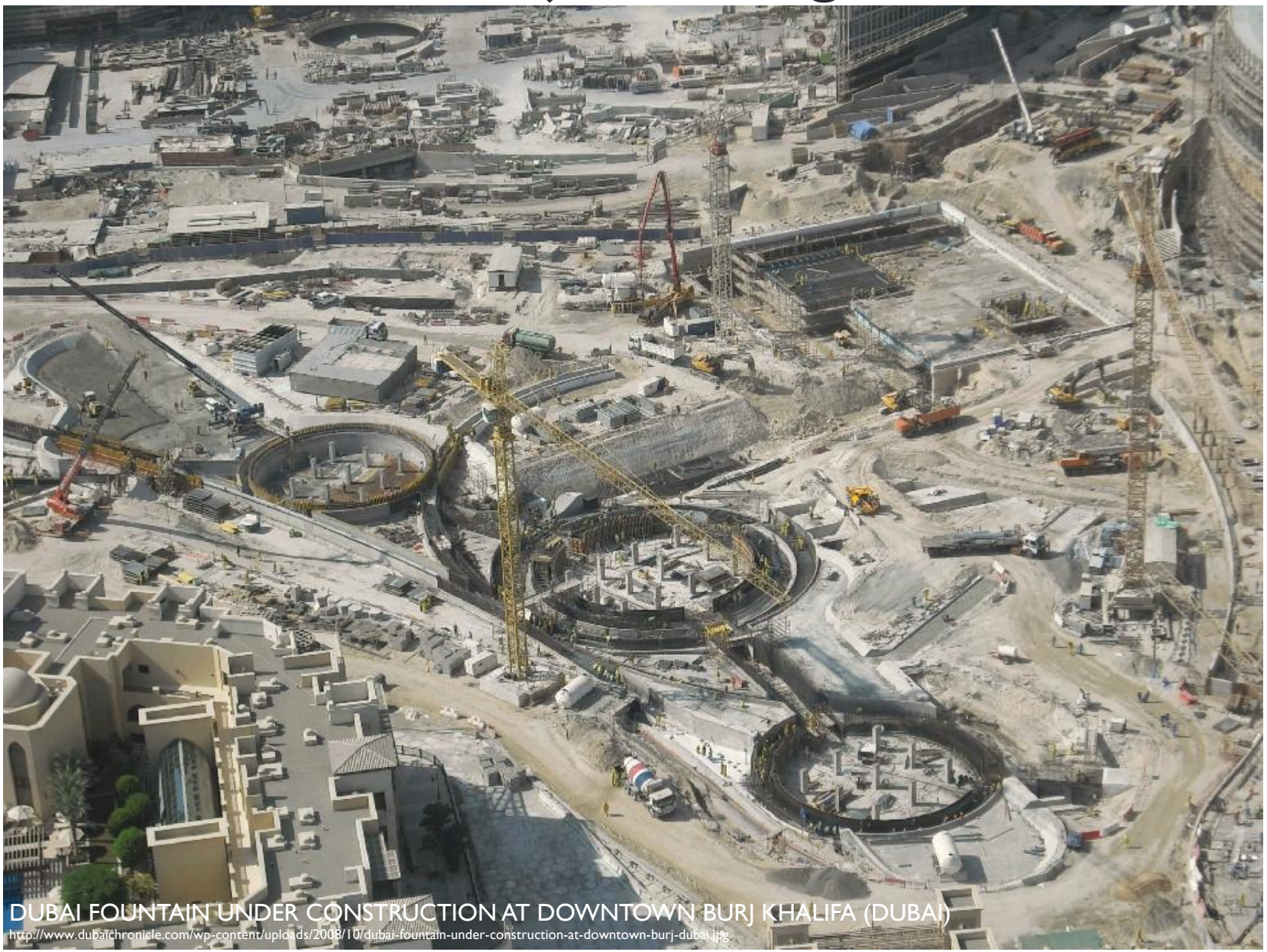
Challenges: object intra-class variation

Challenges: intra-class variation



No intra-class variation: single object recognition

So what does object recognition involve?

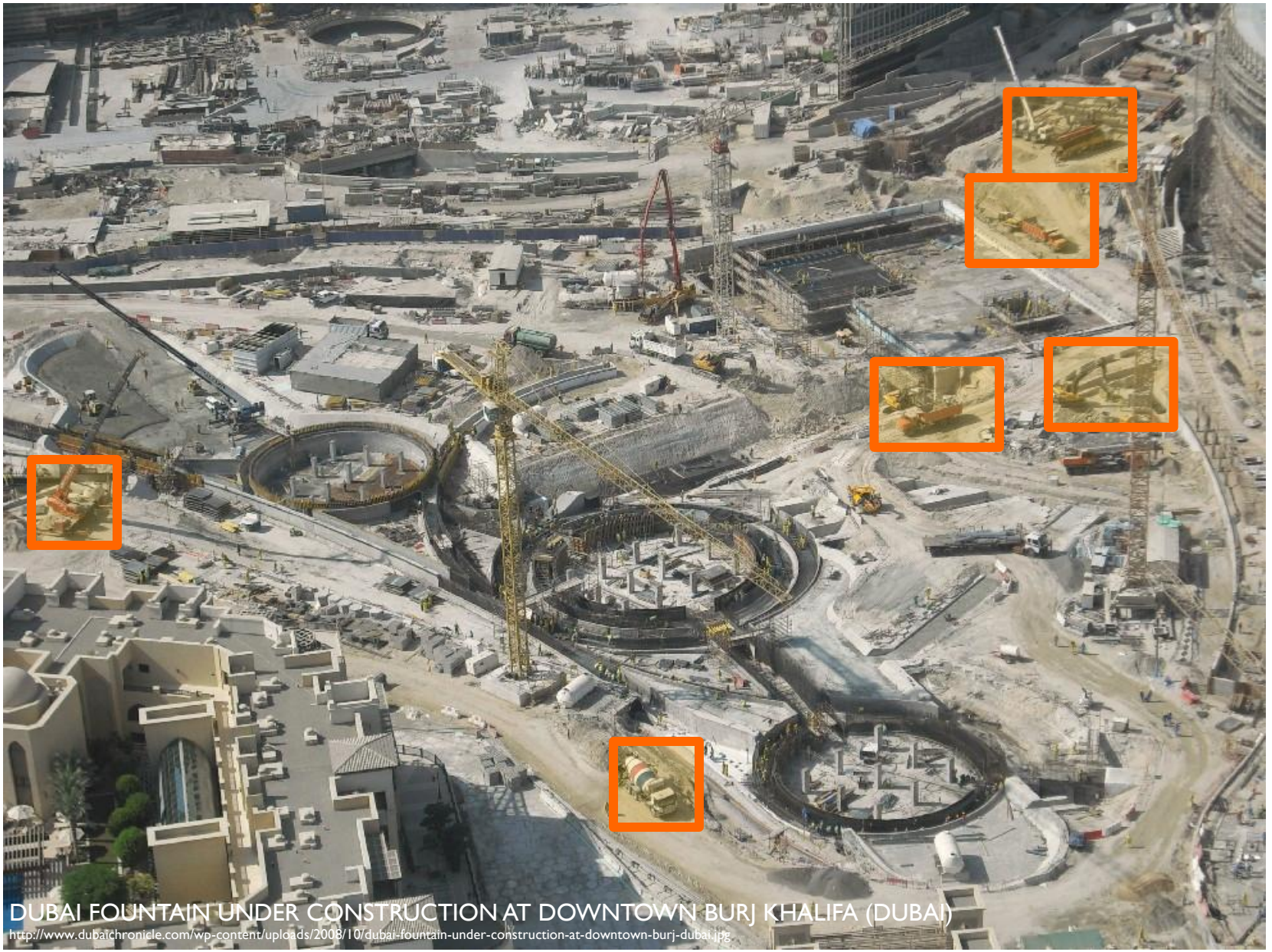


DUBAI FOUNTAIN UNDER CONSTRUCTION AT DOWNTOWN BURJ KHALIFA (DUBAI)
http://www.dubaichronicle.com/wp-content/uploads/2008/10/dubai-fountain-under-construction-at-downtown-burj-dubai.jpg

Identification: does this region contain the Dubai Fountain?

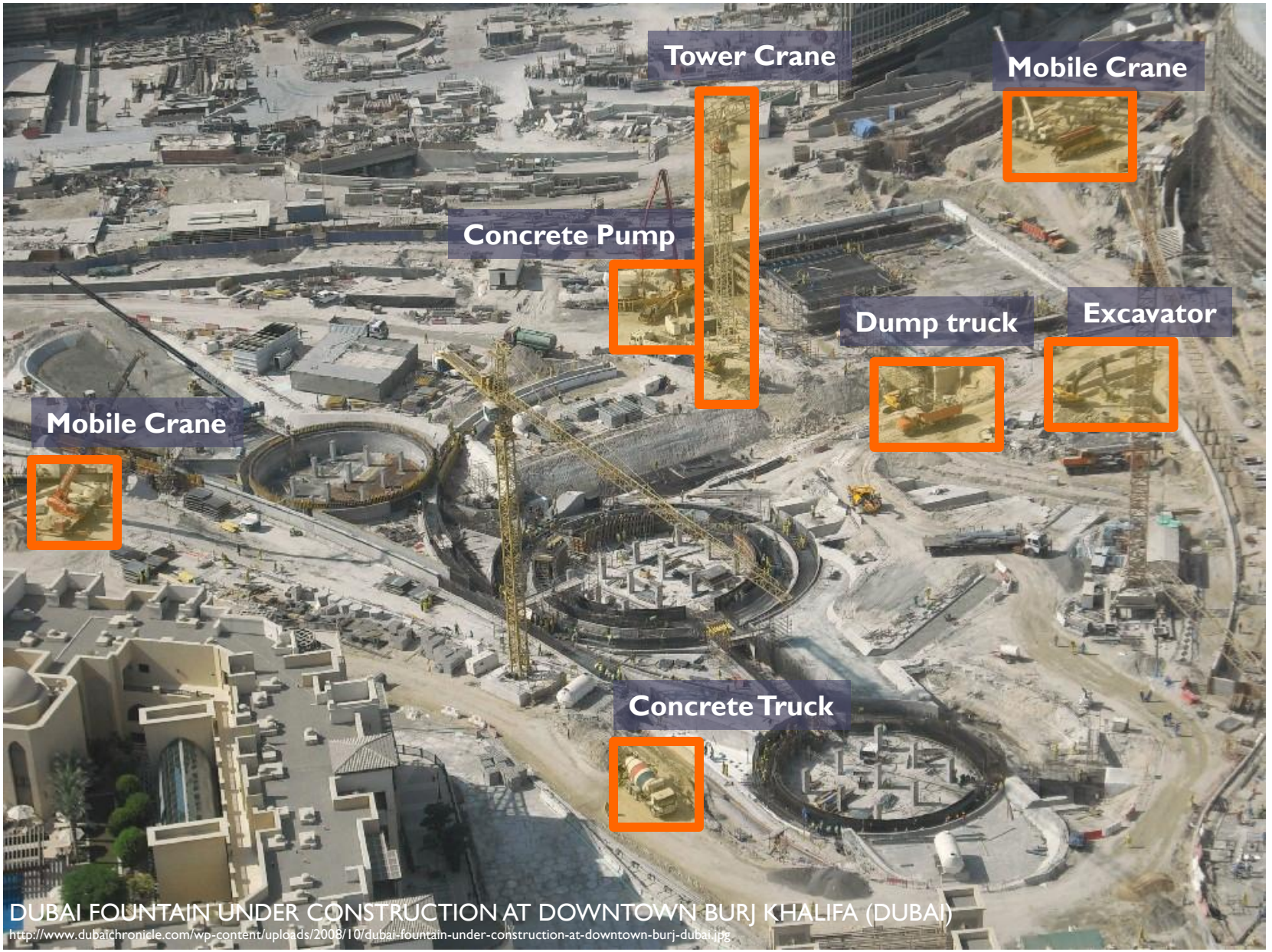
DUBAI FOUNTAIN UNDER CONSTRUCTION AT DOWNTOWN BURJ KHALIFA (DUBAI)
http://www.dubaichronicle.com/wp-content/uploads/2008/10/dubai-fountain-under-construction-at-downtown-burj-dubai.jpg

Object Labeling and Categorization



DUBAI FOUNTAIN UNDER CONSTRUCTION AT DOWNTOWN BURJ KHALIFA (DUBAI)
http://www.dubaichronicle.com/wp-content/uploads/2008/10/dubai-fountain-under-construction-at-downtown-burj-dubai.jpg

Object Labeling and Categorization



DUBAI FOUNTAIN UNDER CONSTRUCTION AT DOWNTOWN BURJ KHALIFA (DUBAI)
http://www.dubaichronicle.com/wp-content/uploads/2008/10/dubai-fountain-under-construction-at-downtown-burj-dubai.jpg

Why Object Recognition?

- Tracking
- Action Recognition
- Events

Some early works on object categorization

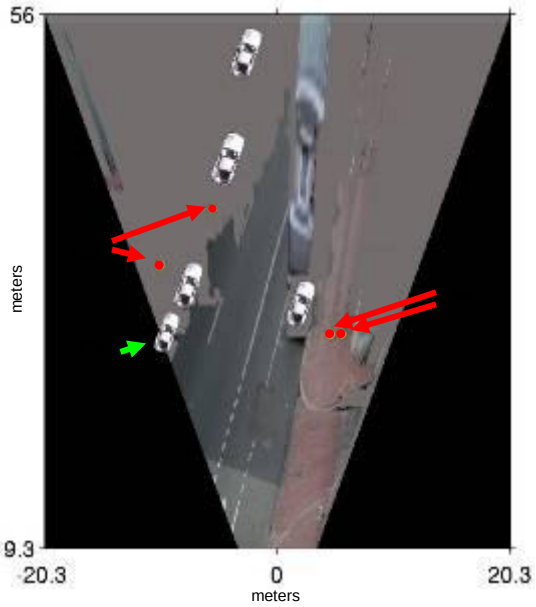
- Turk and Pentland, 1991
- Belhumeur, Hespanha, & Kriegman, 1997
- Schneiderman & Kanade 2004
- Viola and Jones, 2000

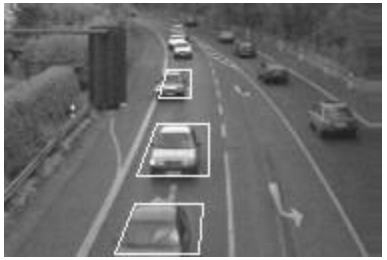
- Amit and Geman, 1999
- LeCun et al. 1998
- Belongie and Malik, 2002

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

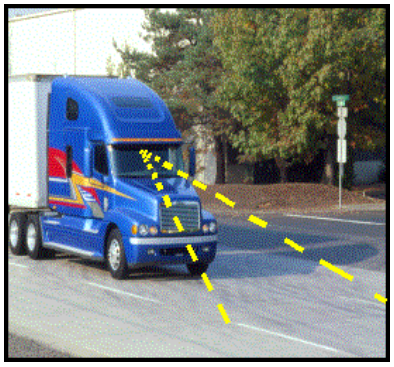
Applications: Assisted driving

Pedestrian and car detection





Lane detection



- Collision warning systems with adaptive cruise control
- Lane departure warning systems
- Rear object detection systems

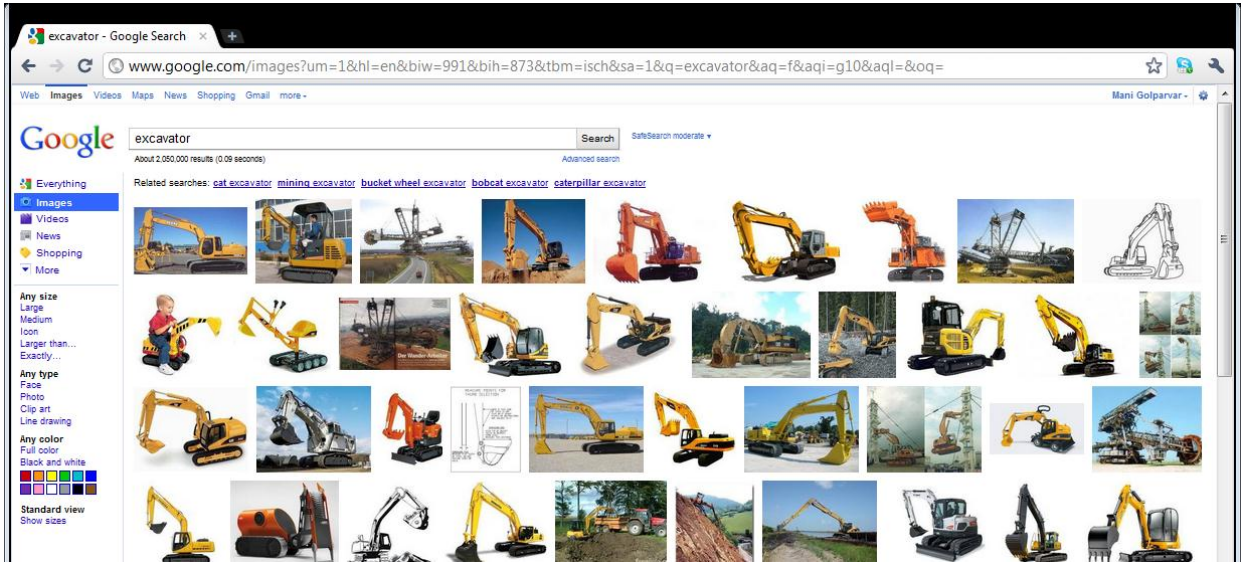
Computational photography

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

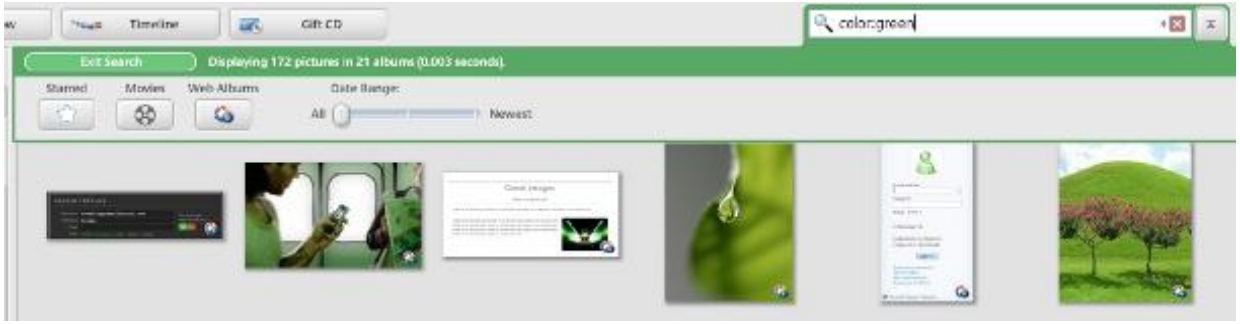
Improving online search

flickr webshots Ask Images Cydral Image & Site Search picsearch Google Image Search altavista

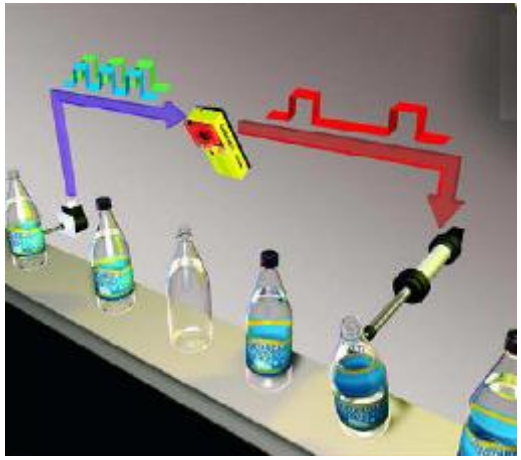
Query:
STREET



Organizing photo collections



Applications of computer vision



Factory inspection

Assistive technologies

Surveillance

Autonomous driving,
robot navigation

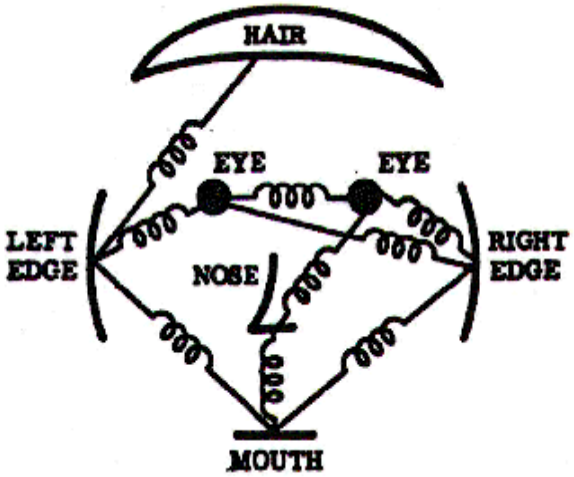
Security

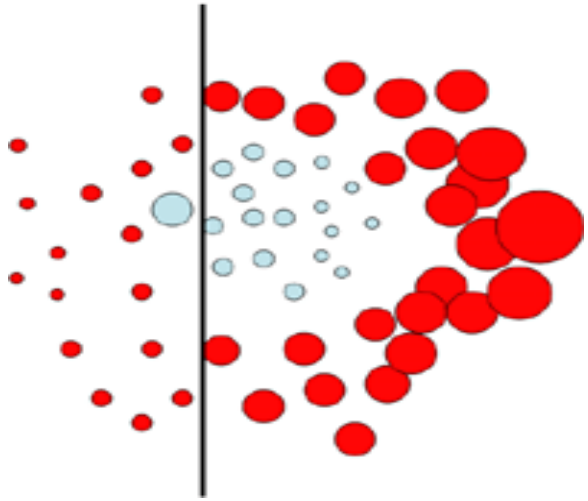
Three main issues

- 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

- Generative / discriminative / hybrid





Object categorization: the statistical viewpoint

$$p(excavator \mid image)$$

vs.

$$p(no\ excavator \mid image)$$

$$\frac{p(excavator \mid image)}{p(no\ excavator \mid image)}$$

- Bayes rule:

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$$

Object categorization: the statistical viewpoint

$$p(excavator \mid image)$$

vs.

$$p(no\ excavator \mid image)$$

- Bayes rule:

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

Object categorization: the statistical viewpoint

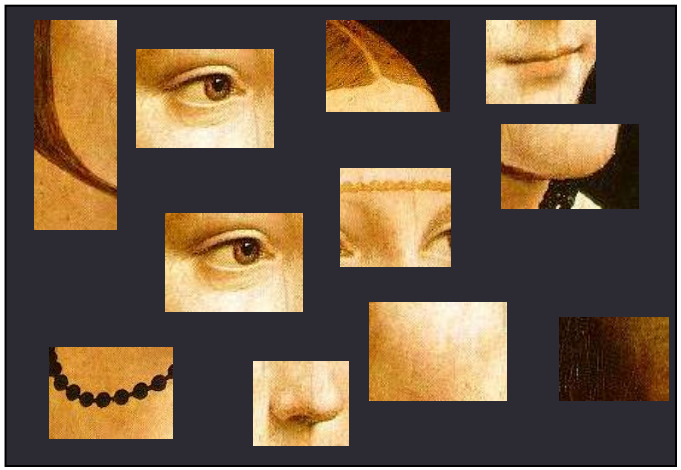
- **Discriminative methods** model posterior
- **Generative methods** model likelihood and prior

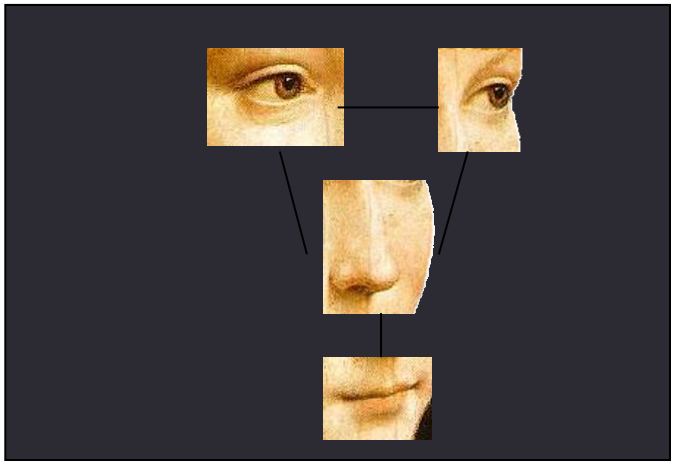
- Bayes rule:

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

Representation

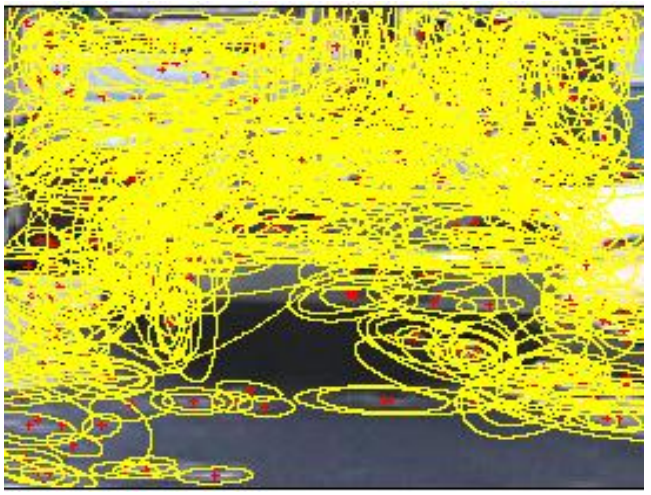
- Generative / discriminative / hybrid
- Appearance only or location and appearance





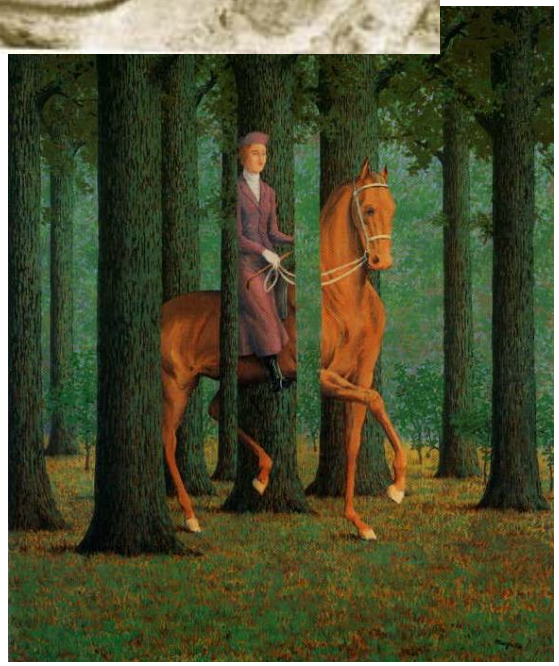
Representation

- Generative / discriminative / hybrid
- Appearance only or location and appearance



Representation

- Generative / discriminative / hybrid
- Appearance only or location and appearance
- Invariances
  - View point
  - Illumination
  - Occlusion
  - Scale
  - Deformation
  - Clutter
  - etc.



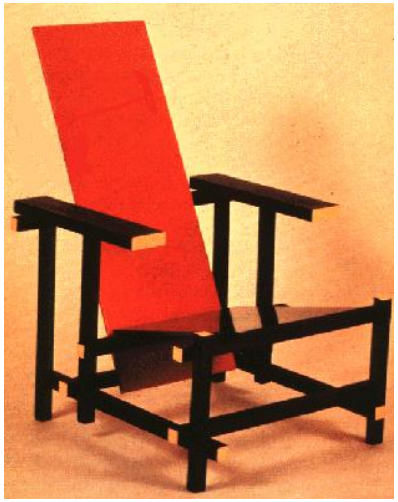
Learning

- Learning model parameters

– Degree of supervision

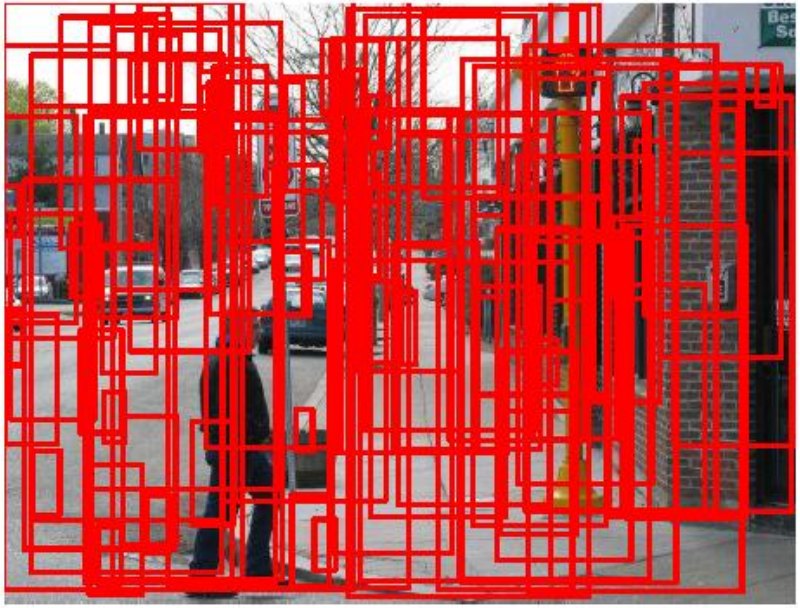
– Batch-vs-online

– etc…



Recognition

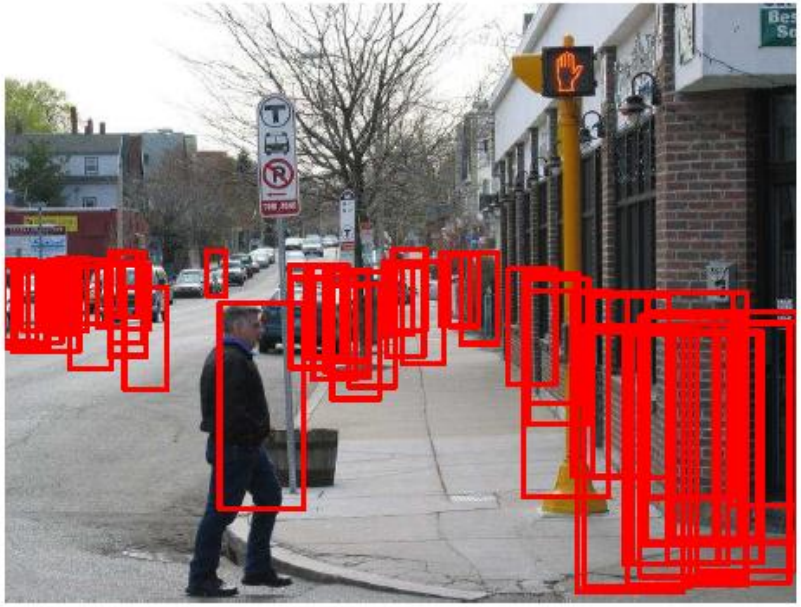
- Scale / orientation range to search over
- Speed
- Context



(b) P(person) = uniform

(d) P(person | geometry)

(f) P(person | viewpoint)



(g) P(person|viewpoint,geometry)

Outline

- Object Recognition

- Introduction
- Recognition of single 3D objects
  - Bag of world models
  - Part based models
  - Models for 3D objects categorization

train

test

Car: front-right

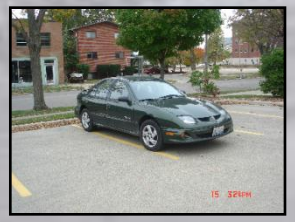
Iron: top-rear-left

•Various level of supervision



















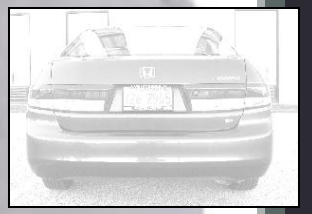
Single 3D object recognition



















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

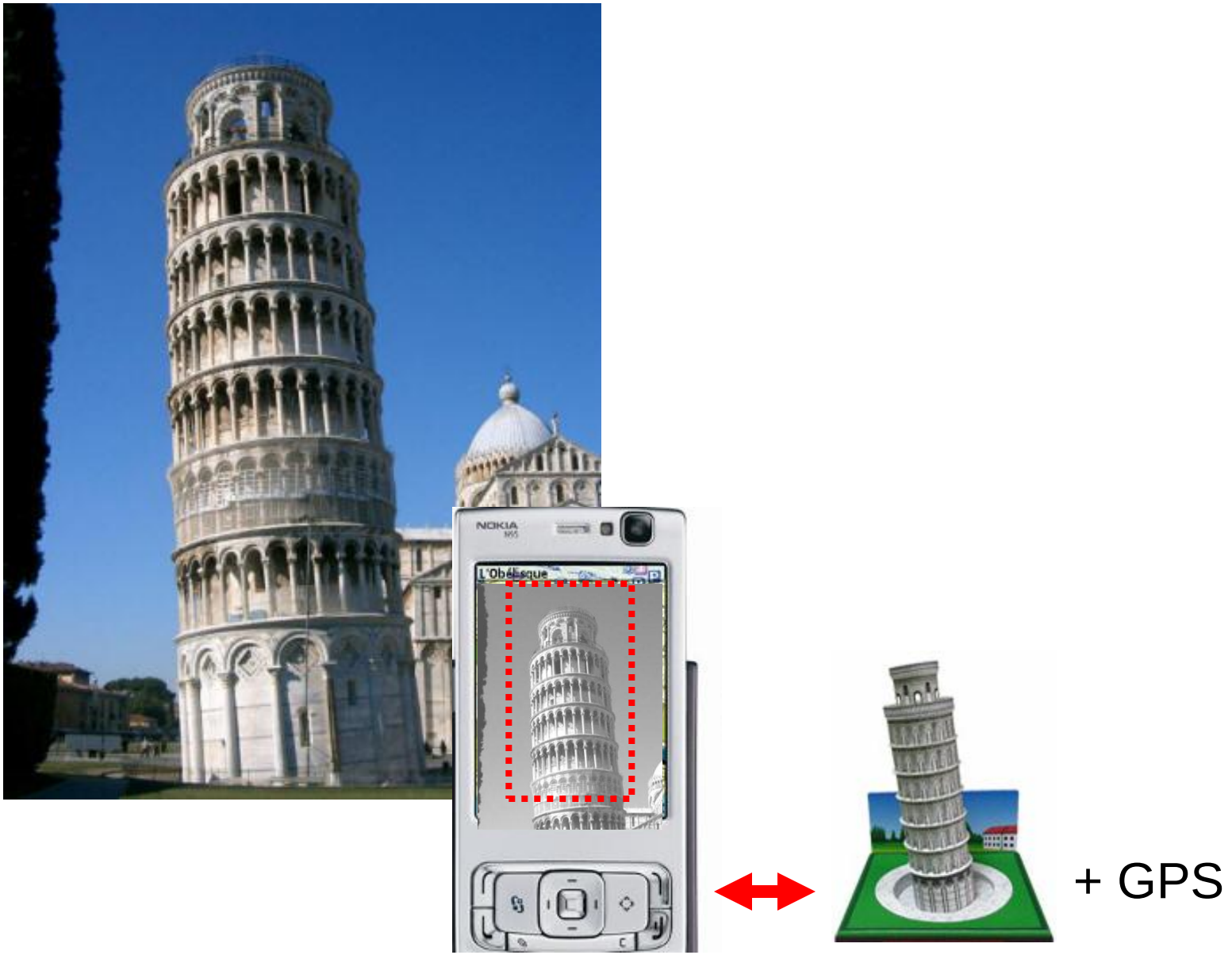
- Edelman et al. '91
- Ullman & Barsi, '91
- Rothwell '92
- Linderberg, '94
- Murase & Nayar '94

- 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**

AIBO® Entertainment Robot

Official U.S. Resources and Online Destinations


NOKIA
L'Obélisque
+ GPS

Usual Challenges

- Variability due to:
  - View point
  - Illumination
  - Occlusions

Basic Scheme

- Representation
  - Features
  - 2D/3D Geometrical constraints
- Model learning
- Recognition
  - Hypothesis generation
  - Validation

1. Rothganger et al. '04, '06
2. Brown et al, '05
3. Lowe '99, '04
4. Ferrari et al. '04, '06
5. Lazebnick et al '04

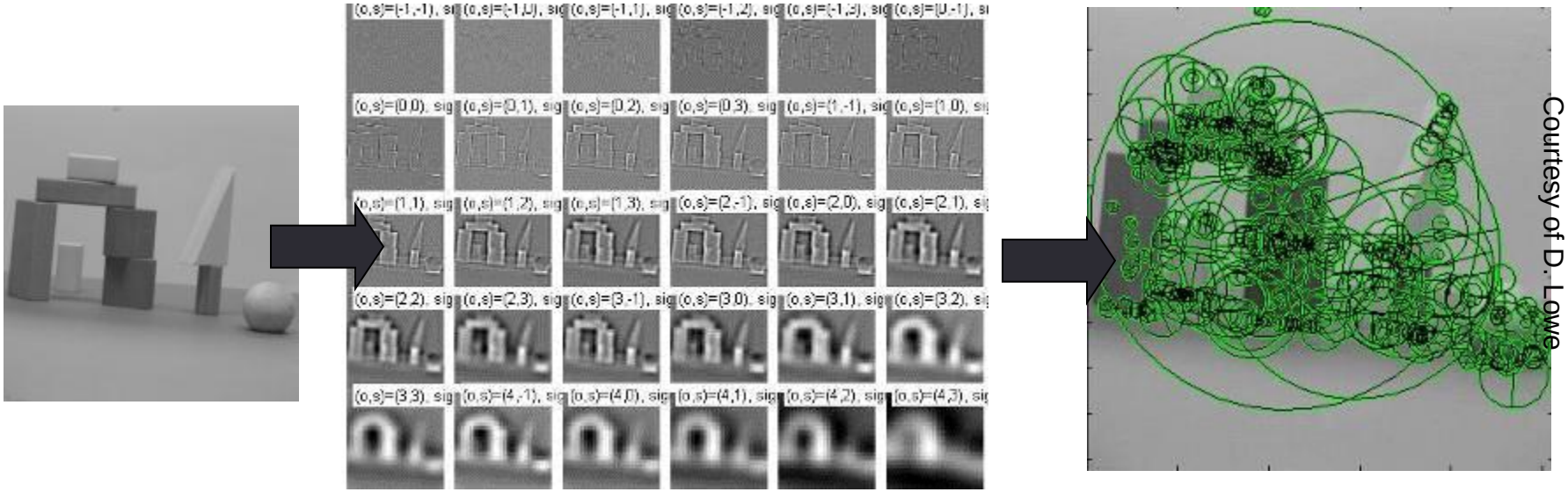
Representation

- 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

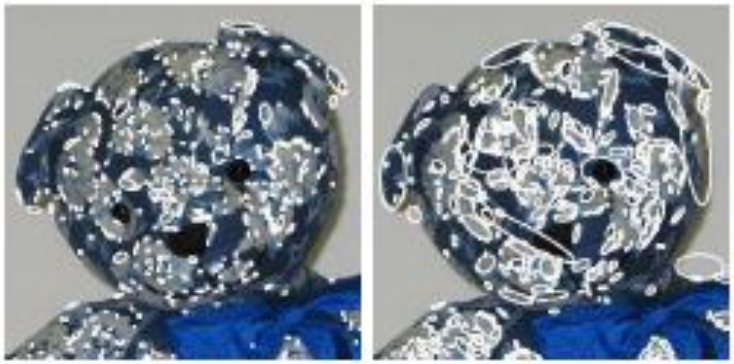
- Geometrical constraints
  - 2D spatial layout of keypoints
  - Tracks of keypoints (regions) across views
  - 3D locations and/or surface normal

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



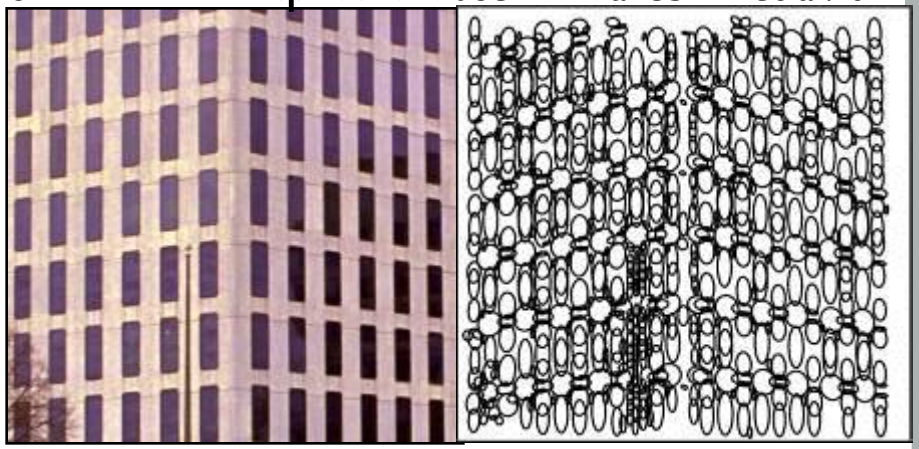
Courtesy of D. Lowe

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



Courtesy of Rothganger et al

Laplacian: used in Lazebnik 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

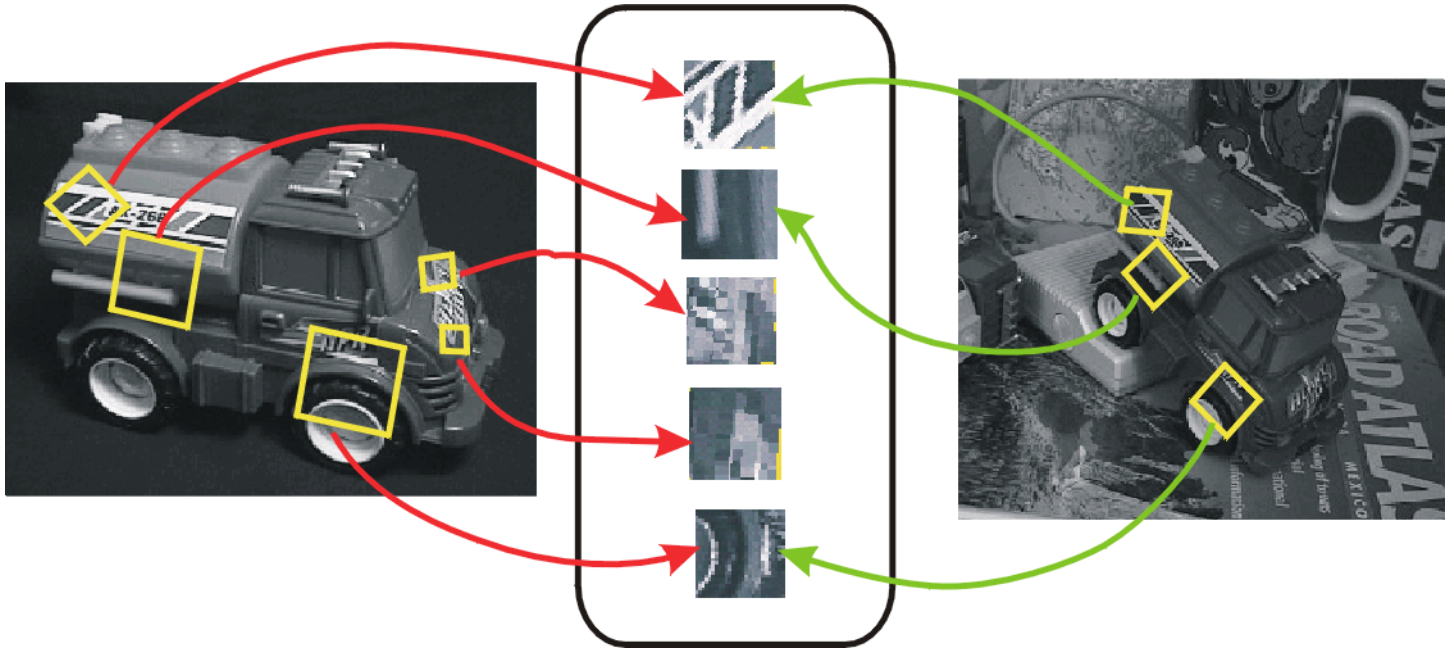
- x,y
- Scale
- Orientation
- Affine structure

- Description
  - SIFT
  - Color histograms

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

Adaptation

- keypoints are transformed in order to be invariant to translation, rotation, scale, and other geometrical parameters



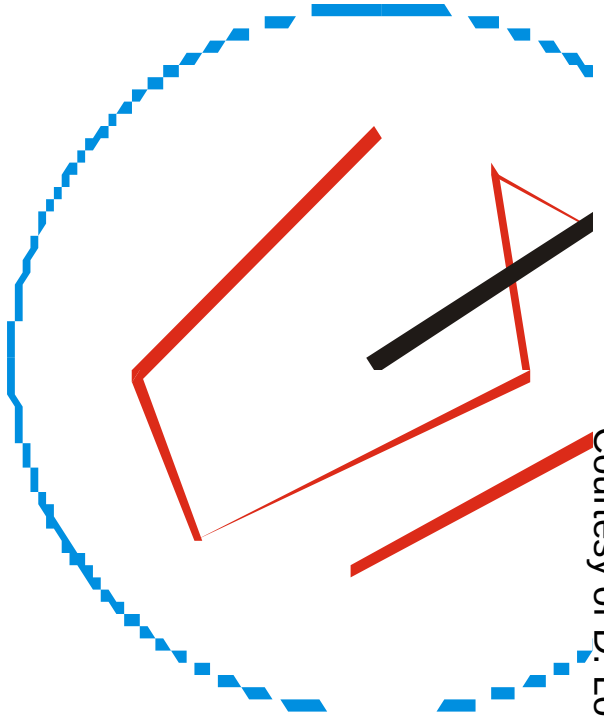
Courtesy of D. Lowe

Change of scale, pose, illumination…

Scale & orientation adaptation

[used in Lowe '99]

detector → • x,y
•Scale
•Orientation



- SIFT: Create histogram of local gradient directions computed at selected scale
- Assign canonical orientation at peak of smoothed histogram

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

- x,y
- Scale
- Orientation
- Affine structure

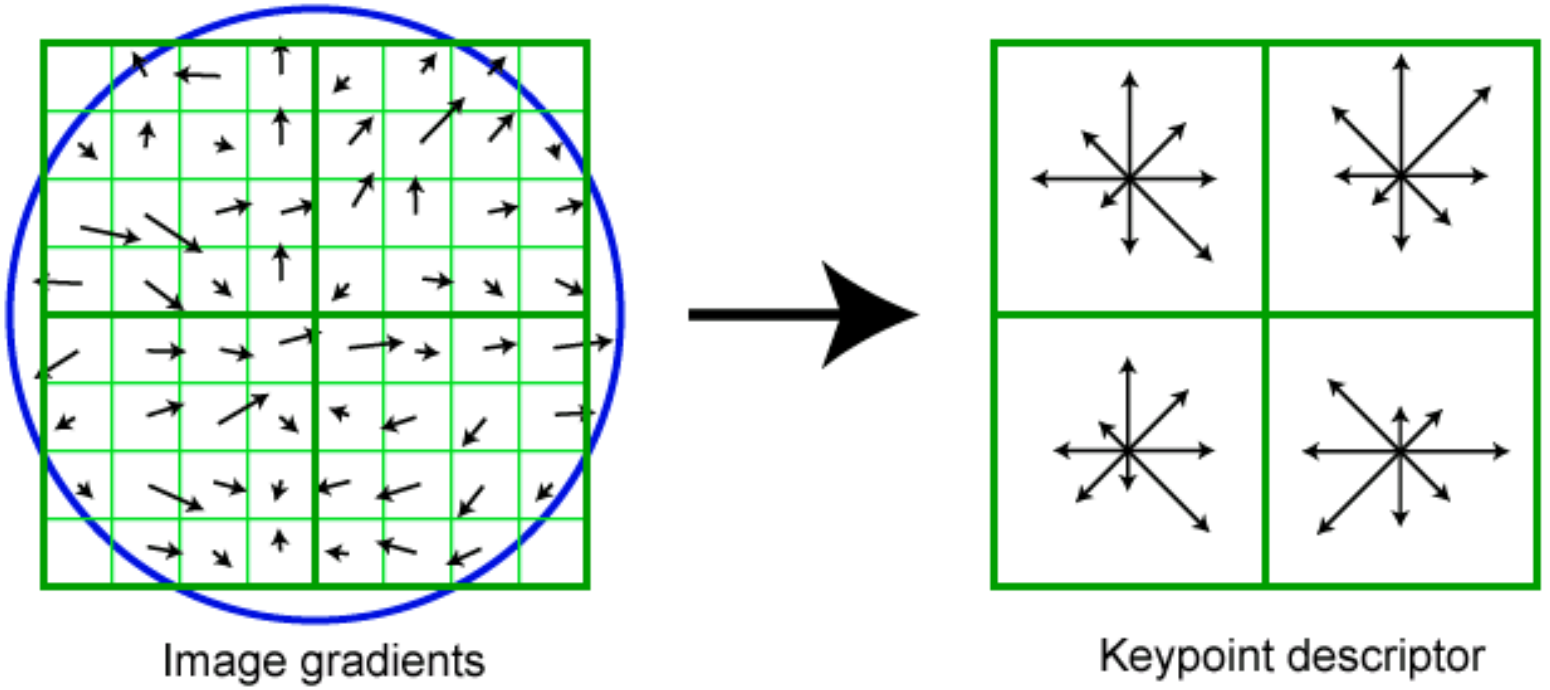
- Description
  - SIFT
  - Color histograms

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

Keypoint description

[Lowe '99]

- Thresholded image gradients are sampled over 16x16 array of locations in scale space
- Create array of orientation histograms
- 8 orientations x 4x4 histogram array = 128 dimensions



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

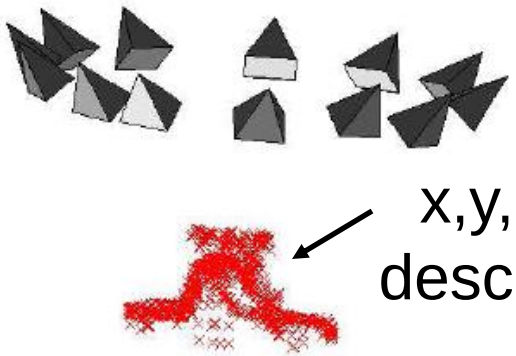
- x,y
- Scale
- Orientation
- Affine structure

- Description
  - SIFT
  - Color histograms

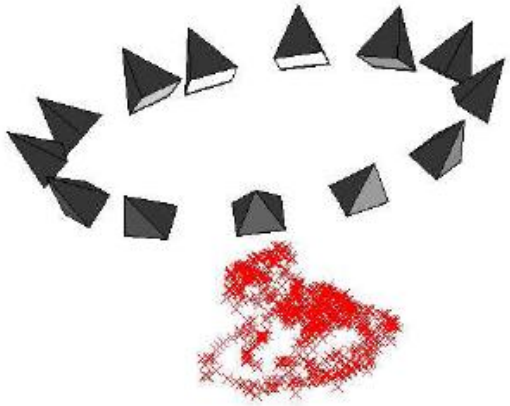
- Geometrical constraints
  - 2D spatial layout of keypoints
  - Tracks of keypoints (regions) across views
  - 3D locations and/or surface normal

Object representation: 3D location of key points

Brown & Lowe '05

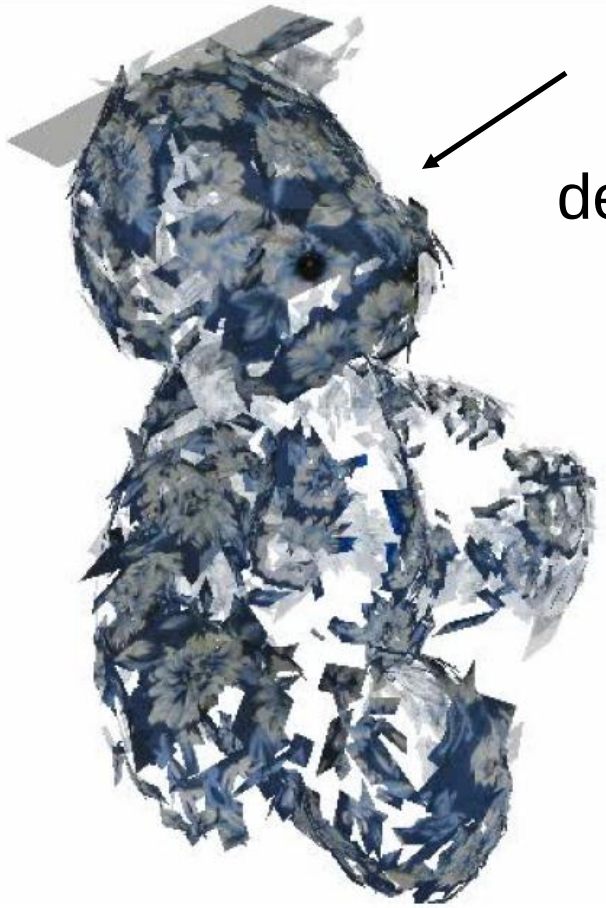


x,y,z + descriptor



Courtesy of Brown & Lowe

Rothganger et al. '06

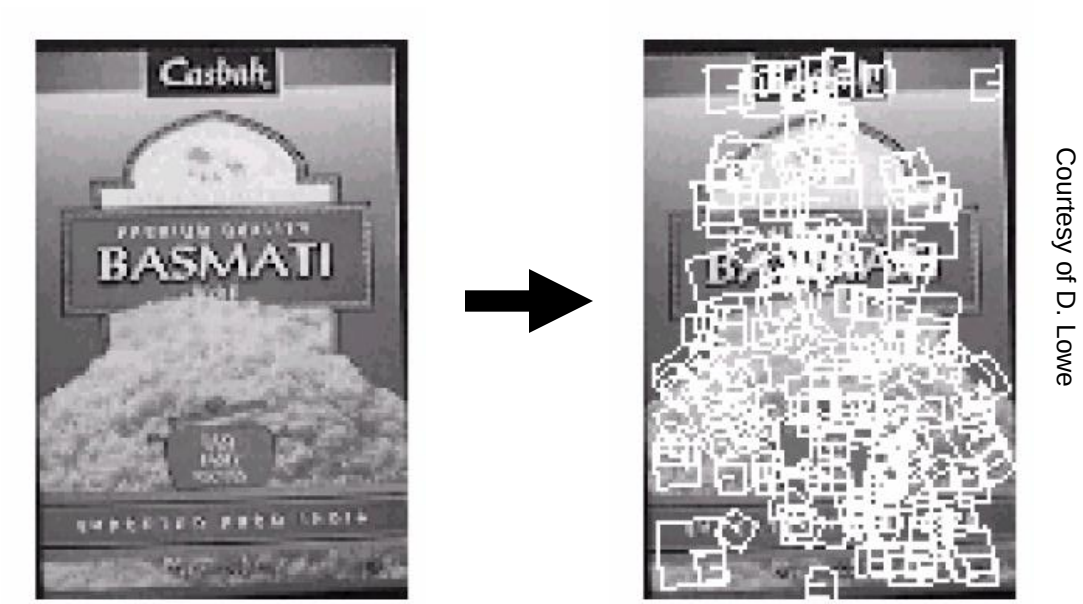


x,y,z + h,v + descriptor

Courtesy of Rothganger et al.

Object representation

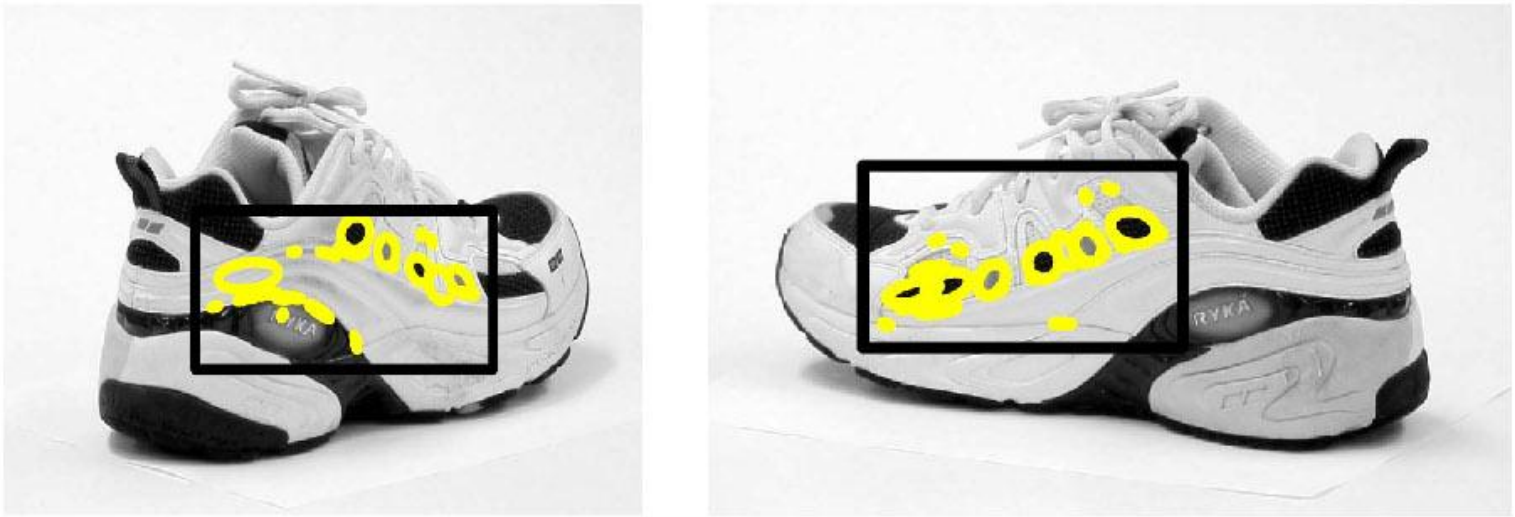
- 2D layout of key points [Lowe '99]



Courtesy of D. Lowe

Object representation: Collections of semi-local affine parts

[Lazebnick et al '04]



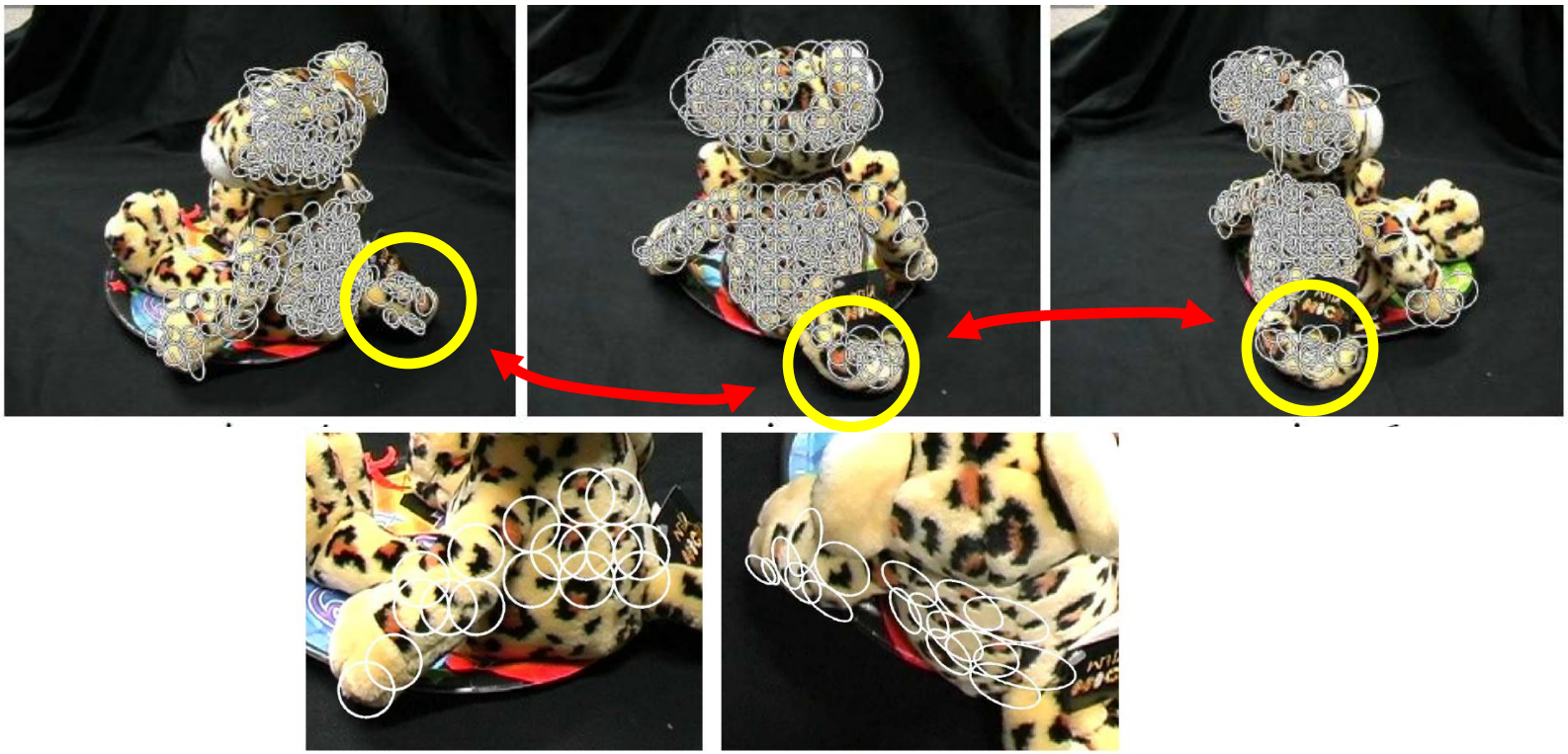
Courtesy of Lazebnick et al

Parts = group of key points that share 'consistent' affine configuration across views

The learning technique establishes the meaning of consistency

Object representation: Collections of GAMs and tracks

[Ferrari et al '04]



Courtesy of Ferrari et al

GAM (Group of Aggregated Matches): "A set of key point matches between two images, which are distributed over a smooth connected surface of the object"

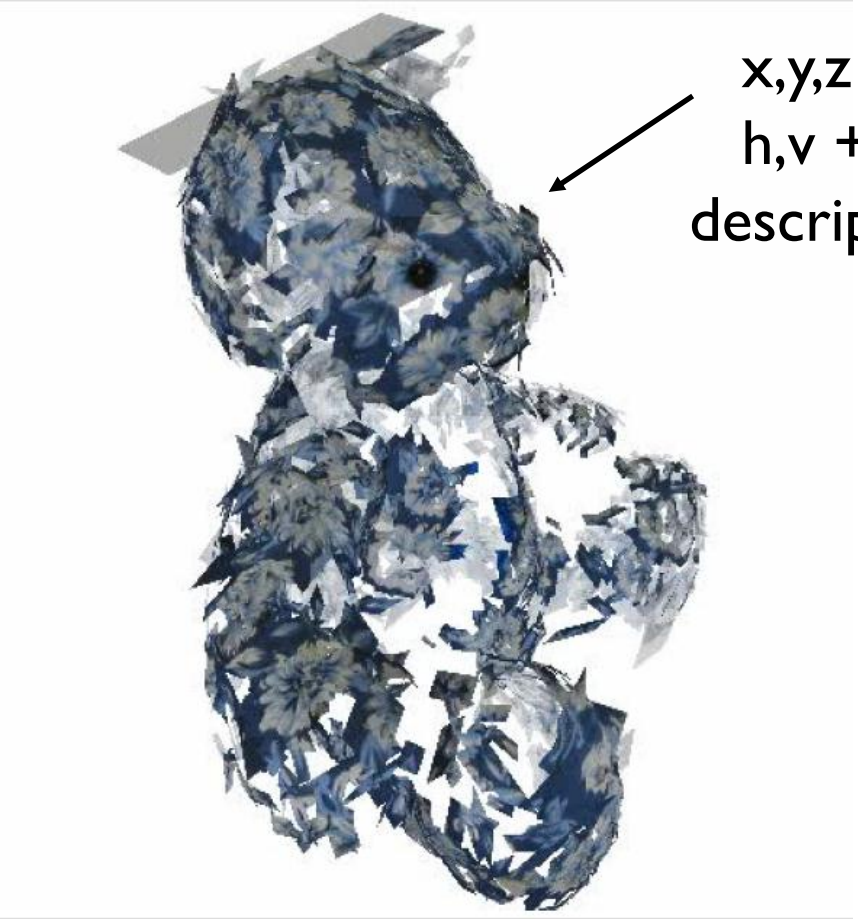
See also: Tuytelaars & Van Gool, 2004

Basic Scheme

- Representation
  - Features
  - 2D/3D Geometrical constraints
- Model learning
- Recognition
  - Hypothesis generation
  - Validation

Model learning

Rothganger et al. '03 '06

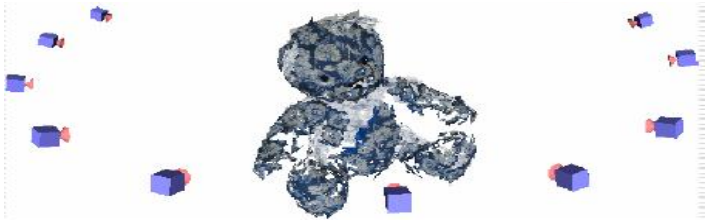


Courtesy of Rothganger et al

Model learning

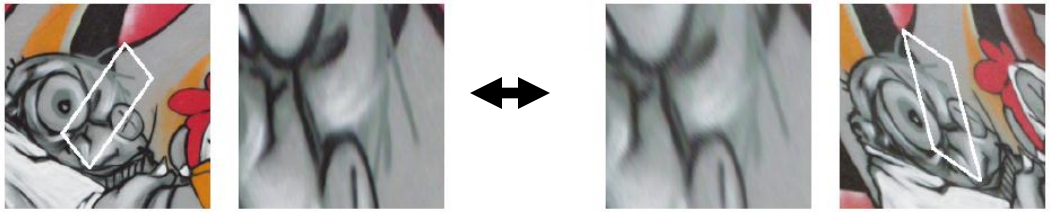
Rothganger et al. '03 '06

Build a 3D model:



- N images of object from N different view points

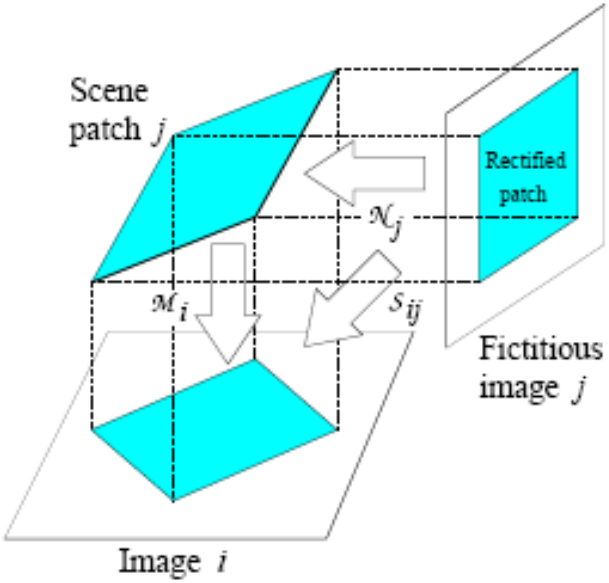
- Match key points between consecutive views
[create sample set]

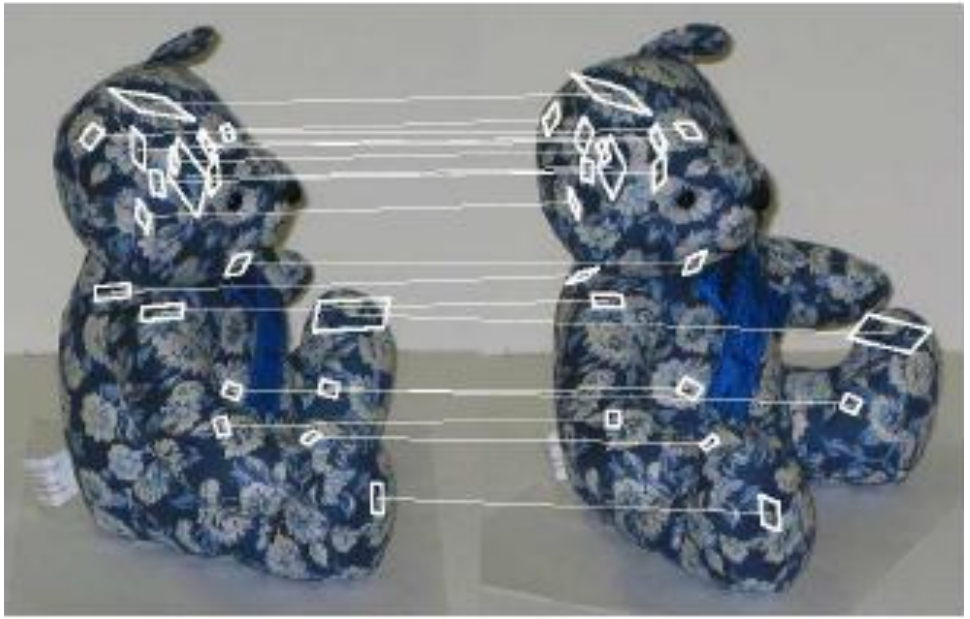


- Use affine structure from motion to compute 3D location and orientation + camera locations

- Affine factorization Tomasi & Kanade '92
- RANSAC
- 2 matches are needed rather than 4 thanks to affine invariant patches

RANSAC Rothganger et al. '03 '06





Courtesy of Rothganger et al

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

$$\mathcal{N}_j = \begin{bmatrix} \boldsymbol{H}_j & \boldsymbol{V}_j & \boldsymbol{C}_j \\ 0 & 0 & 1 \end{bmatrix}$$

Algorithm:

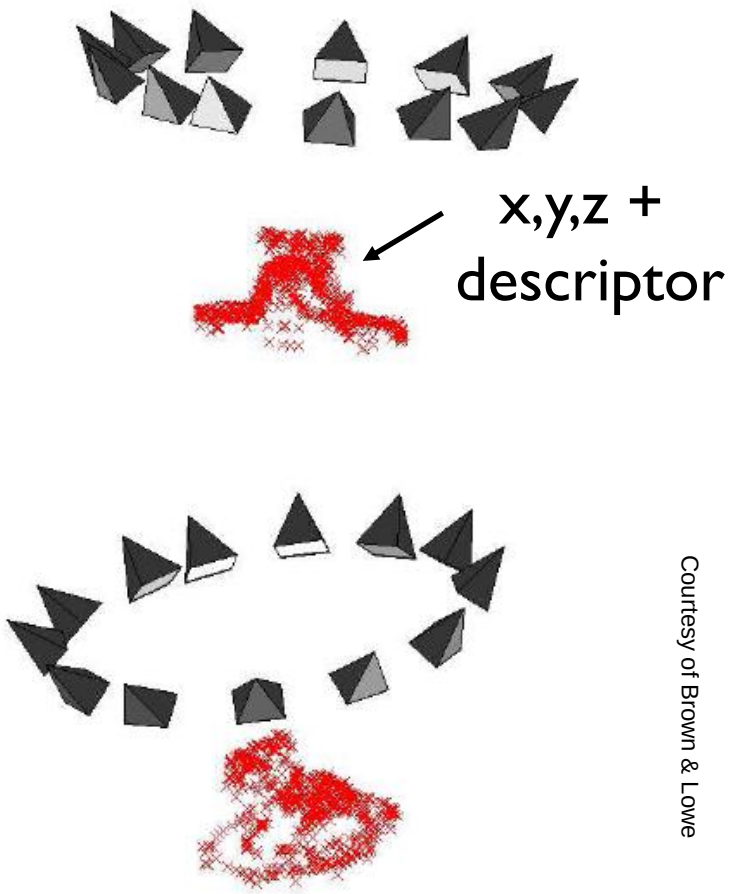
[Affine factorization Tomasi & Kanade '92]

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

Model learning

[Brown & Lowe '05]



Courtesy of Brown & Lowe

Model learning

[Brown & Lowe '05]

- N images of an object from N different view points

- Match key points between view pairs [create sample set]

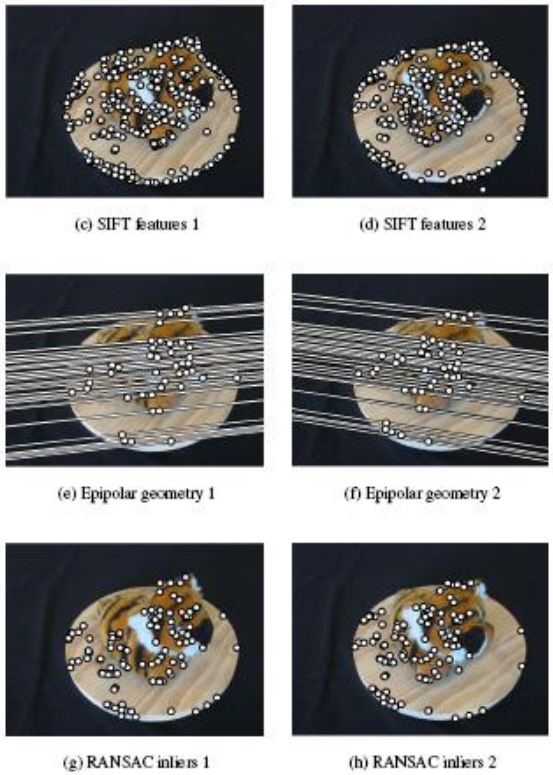
- Computer fundamental matrix F between view pair [RANSAC]

→ Full perspective model! (not affine)

- Find connected components across views

- Bundle adjustment & Metric upgrade

→ Photosynth [Snavely et al.'06]

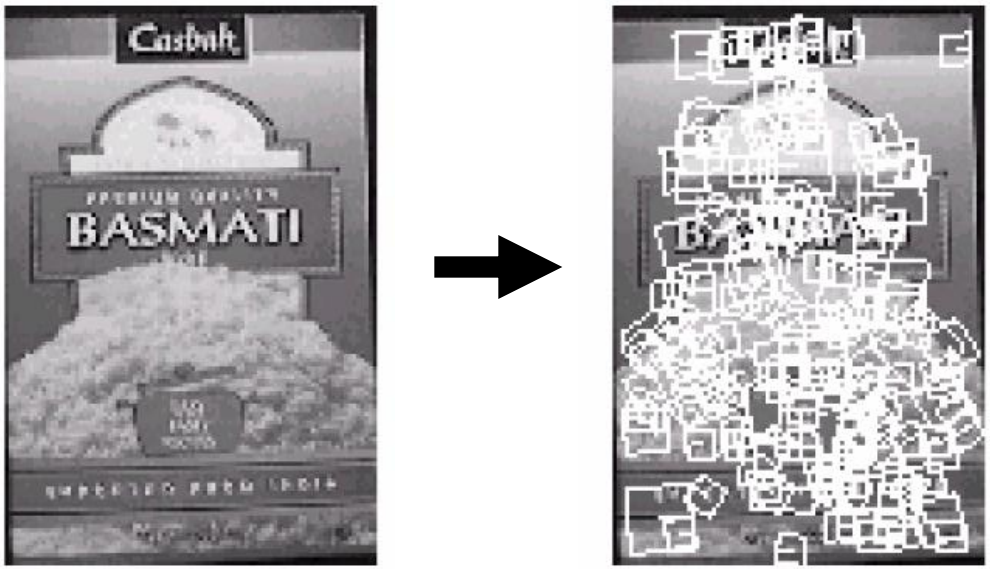


Courtesy of Brown & Lowe

Model learning

- **The model is a template of 2D layout of key points**

[Lowe '99]



Model learning

[Lazebnick et al '04]

- **Collections of semi-local affine parts**

Courtesy of Lazebnick et al

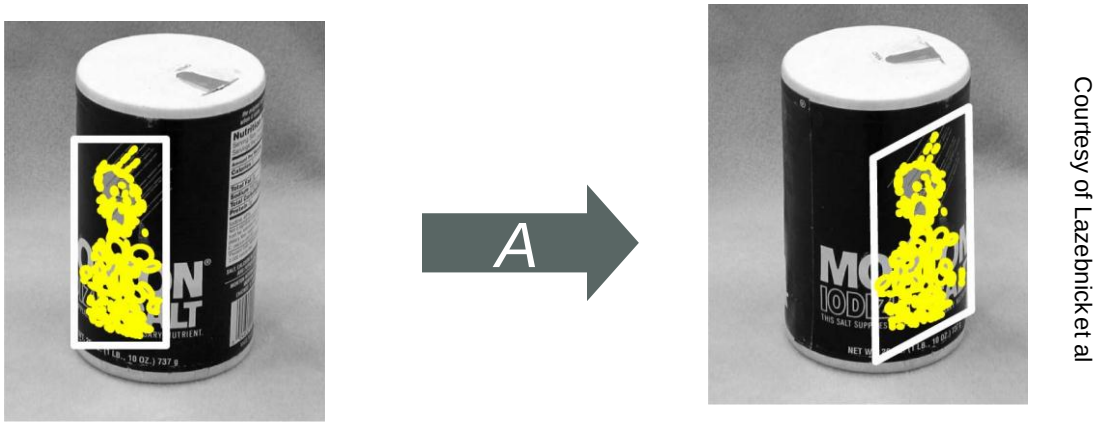
Parts = group of key points that share 'consistent' affine configuration across views

Model learning

[Lazebnick et al '04]

- **Collections of semi-local affine parts**

Goal: to find collections of local affine regions that can be mapped onto each other using a single affine transformation



Courtesy of Lazebnick et al

•Implementation: greedy search based on geometric and photometric consistency constraints

–Returns multiple correspondence hypotheses

–Automatically determines number of regions in correspondence

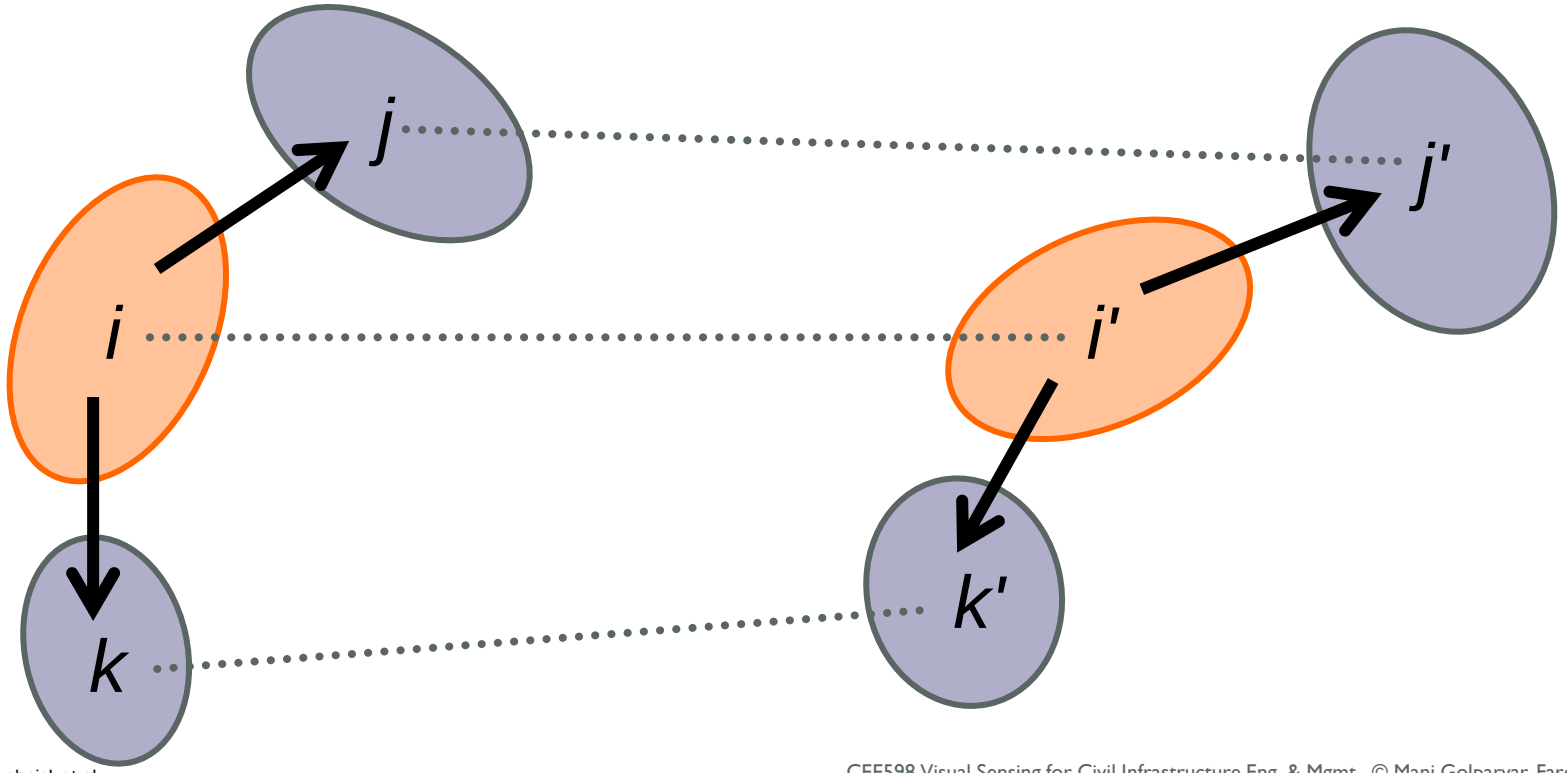
Model learning

- Collections of semi-local affine parts

Initialization:

- Identify triples of neighboring regions (i, j, k) in first image

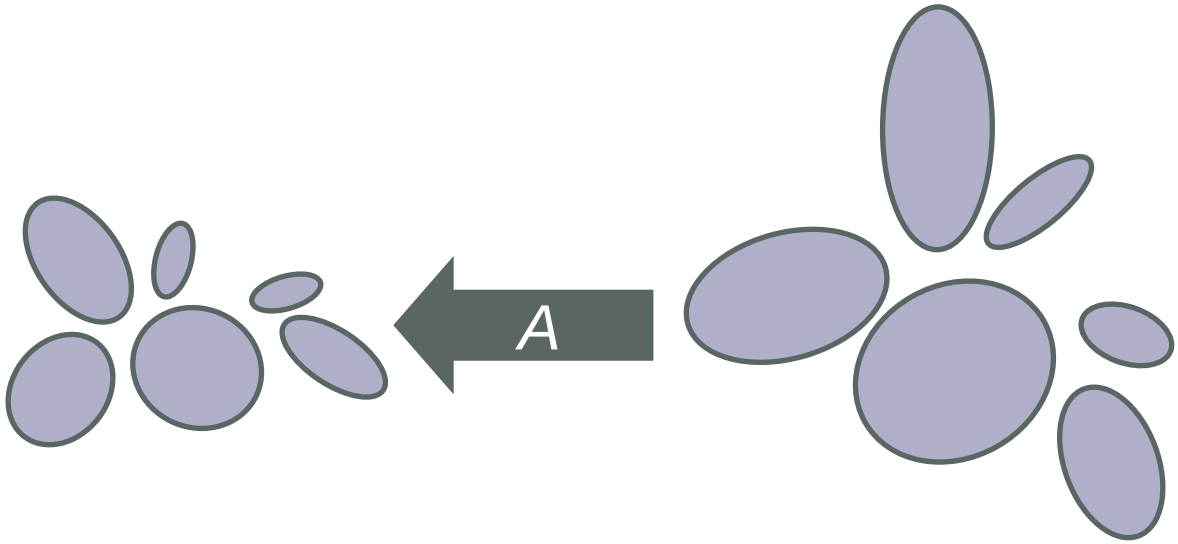
–Find all triples (i', j', k') in the second image such that i' (resp. j', k') is a potential match of i (resp. j, k), and j', k' are neighbors of i'



Courtesy of Lazebnick et al

Model learning

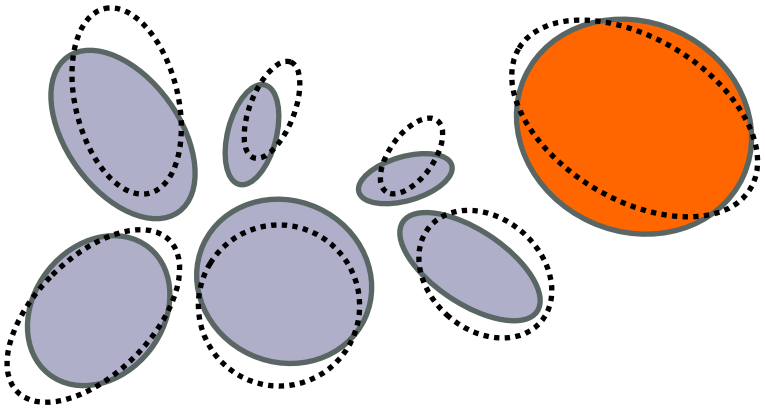
- Collections of semi-local affine parts
- Beginning with each seed triple, iterate:
  - Estimate the affine transformation between centers of corresponding regions in current group of matches



Courtesy of Lazebnick et al

Model learning

- Collections of semi-local affine parts
- Beginning with each seed triple, iterate:
  - Estimate the affine transformation between centers of corresponding regions in current group of matches
  - Determine geometric consistency of current group of matches
  - Search for additional matches in the neighborhood of the current group

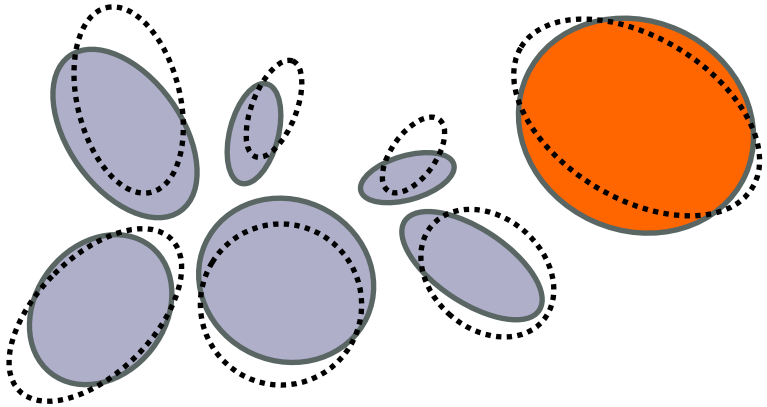


- Geometric consistency criteria:
  - Distance between ellipse centers (residual)
  - Difference of major and minor axis lengths
  - Difference of ellipse orientations

Model learning

- Collections of semi-local affine parts

- Beginning with each seed triple, iterate:
  - Estimate the affine transformation between centers of corresponding regions in current group of matches
  - Determine geometric consistency of current group of matches
  - Search for additional matches in the neighborhood of the current group



- Stop when residual error is bigger than threshold

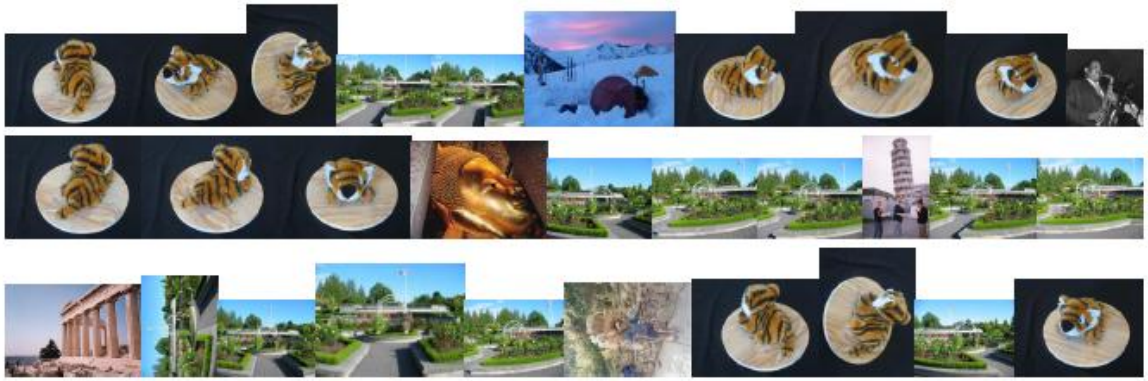
Learnt models

Rothganger et al. '03 '06

Courtesy of Rothganger et al

Learnt models

Brown & Lowe '05

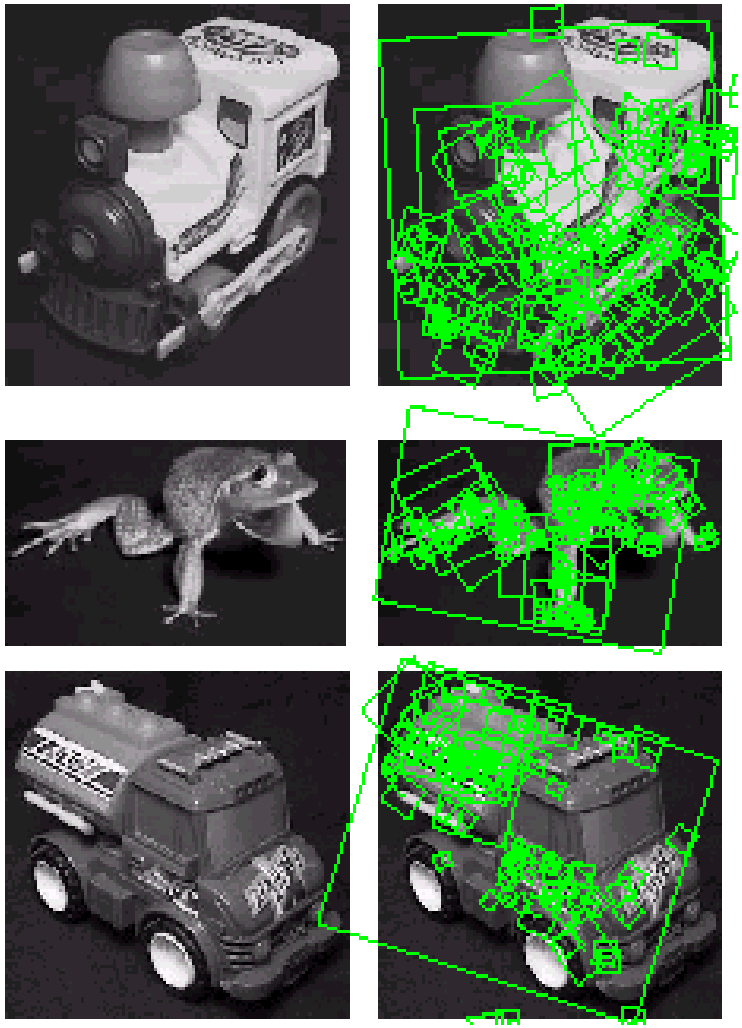


(b) Output 3D model 1 - Tiger

Courtesy of Brown & Lowe

Learnt models

[Lowe '99]



Learnt models

[Lazebnick et al '04]

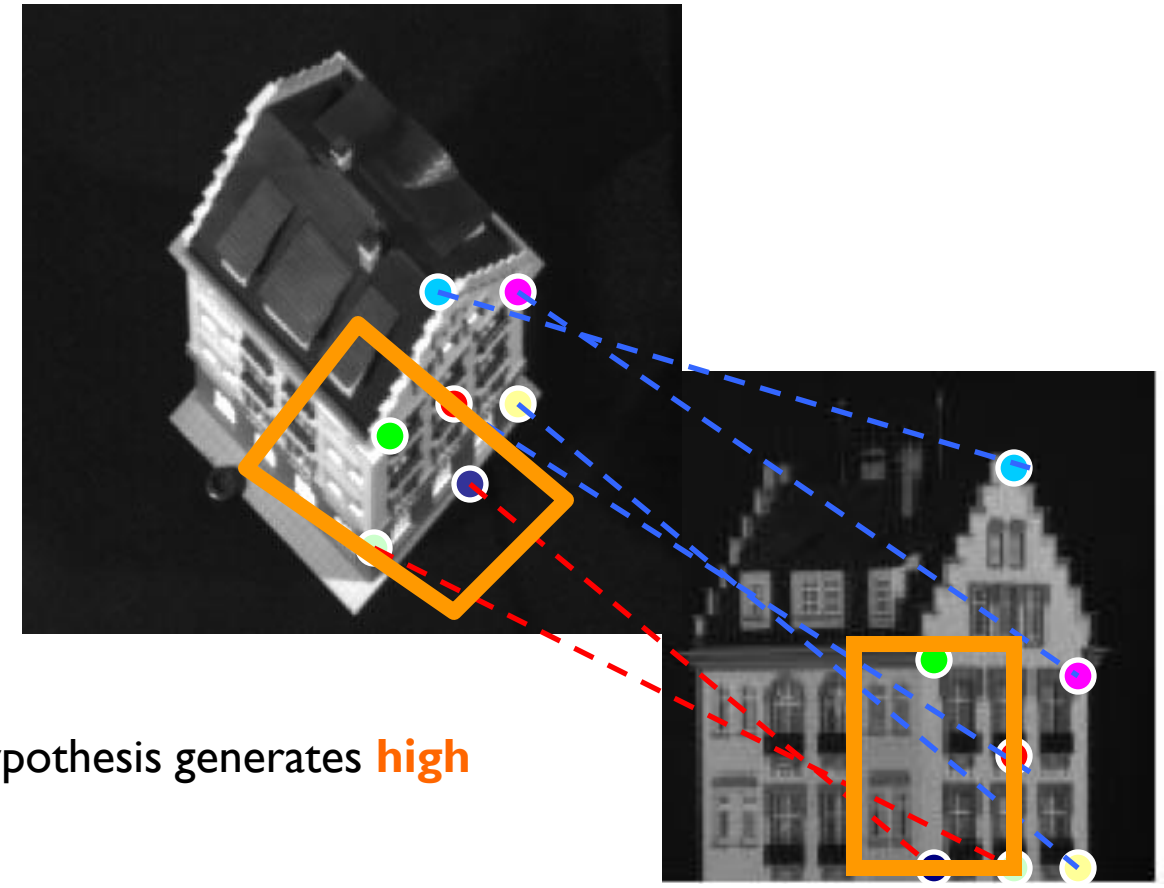
[Ferrari et al '06]

Basic Scheme

- Representation
  - Features
  - 2D/3D Geometrical constraints
- Model learning
- Recognition
  - Hypothesis generation
  - Validation

Recognition

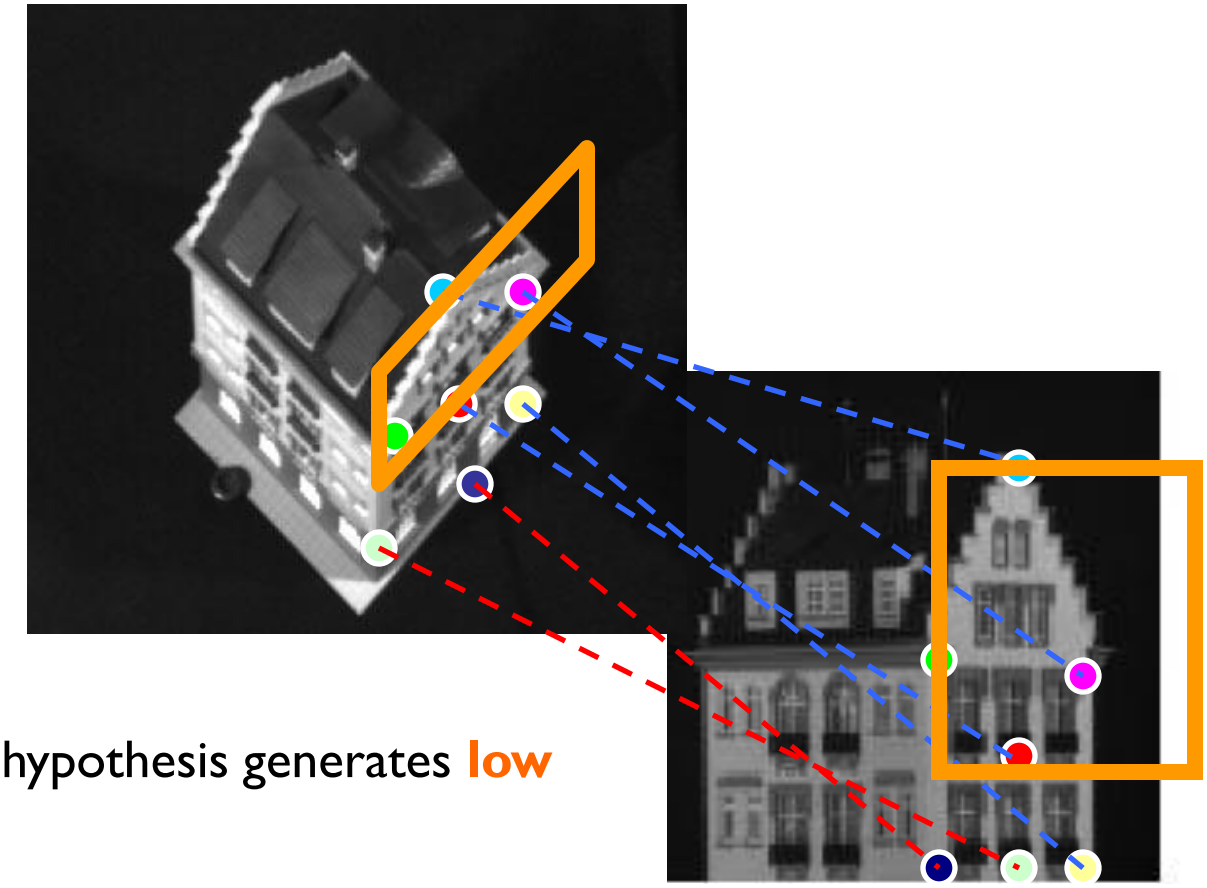
- Hypothesis generation & model verification
- **Basic idea**



Verification: The hypothesis generates **high fitting error**

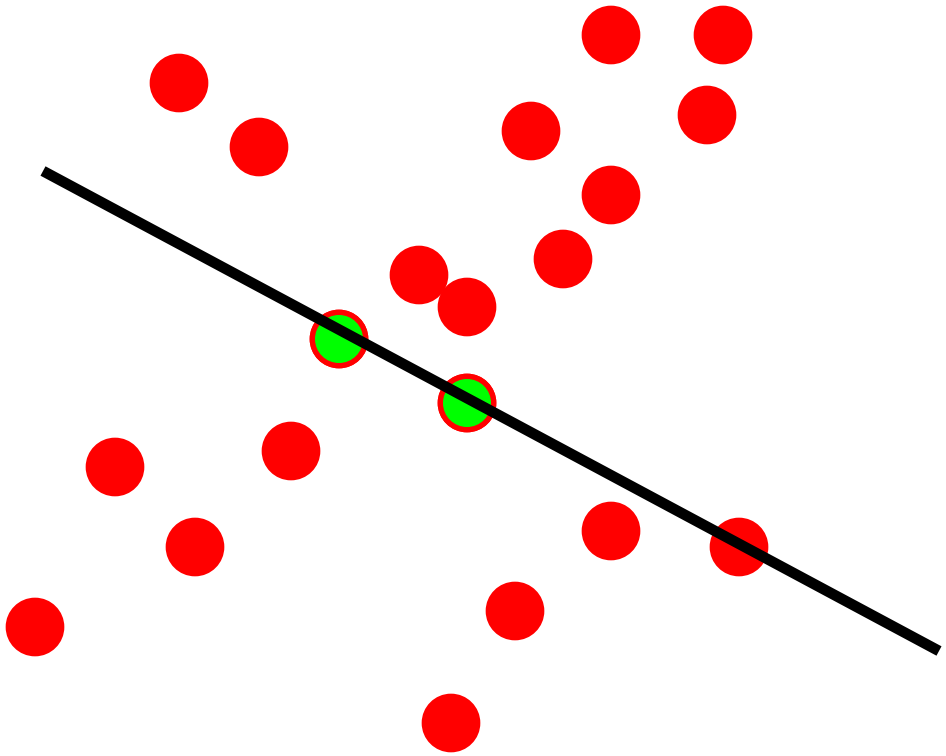
Recognition

- Hypothesis generation & model verification
- **Basic idea**



Verification: The hypothesis generates **low fitting error**

RANSAC



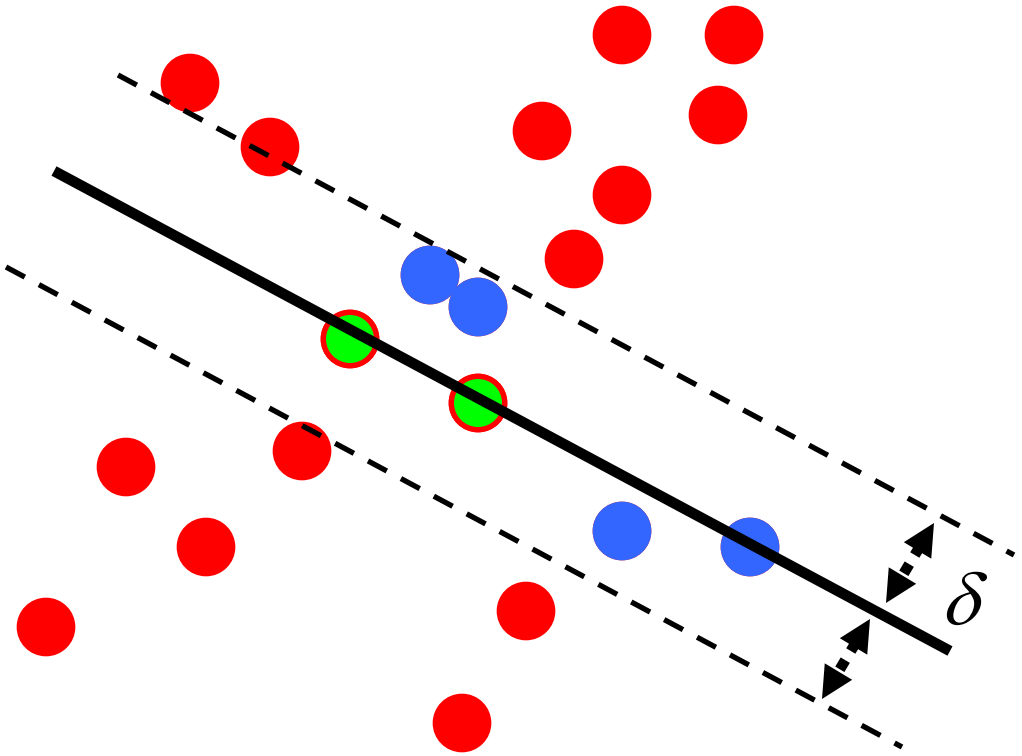
Sample set = set of points in 2D

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

RANSAC



Sample set = set of points in 2D

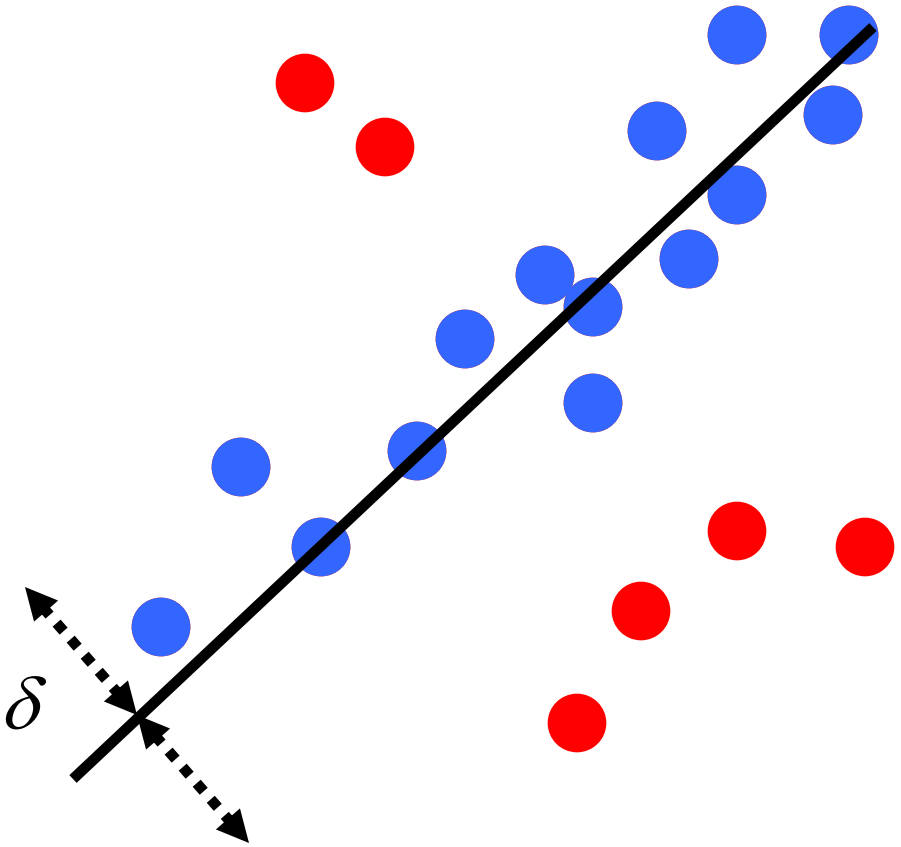
$|O| = 14$

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



$$\pi : \boldsymbol{I} \rightarrow \{\boldsymbol{P}, \boldsymbol{O}\} \qquad \min_{\pi} |\boldsymbol{O}|$$

such that:

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

Model parameters

Recognition

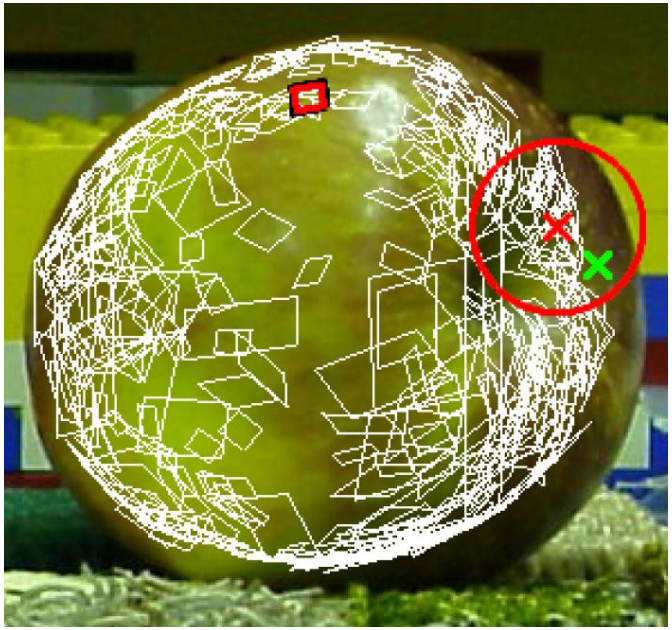
- Hypothesis generation & model verification

[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:

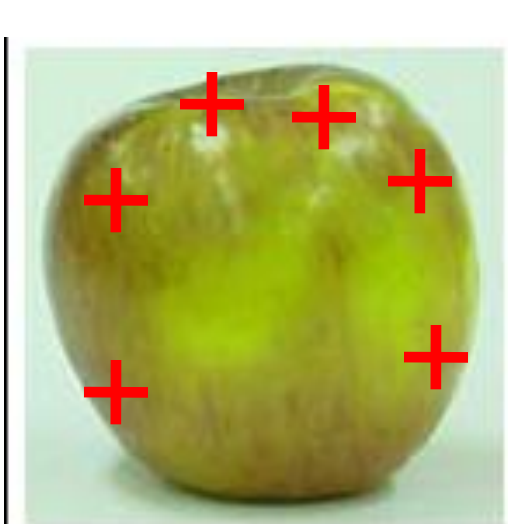
  • For every 2 pairs of matches
  • Compute camera
  • Use camera to project other matched 3D model patches into test image
  • Verification test

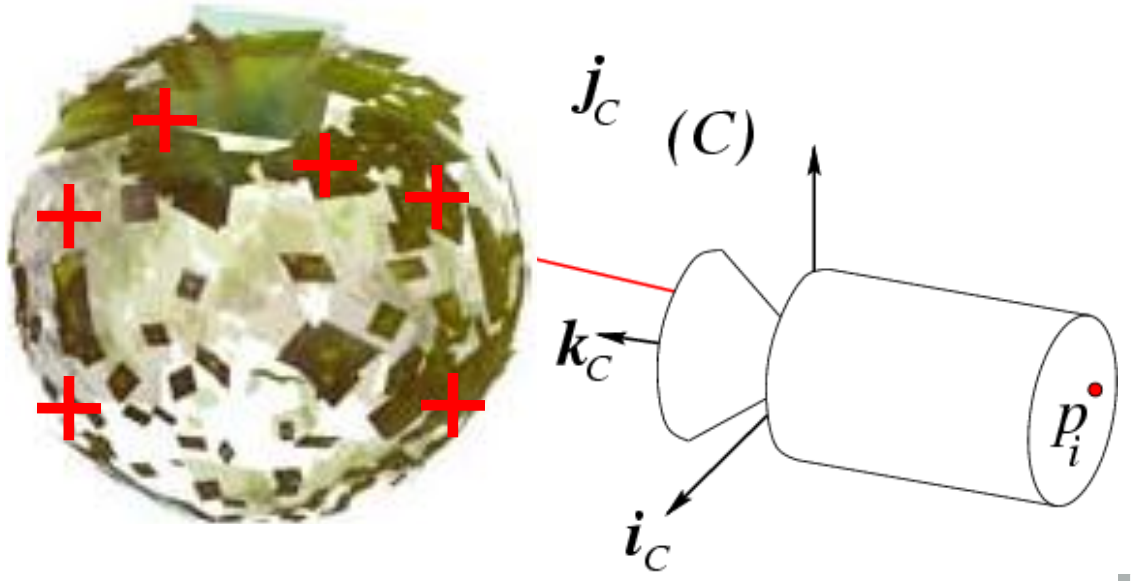


Courtesy of Rothganger et al

Recognition

[Rothganger et al. '03 '06]

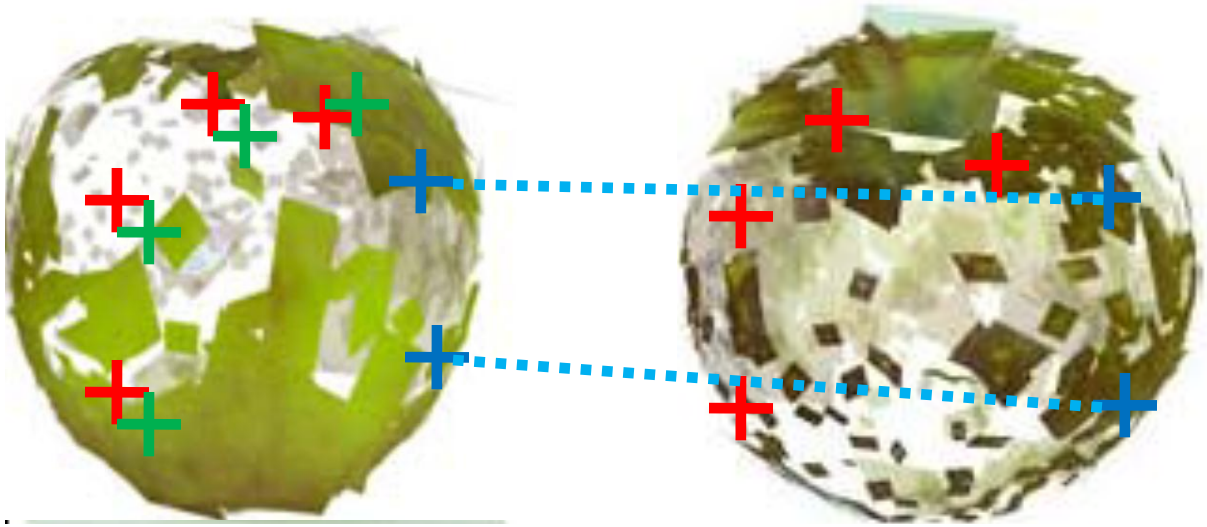




1. Find matches between model and test image features

Recognition

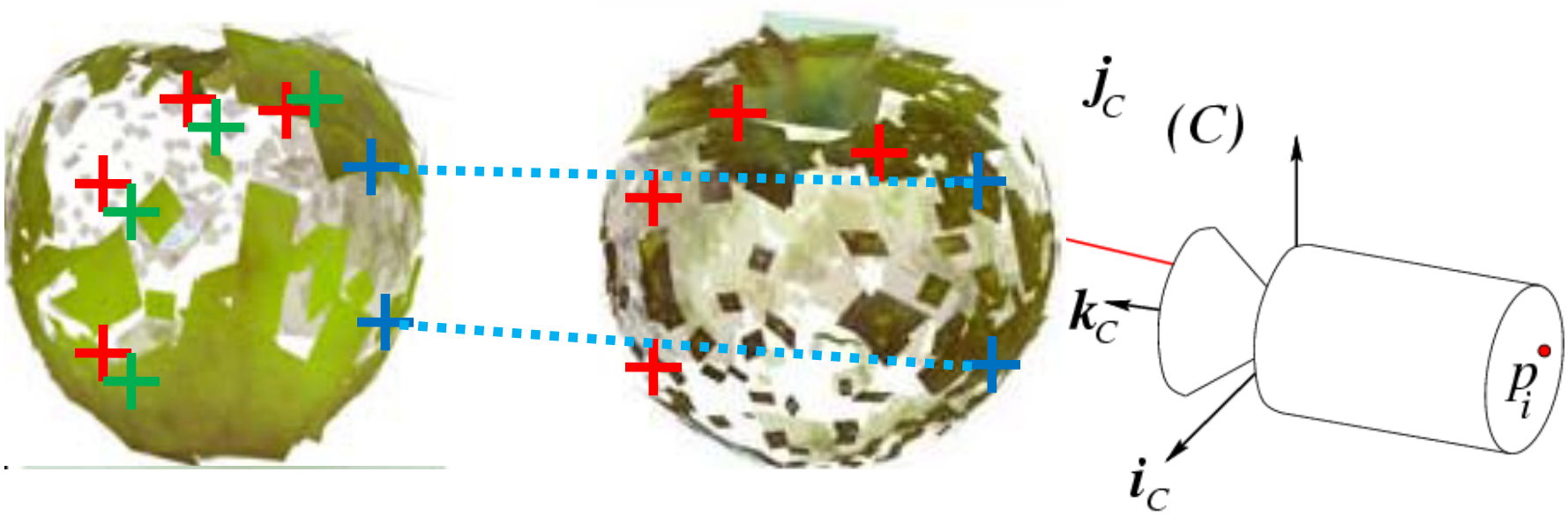
[Rothganger et al. '03 '06]



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(projections, measurements)

Recognition

[Rothganger et al. '03 '06]



Goal:

Estimate (fit) the best M in presence of outliers

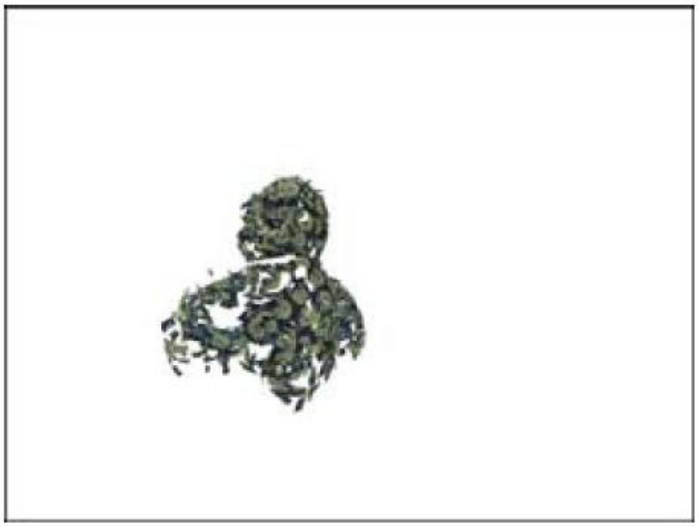
Object to recognize

Initial matches based on appearance

Courtesy of Rothganger et al

Matches verified with geometrical constraints

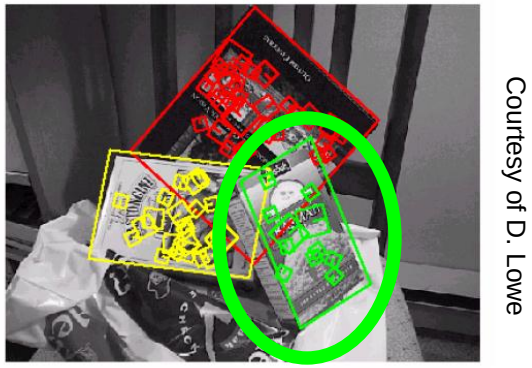
Recovered pose



Recognition

[Lowe '99, '01, '04]

Model



Courtesy of D. Lowe

Test image

- 3 matches generate an hypothesis
  (6 parameters model 2D affine transformation)

Recognition - generating hypothesis

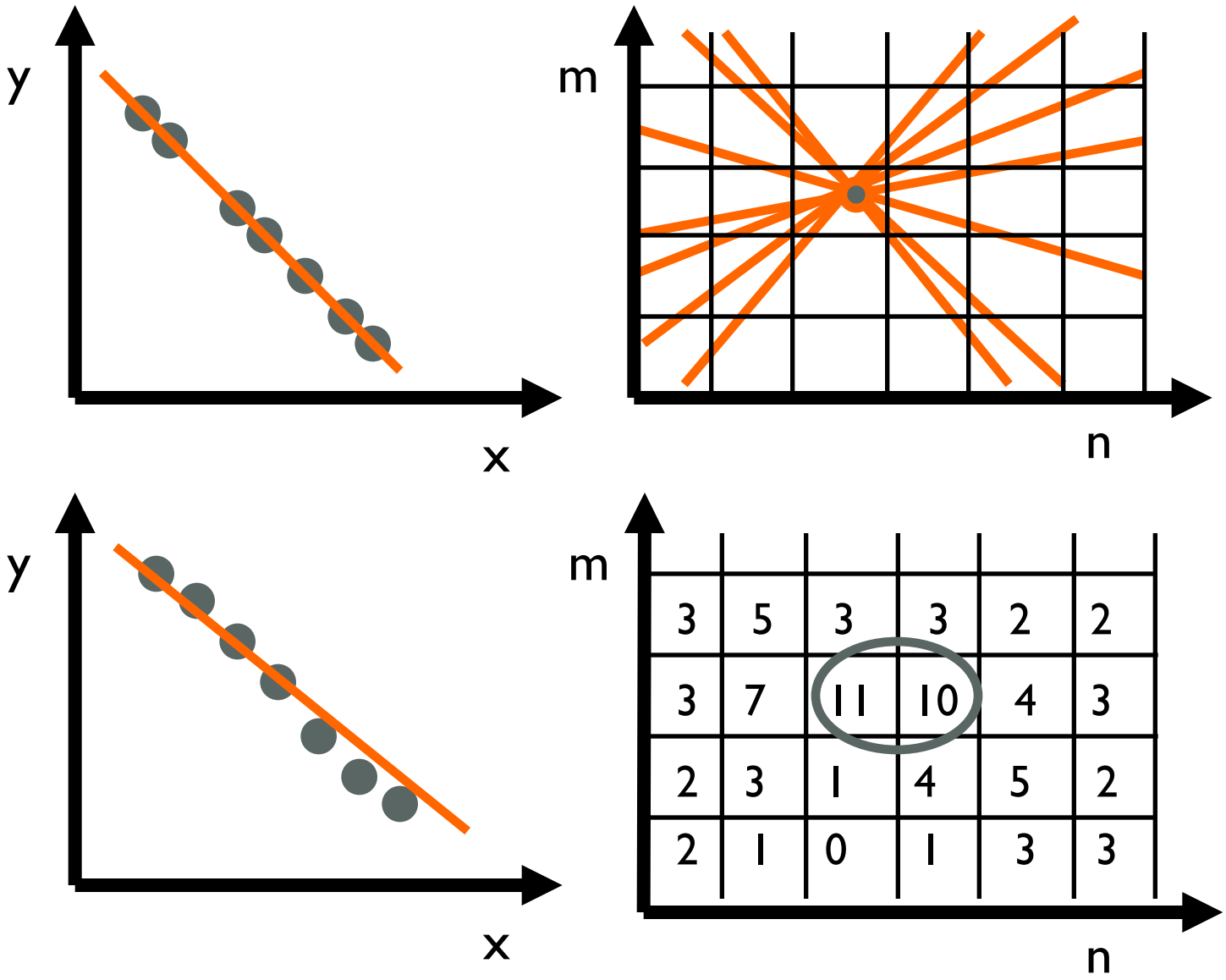
[Lowe '99, '01, '04]

- If the inlier/outlier ratio is too small RANSAC doesn't work…

Courtesy of D. Lowe

- SOLUTION: **Hough transform**
  - –Vote for each potential match according to model ID and pose

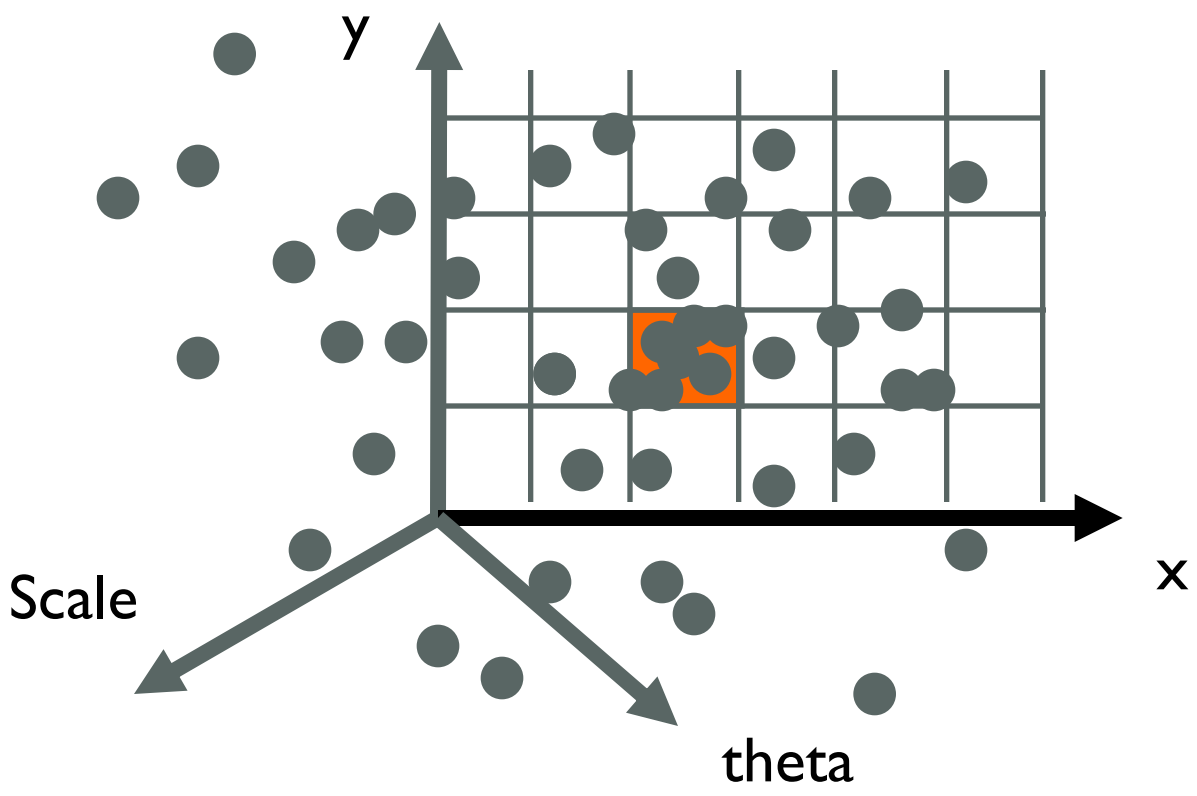
Hough Transform



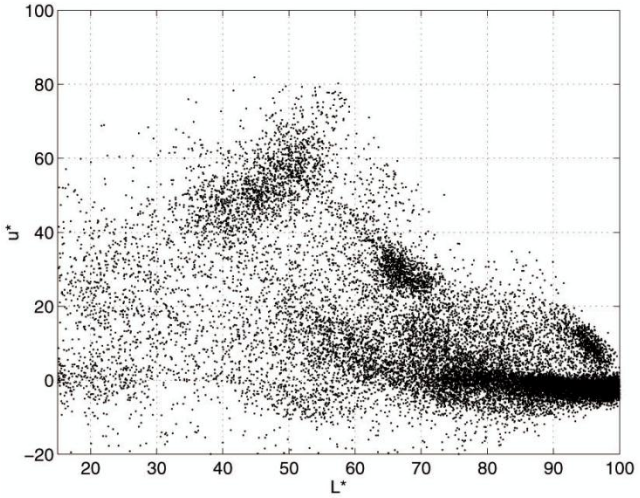
Hough transform

Each matched keypoint ↔

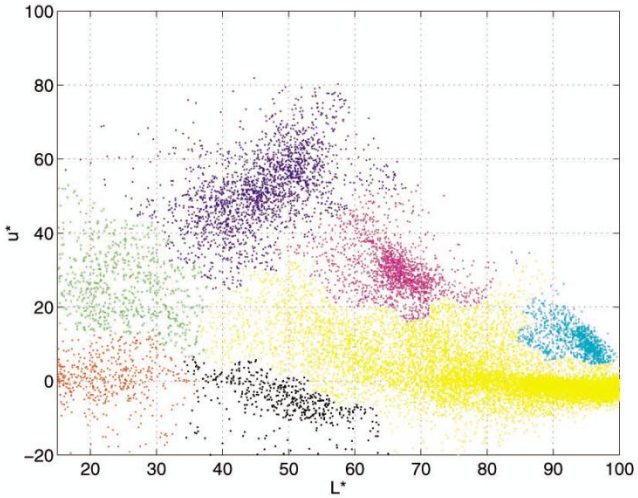
Entry in Hough transform space



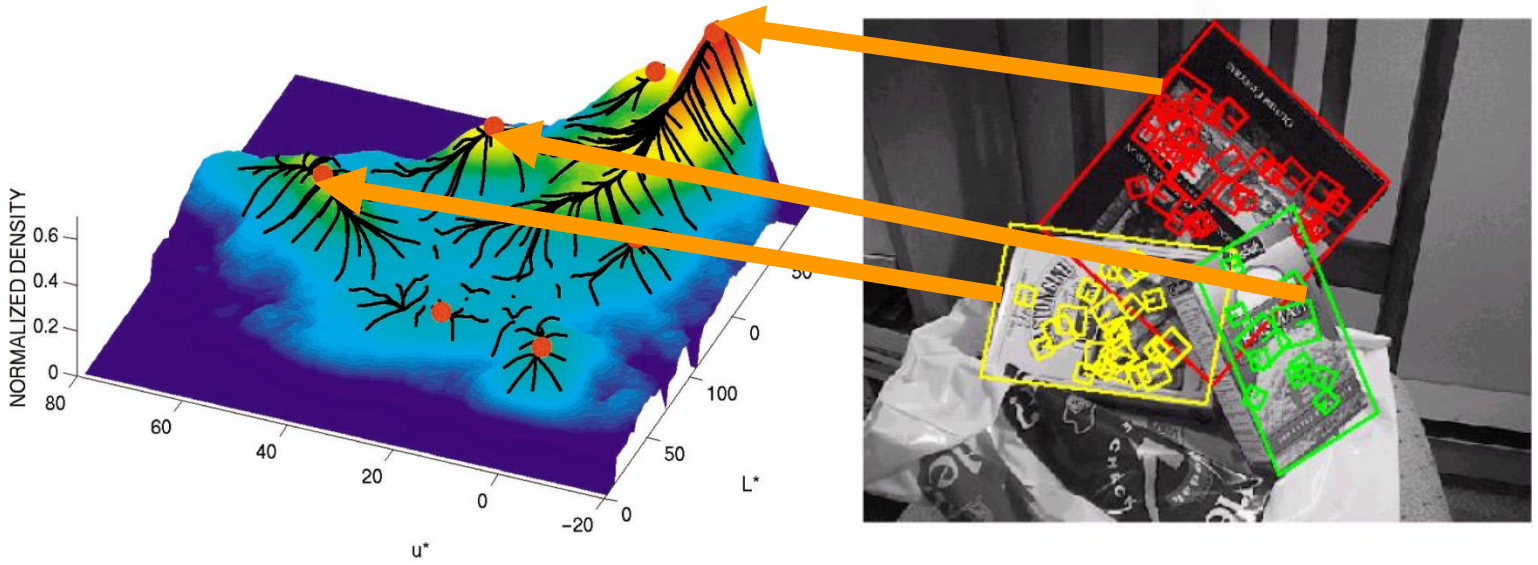
Hough transform



(a)



(b)



Recognition - model verification

[Lowe '99, '01, '04]

1. Examine all clusters with at least 3 features → Consistent configuration
2. Perform least-squares affine fit to model.
3. Discard outliers and perform top-down check for additional features.
4. Evaluate probability that match is correct
   - Use Bayesian model, with probability that features would arise by chance if object was *not* present (Lowe, CVPR 01)

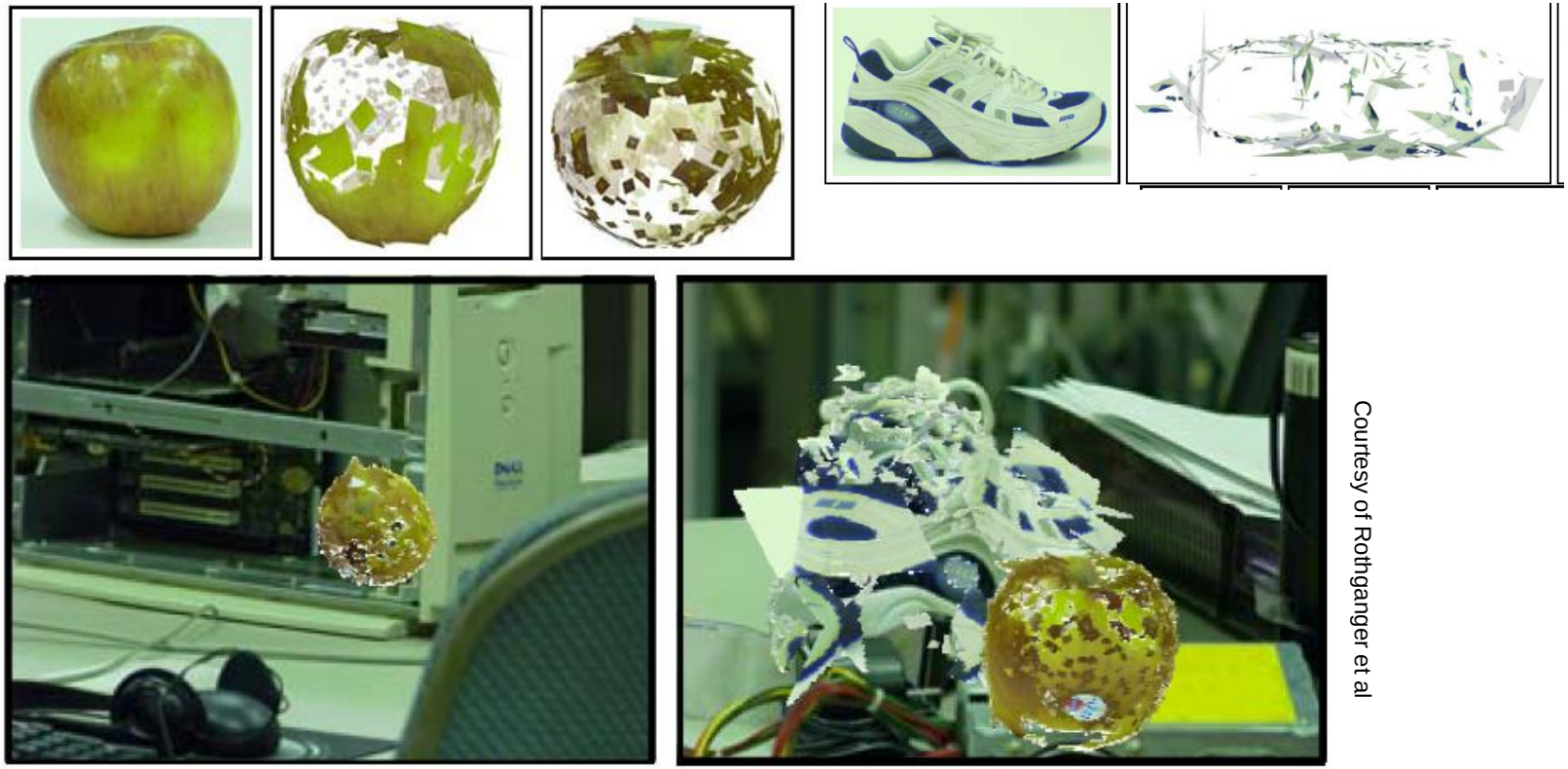
Basic Scheme

- Representation
  - Features
  - 2D/3D Geometrical constraints
- Model learning
- Recognition
  - Hypothesis generation
  - Validation

Let's see some results!

3D Object Recognition results

Rothganger et al. '03 '06

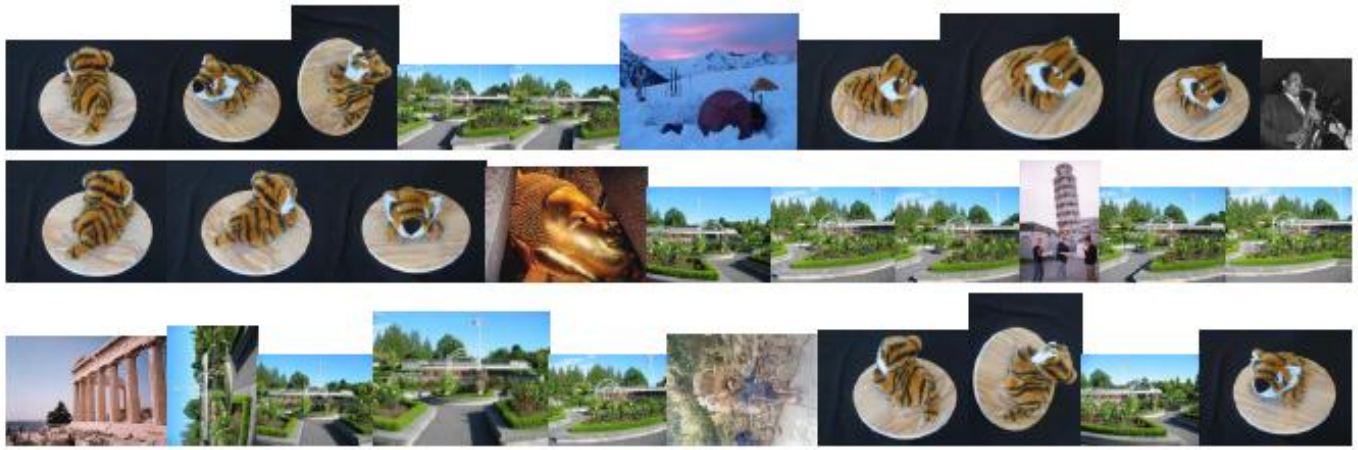


Courtesy of Rothganger et al

- Handle severe clutter

Recognition

Brown & Lowe '05

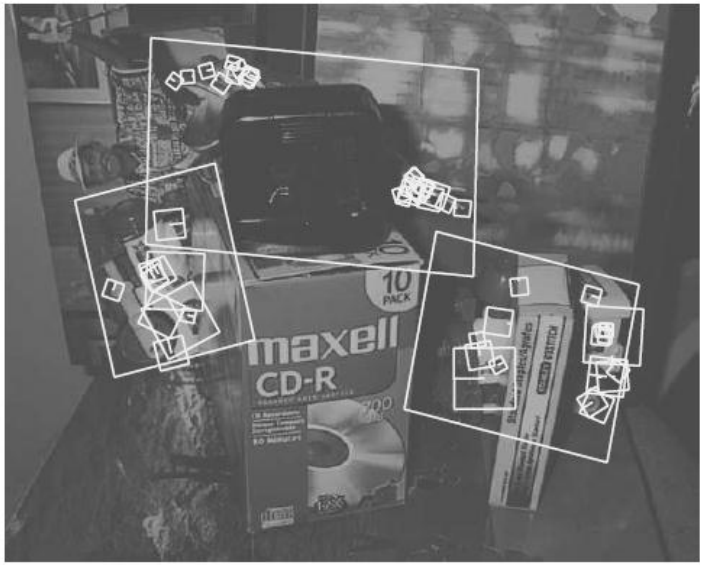


(b) Output 3D model 1 - Tiger

Courtesy of Brown & Lowe

3D Object Recognition results

Lowe. '99, '04

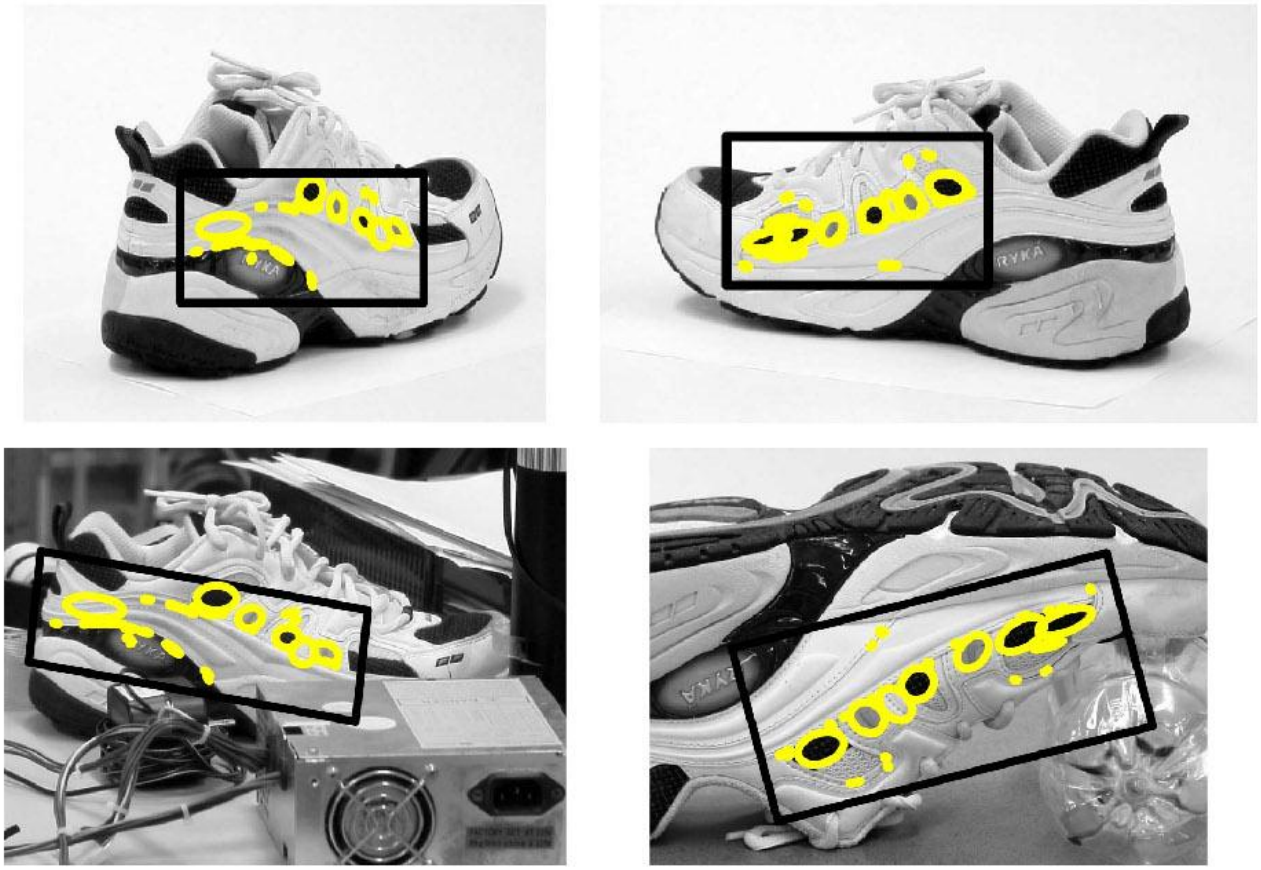


Courtesy of D. Lowe

- Handle severe occlusions
- Fast!

3D Object Recognition results

[Lazebnick et al '04]

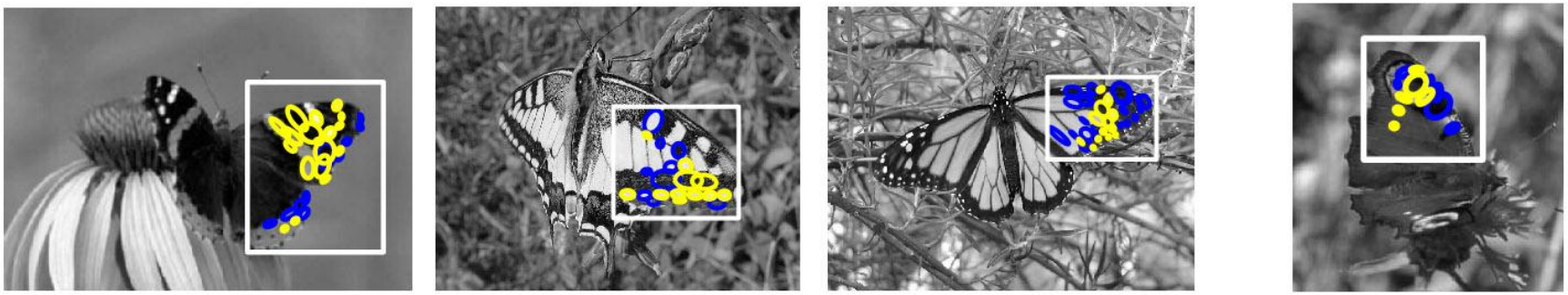


Courtesy of Lazebnick et al

3D Object Recognition results

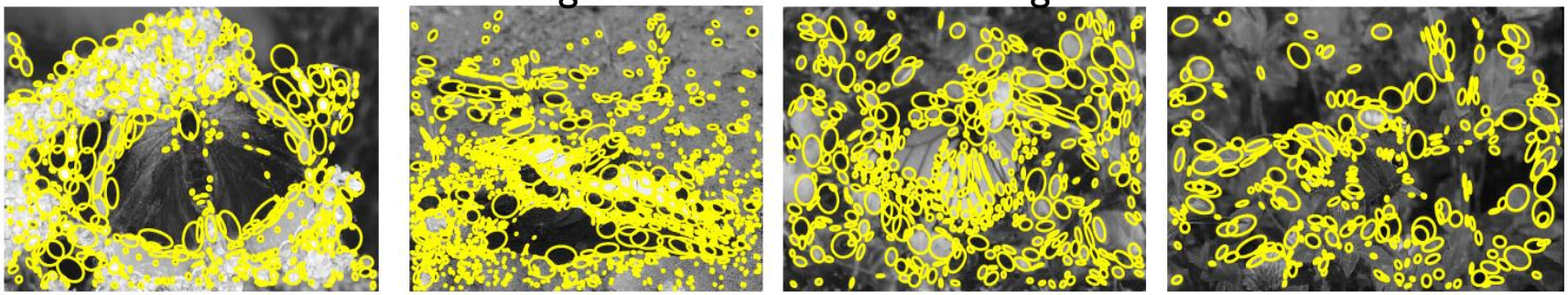
[Lazebnick et al '04]

Training images



Courtesy of Lazebnick et al

All regions found in the test images



Test images (blue: occluded regions)

3D Object Recognition results

[Ferrari et al '04]

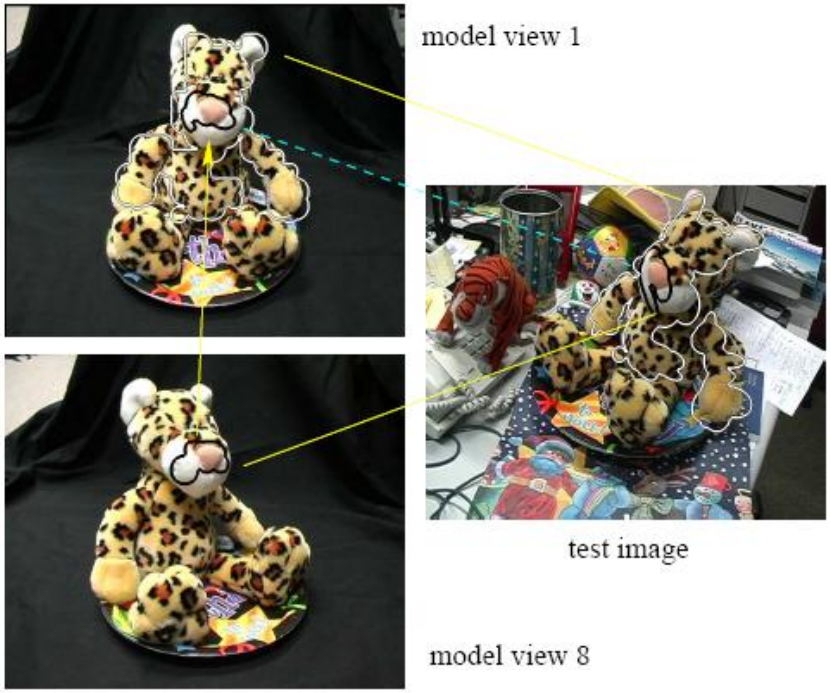
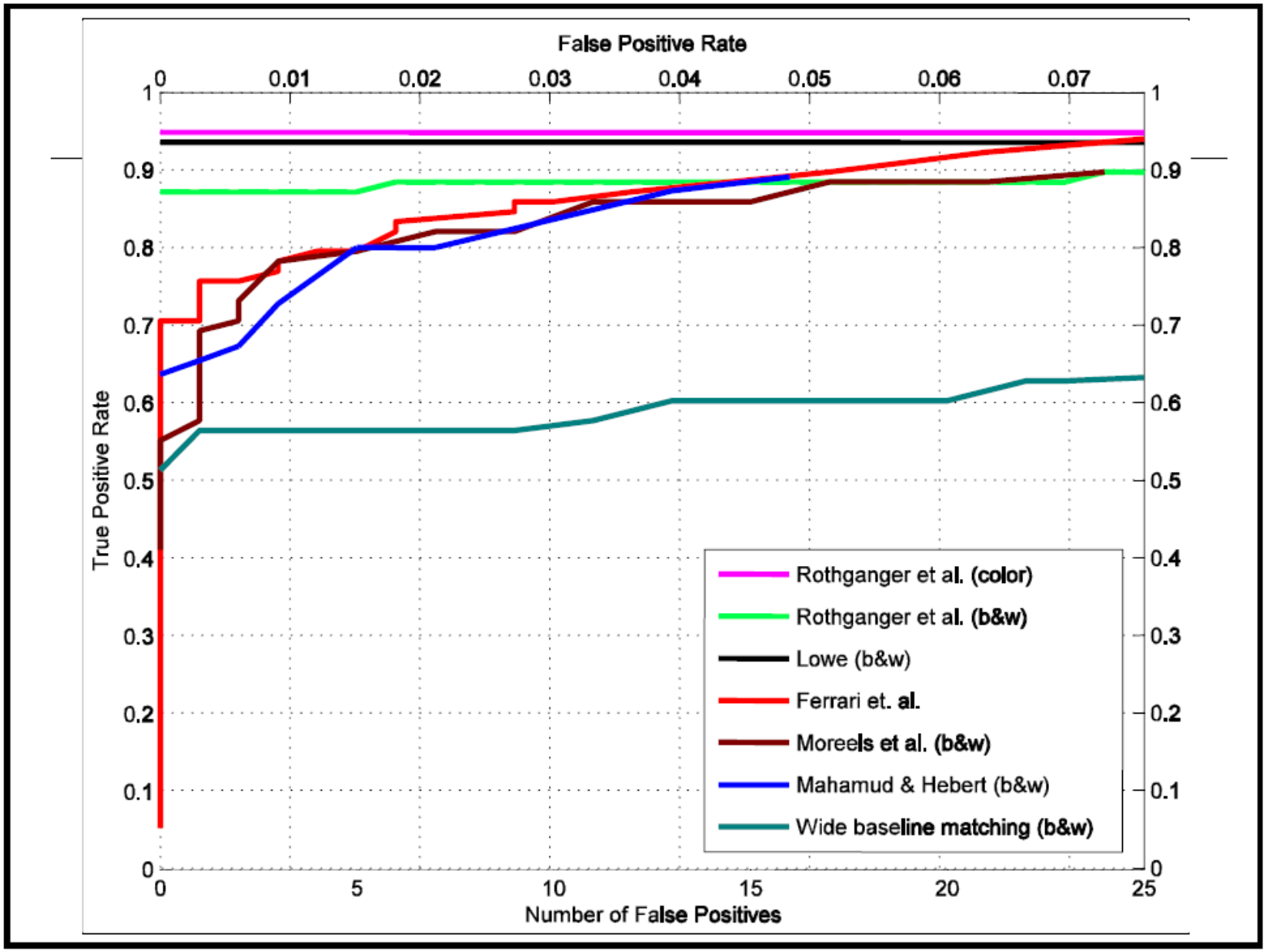
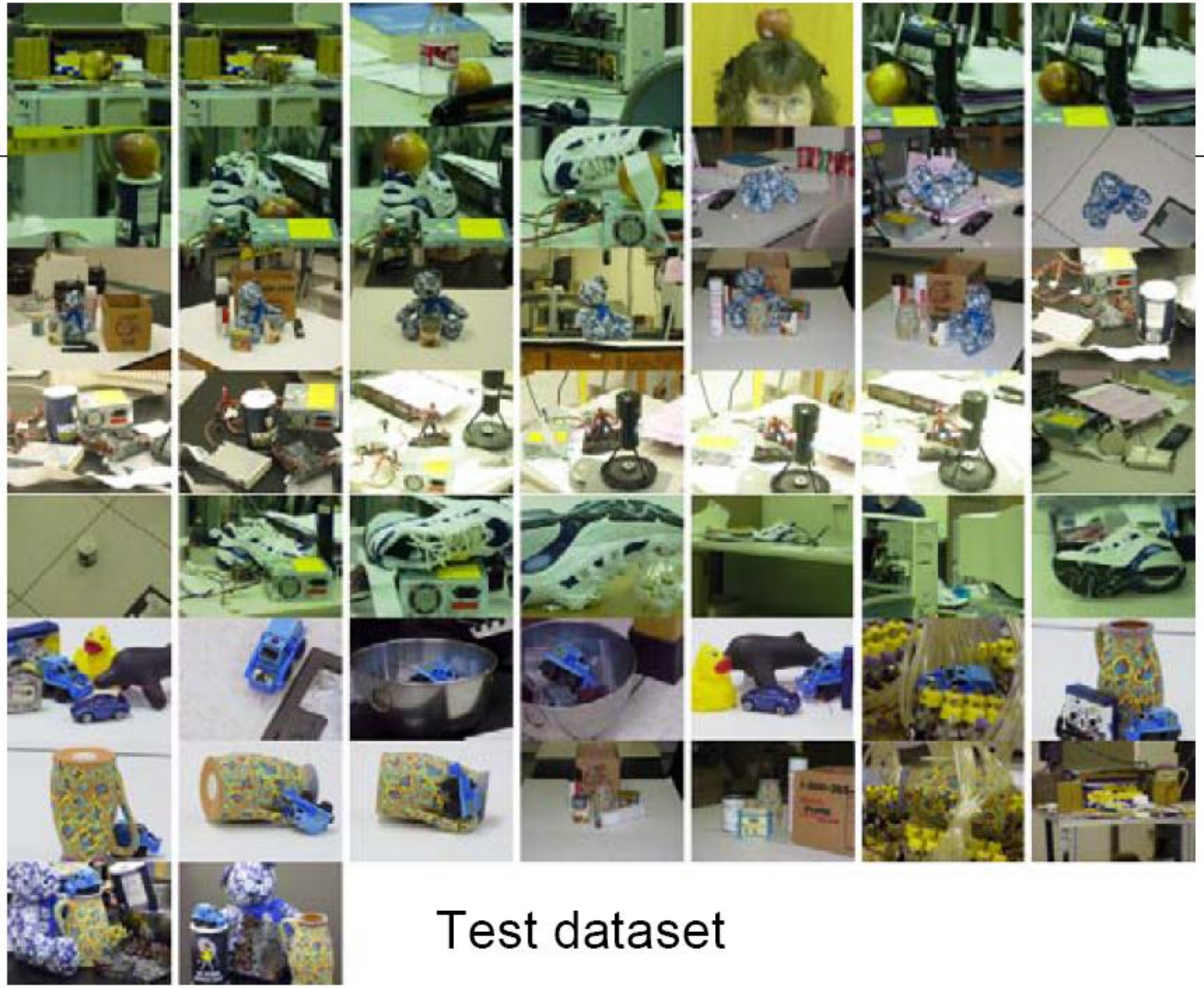


Figure 17: *Two compatible (and correct) GAMs. The nose GAM (black) is initially matched from model view 8, and is transferred to model view 1. Note how the other GAM (white) is very large and covers the head, arms and chest. A GAM can extend over multiple facets when the combination of viewpoints and 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 Ferrariet al

A comparative experiment





Test dataset

Outline

- Object Recognition

- Introduction
- Recognition of single 3D objects
  - Bag of word models
  - Part based models
  - Models for 3D objects categorization