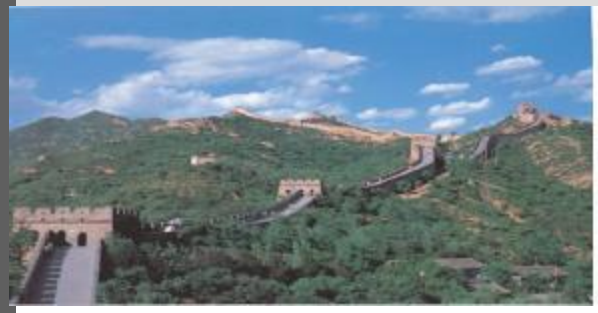
Methods for Representing and Recognizing 3D objects

part 1



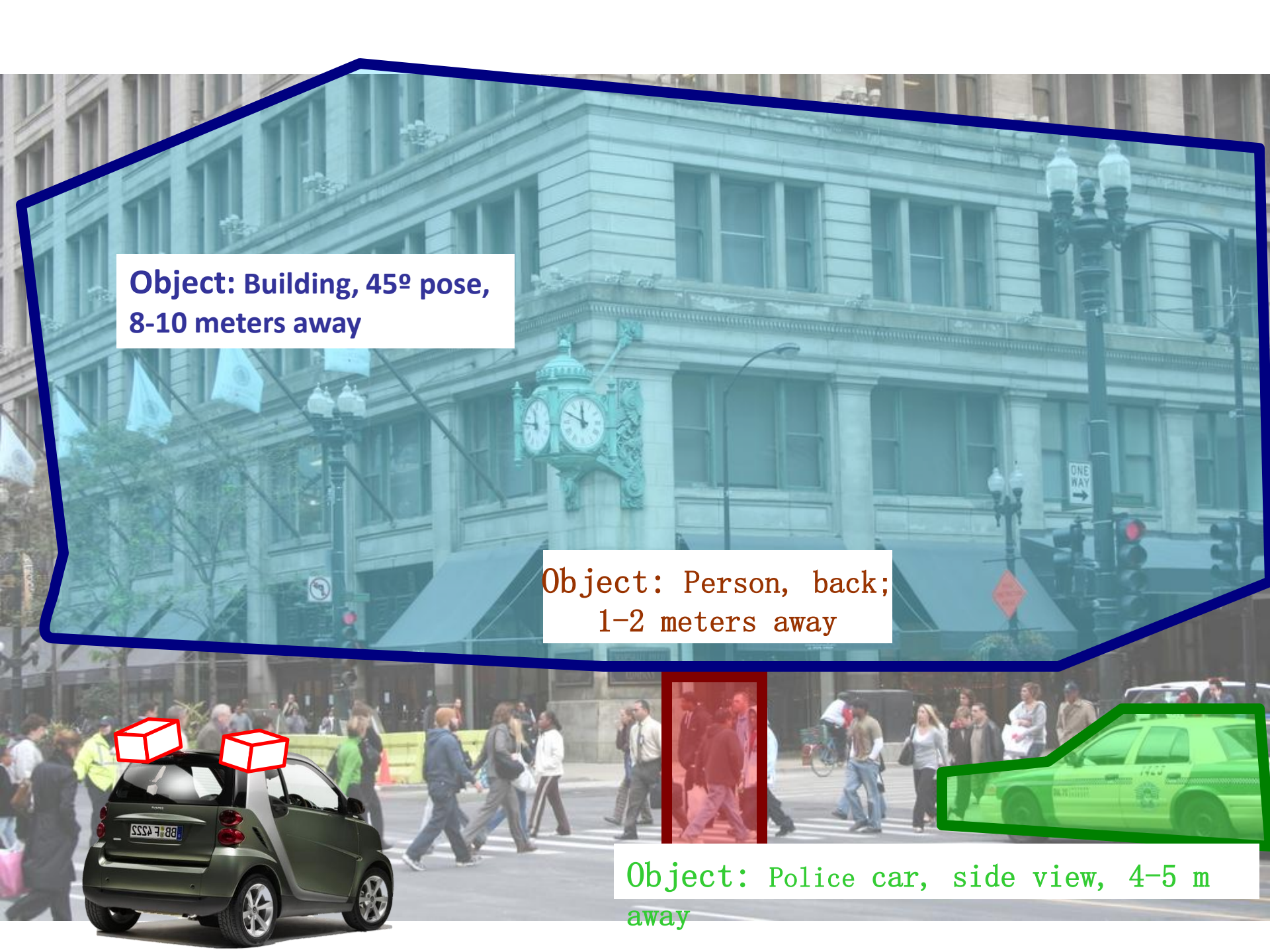
1st Sino-USA Summer School in Vision, Learning, and Pattern Recognition

VLPR 2009 • July 20-27, 2009 • Peking University, Beijing, China

Silvio Savarese

University of Michigan at Ann Arbor


NOKIA
N95
ONE
WAY
macy's

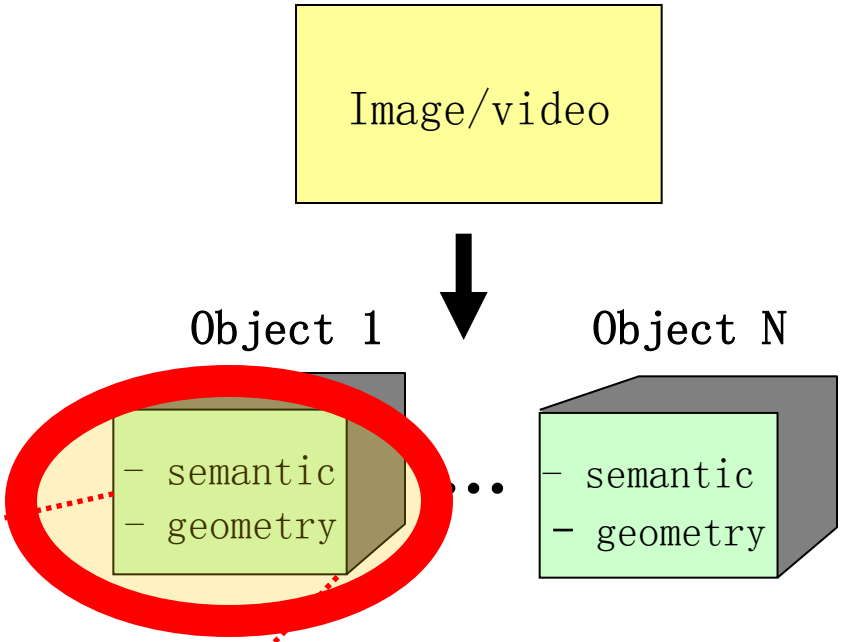

Object: Building, 45º pose, 8-10 meters away
Object: Person, back; 1-2 meters away
Object: Police car, side view, 4-5 m away

Object: Building’ s name;
What are the businesses inside?

Street or intersection name

Visual technology

- Scene understanding
- Navigation
- Interaction
- Augmentation
- Manipulation



- Recognizing objects under arbitrary viewing conditions
- Recognize their pose

•Minimal supervision



















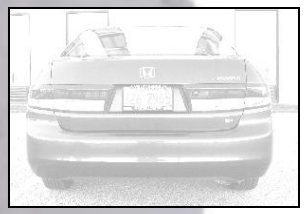
Single 3D object recognition













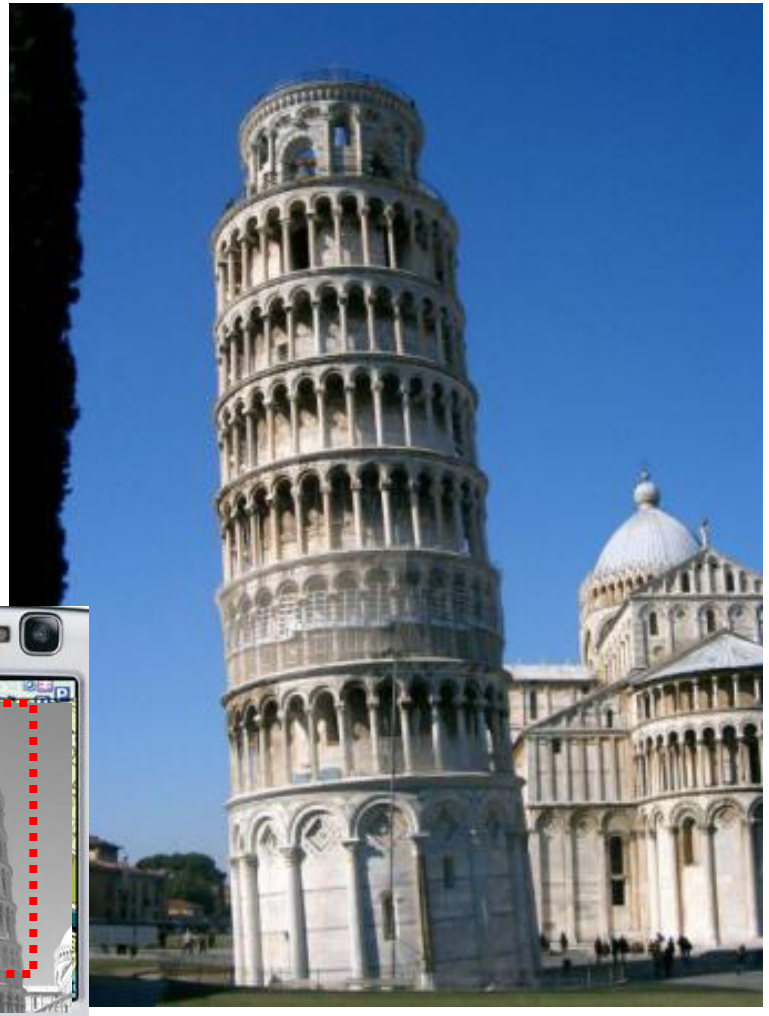


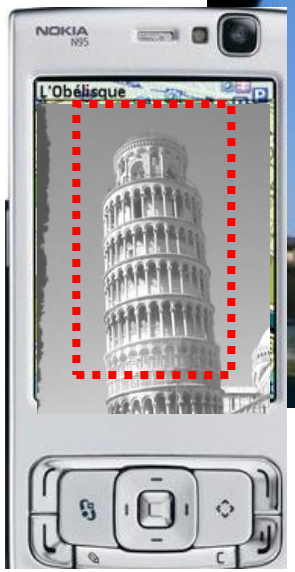




- Ballard, '81
- Grimson & L.-Perez, '87
- Lowe, ' 87
- Edelman et al. ' 91
- Ullman & Barsi, ' 91
- Rothwell '92
- Linderberg, ' 94
- Murase & Nayar '94
- Zhang et al ' 95
- Schmid & Mohr, '96
- Schiele & Crowley, ' 96
- Lowe, '99
- Jacob & Barsi, '99
- Rothganger et al., '04
- Ferrari et al, ' 05
- Brown & Lowe ' 05
- Snavely et al ' 06

Recognition for Virtual sightseeing





NOKIA

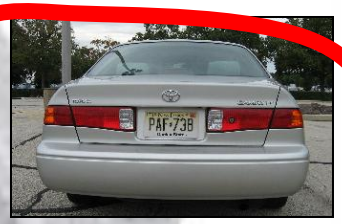
Object manipulations

Robot arm for automatic gas fill up

Object manipulations

Single view object categorization



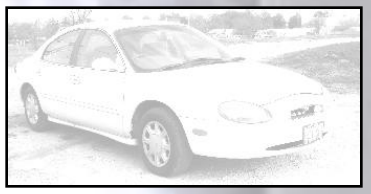











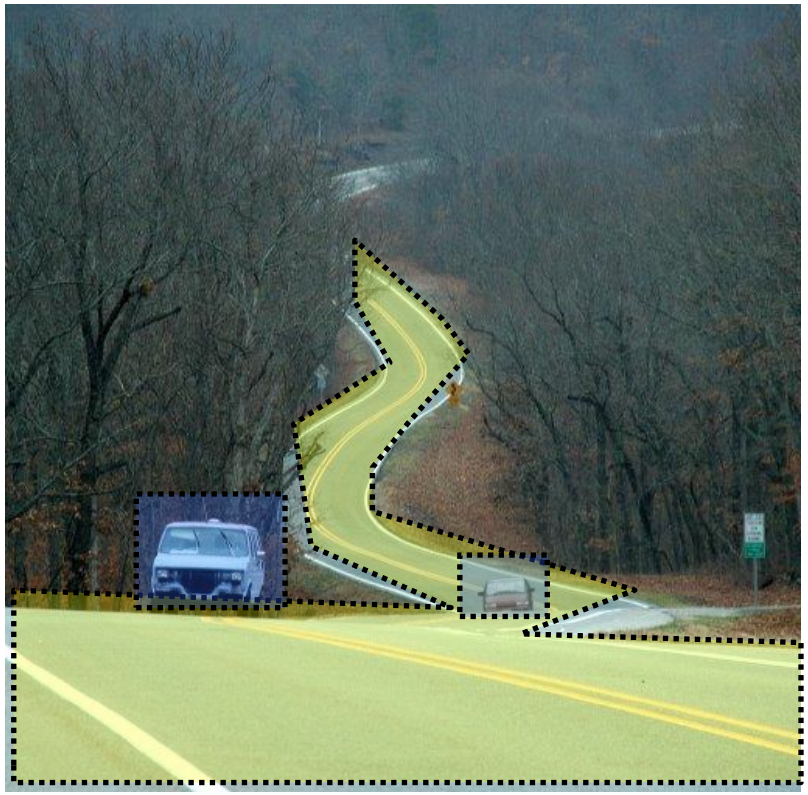


- Leung et al '99
- Weber et al. '00
- Ullman et al. 02
- Fergus et al. ' 03
- Torralba et al. '03

- Felzenszwalb & Huttenlocher '03
- Fei-Fei et al. '04
- Leibe et al. '04

- Kumar & Hebert '04
- Sivic et al. ' 05
- Shotton et al '05
- Grauman et al. '05

- Sudderth et al '05
- Torralba et al. '05
- Lazebnik et al. '06
- Todorovic et al. ' 06
- Bosh et al '07
- Vedaldi & Soatto ' 08
- Zhu et al 08

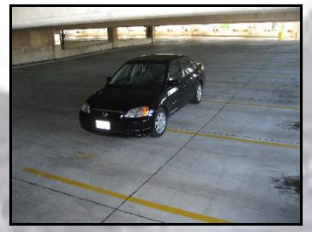


Safe driving

Security

photography

3D Object Categorization
















- Weber et al. '00
- Schneiderman et al. '01
- Capel et al '02
- Johnson & Herbert '99
- Bronstein et al, '03
- Ruiz-Correa et al. '03,
- Funkhouser et al '03
- Bart et al '04
- Thomas et al. '06
- Kushal, et al., '07
- Savarese et al, 07, 08
- Chiu et al. '07
- Hoiem, et al., '07
- Yan, et al. '07

Overview

- Single 3D object recognition
- Single view object categorization
- 3D object categorization

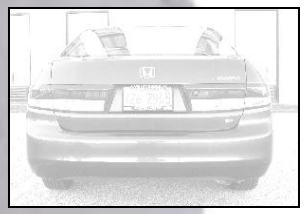
Single 3D object recognition



















- 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
- Rothganger et al., '04
- Ferrari et al, ' 05
- Brown & Lowe ' 05
- Snavely et al ' 06

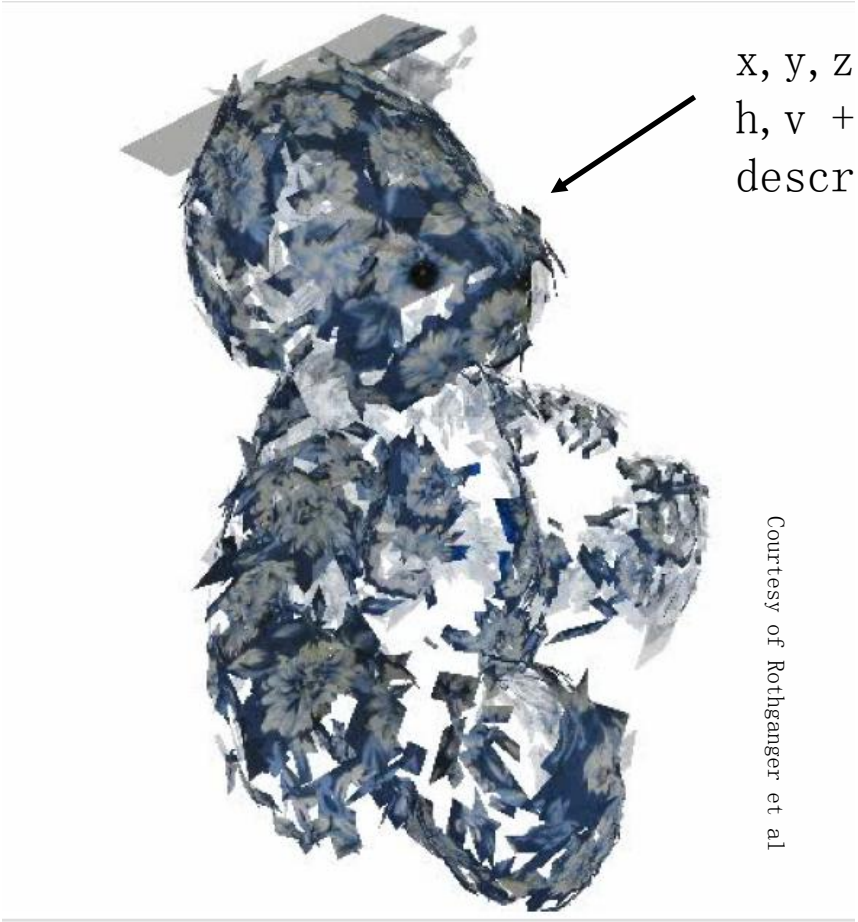
Basic scheme

- Representation
  - Features
  - Descriptors
  - Model
- Model learning

- Recognition
  - Hypothesis generation
  - Model verification

Object representation: Collection of patches in 3D

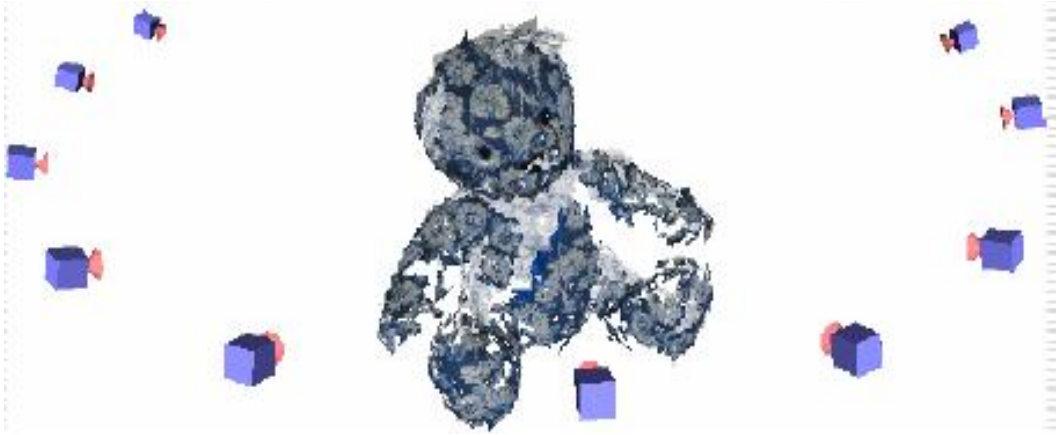
Rothganger et al. ’06



Model learning

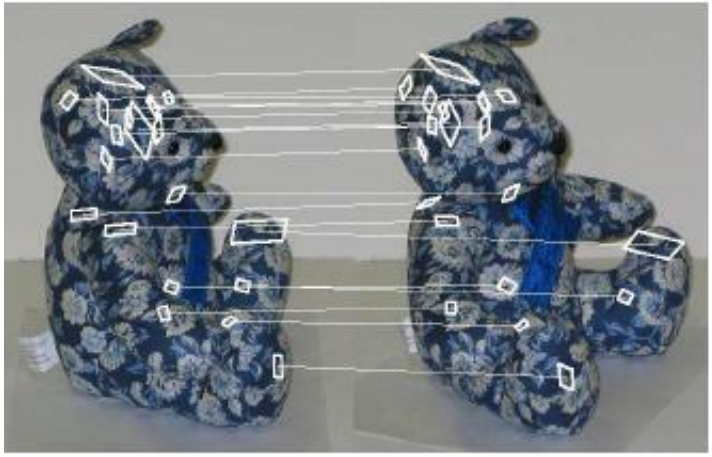
Rothganger et al. ‘03 ’ 06

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 [RANSAC]

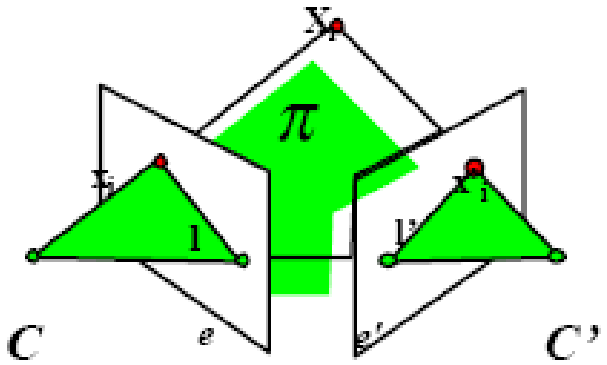


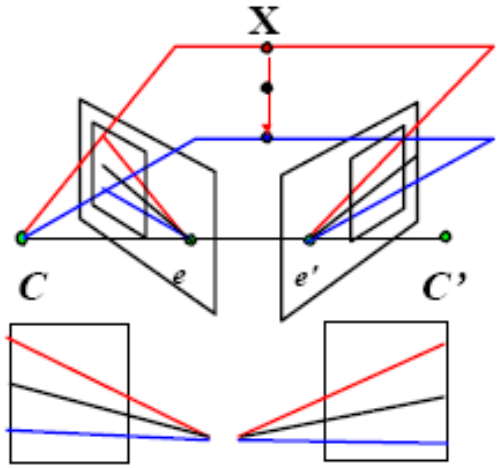
- Find connected components
- Use bundle adjustment to refine model
- Upgrade model to Euclidean assuming zero skew and square pixels

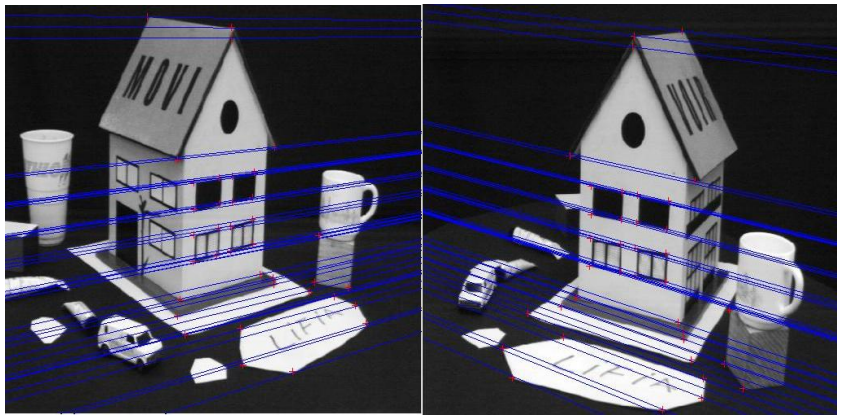
Affine Structure from Motion

Books:

- Faugeras, ‘95
- Zisserman & Hartley, ’00
- Ma, Soatto, et al. ‘05







Affine epipolar geometry between image pairs.

Fundamental matrix F imposes:

x’ Fx = 0 and l’ = Fx

Learnt models

Rothganger et al. ‘03 ’ 06

Basic scheme

- Representation
  - Features
  - Descriptors
  - Model

- Model learning

- Recognition [object instance from a single image]
  - Hypothesis generation
  - Model verification

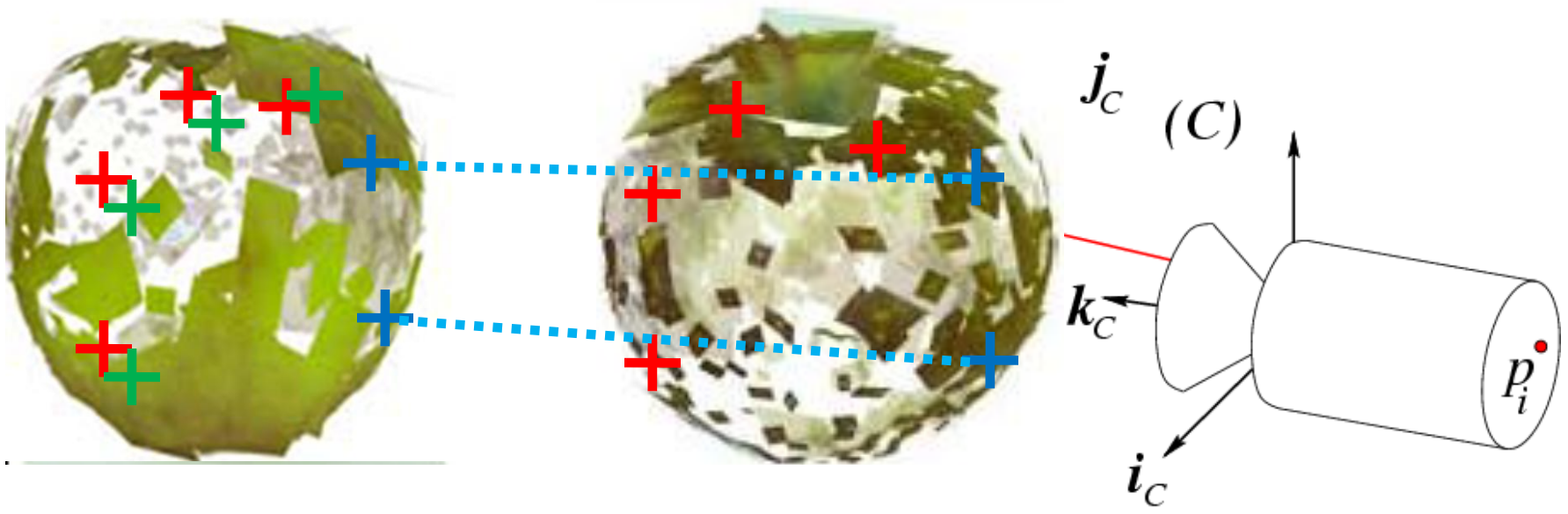
Recognition

[Rothganger et al. '03 ' 06]

1. Find matches between model and test image features

Recognition

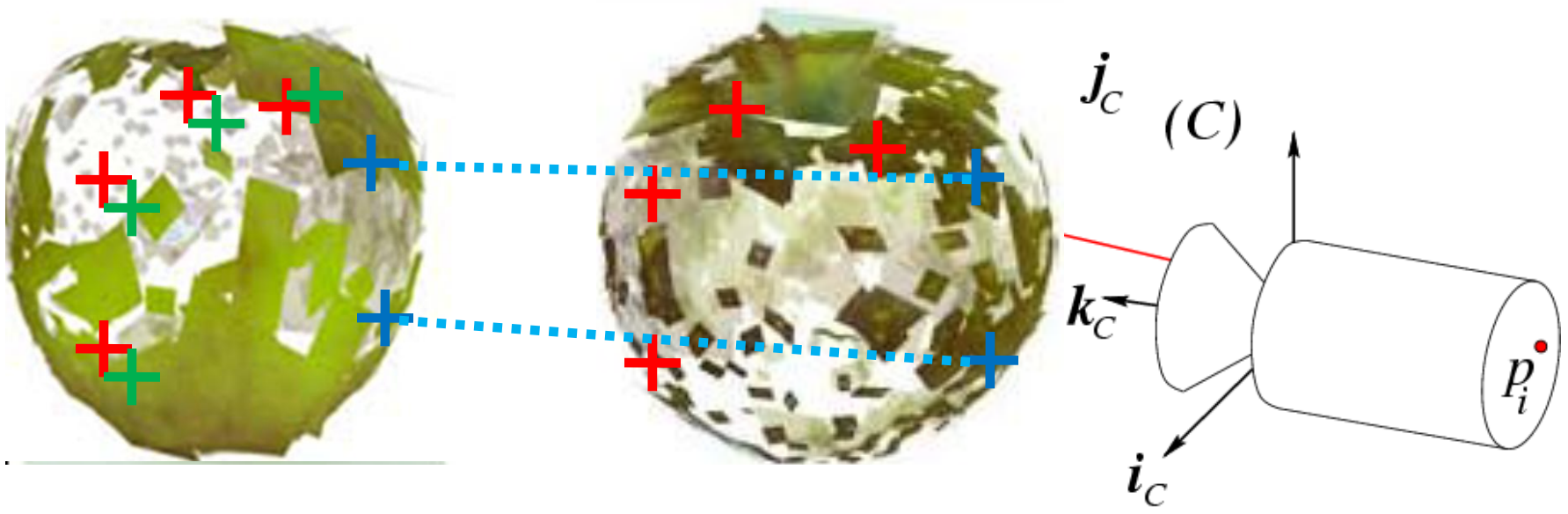
[Rothganger et al. '03 ' 06]



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 2D model features into test image

Recognition

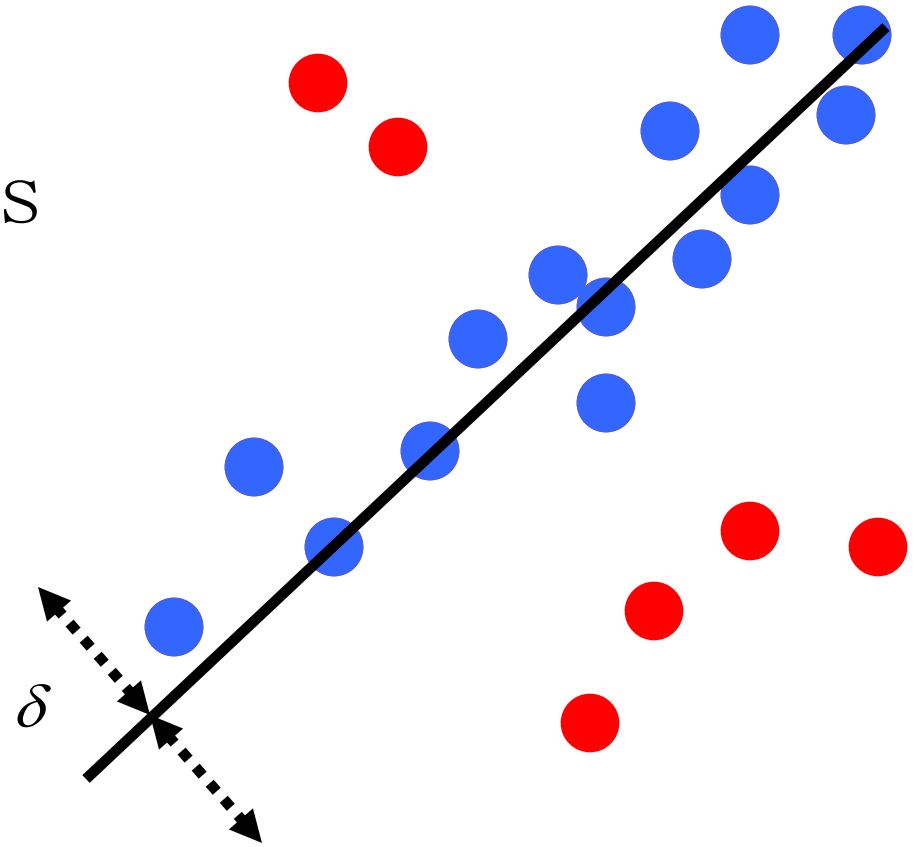
[Rothganger et al. '03 ' 06]



Goal:

Estimate (fit) the best M in presence of outliers

Line fitting with outliers



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

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

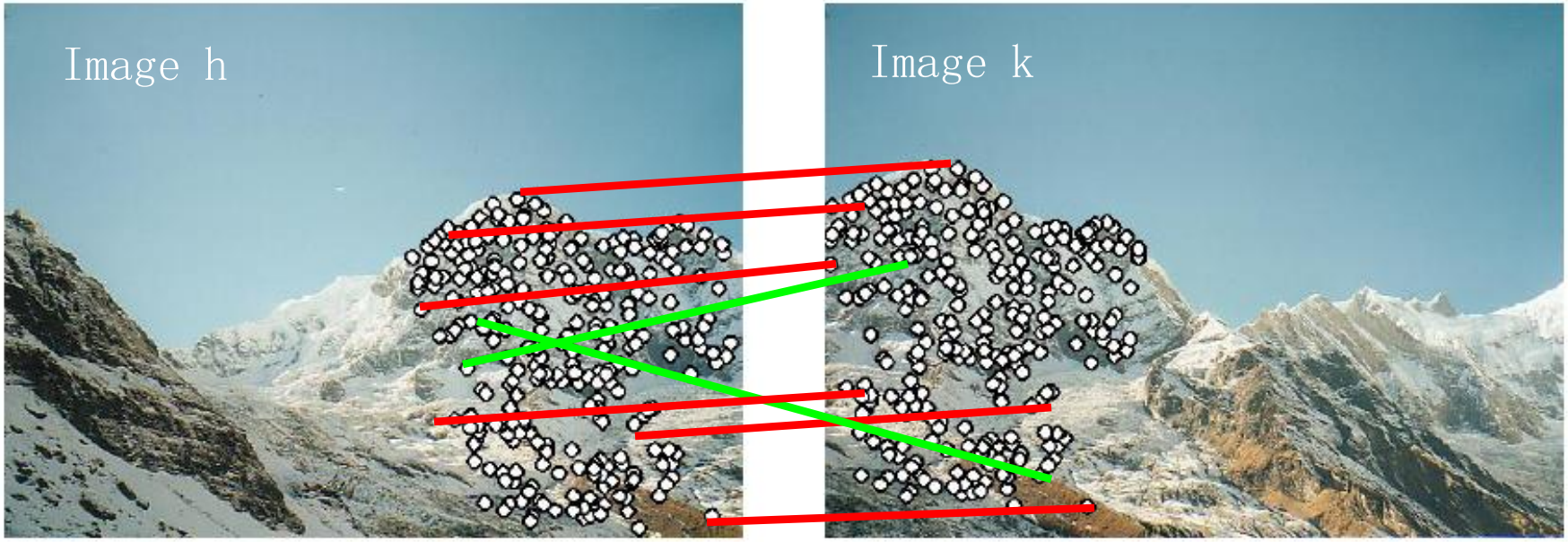
such that:

$$\boldsymbol{f}(\boldsymbol{P}, \beta) < \delta$$

Model parameters

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

Fitting homographies for stitching panorama



$x_1, x_2, \cdots, x_n$

$$\pi : I^h \rightarrow \{P^h, O^h\}$$
$$\tau : I^k \rightarrow \{P^k, O^k\}$$

$$\min \left| O^h \cup O^k \right|$$

such that:

$$f(P^h, P^k, \beta) < \delta$$

$$f(P^h, P^k, \beta) = \left\| P^h - H \quad P^k \right\|$$

Fitting homographies for stitching panorama

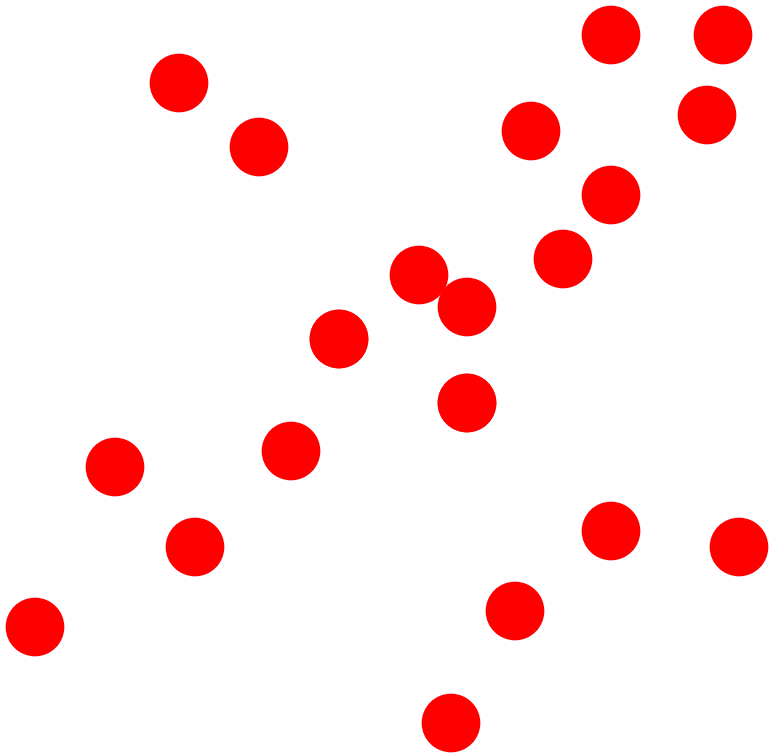
RANSAC – Basic philosophy

(RANdom SAmple Consensus) :
Learning technique to estimate parameters of a model by random sampling of observed data

Fischler & Bolles in '81.

- Data elements are used to vote for one (or multiple) models
- Robust to outliers and missing data
- Assumption1: Noise features will not vote consistently for any single model ("few" outliers)
- Assumption2: there are enough features to agree on a good model ("few" missing data)

RANSAC



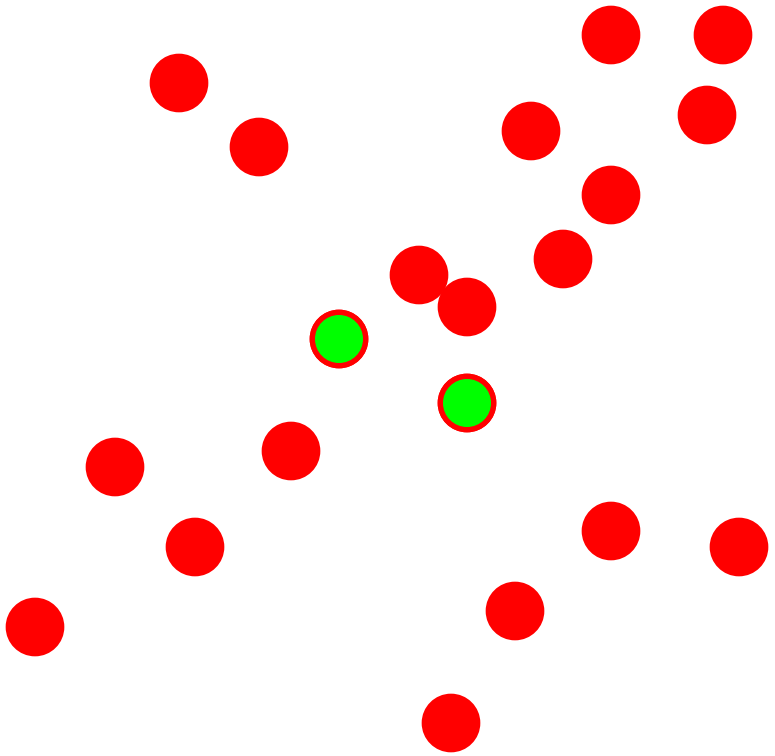
Sample set = set of points in 2D

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

RANSAC



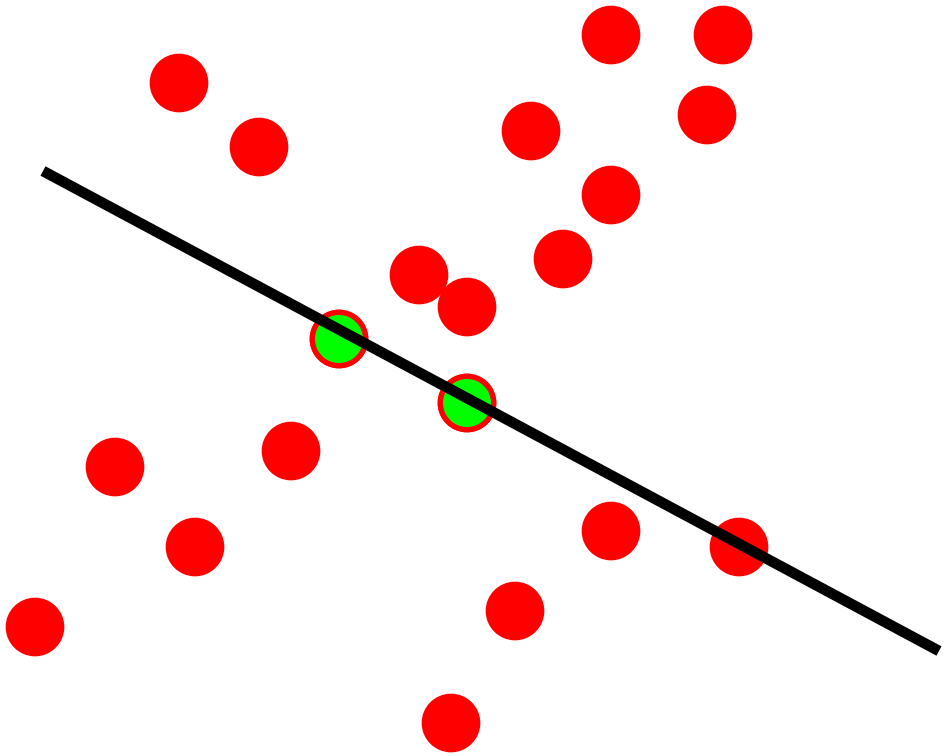
Sample set = set of points in 2D

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



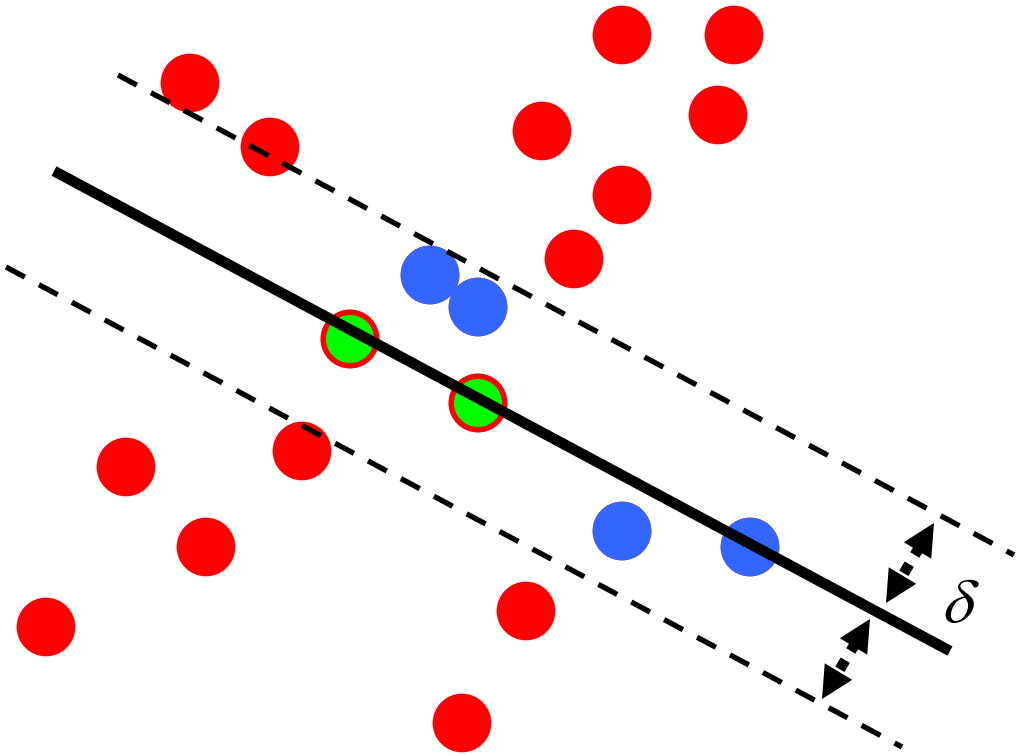
Sample set = set of points in 2D

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:

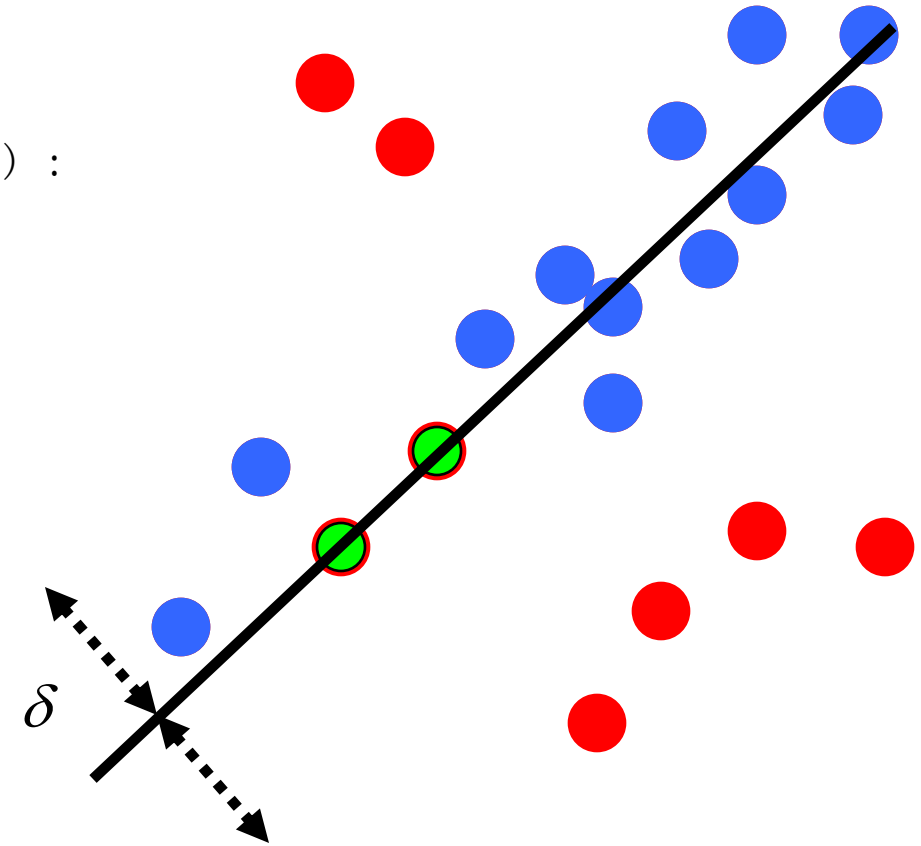
1. Select random sample of minimum required size to fit model=[2]
2. Compute a putative model from sample set
3. Compute the set of inliers to this model from whole data set

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

RANSAC

(RANdom SAmple Consensus) :

Fischler & Bolles in '81.



$|O| = 6$

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

How many samples?

- Number of samples N
  - Choose N so that, with probability p, at least one random sample is free from outliers (e.g. p=0.99) (outlier ratio: e)
- Initial number of points s
  - Typically minimum number needed to fit the model
- Distance threshold δ
  - Choose δ so probability for inlier is p (e.g. 0.95)
  - Zero-mean Gaussian noise with std. dev. σ : $t^2=3.84\sigma^2$

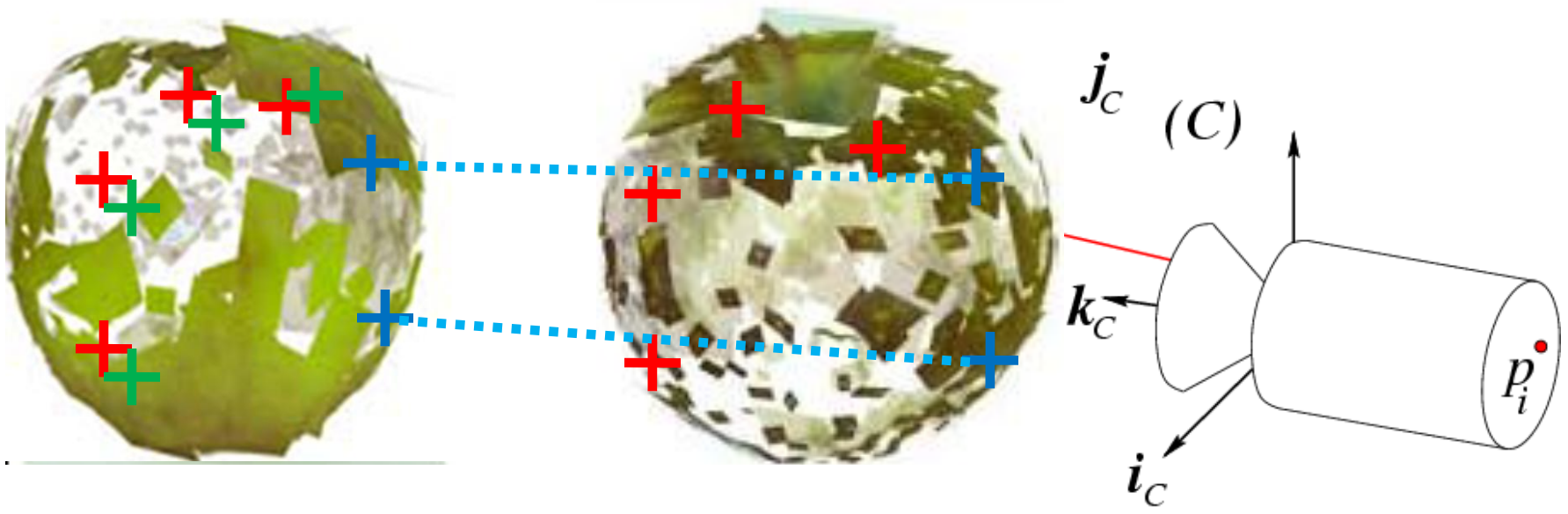
$$N = \log(1-p)/\log\left(1-(1-e)^s\right)$$

	proportion of outliers e						
s	5%	10%	20%	25%	30%	40%	50%
2	2	3	5	6	7	11	17
3	3	4	7	9	11	19	35
4	3	5	9	13	17	34	72
5	4	6	12	17	26	57	146
6	4	7	16	24	37	97	293
7	4	8	20	33	54	163	588
8	5	9	26	44	78	272	1177

Source: M. Pollefeys

Recognition

[Rothganger et al. '03 ' 06]



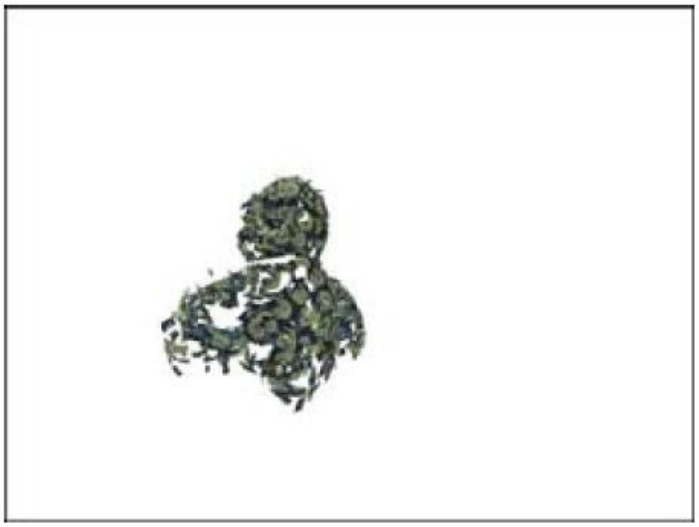
1. Find matches between model and test image features
2. Generate hypothesis:
   - Compute transformation M from N matches (N=2; affine camera; affine key points)
3. Model verification
   - Use M to project other matched 2D model features into test image

Object to recognize

Initial matches based on appearance

Matches after pose verification

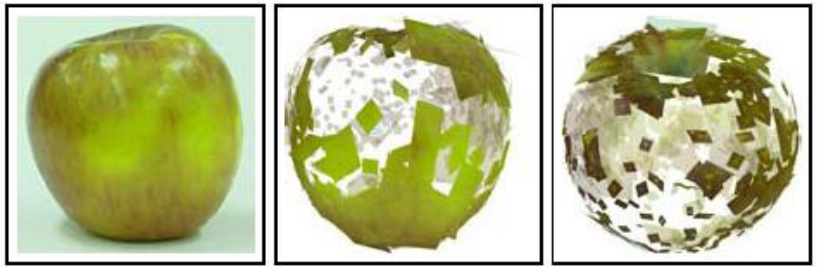
Recovered pose

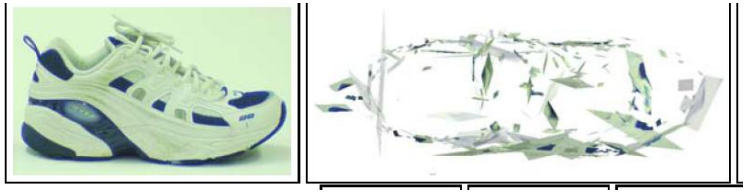


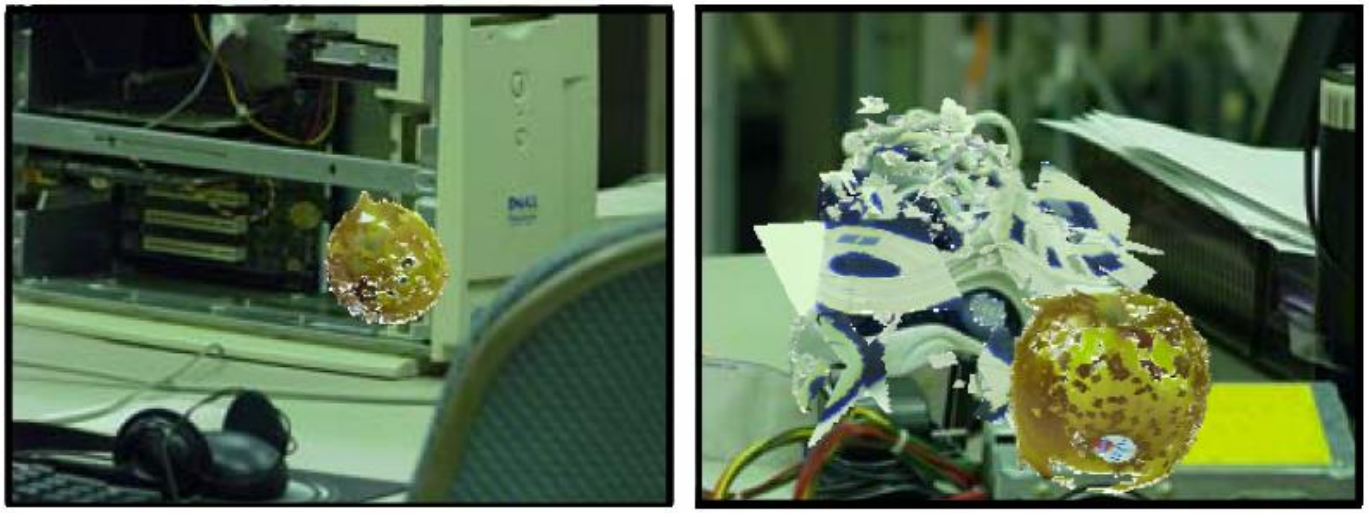
Courtesy of Rothganger et al

3D Object Recognition results

Rothganger et al. ‘03 ’ 06

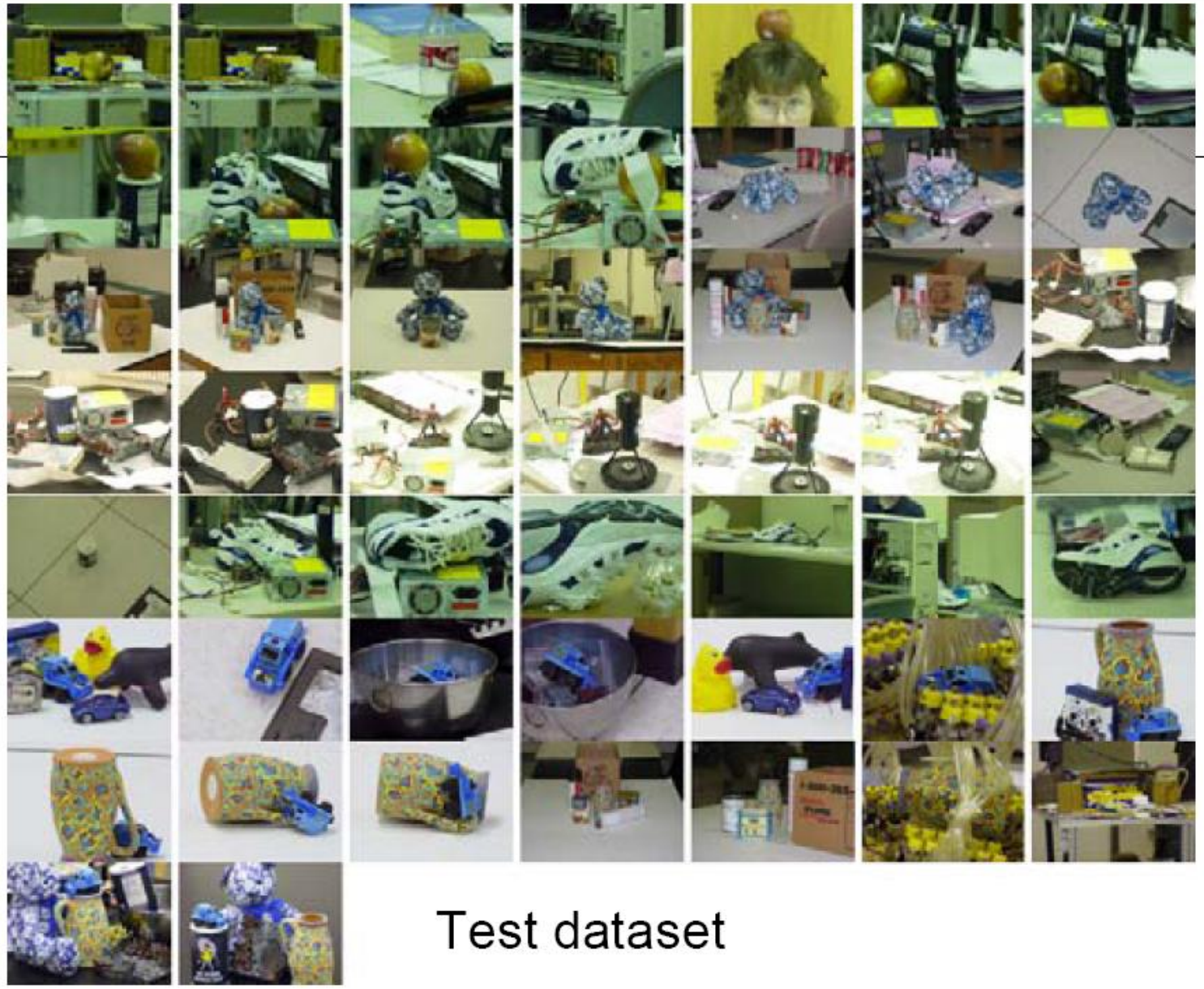






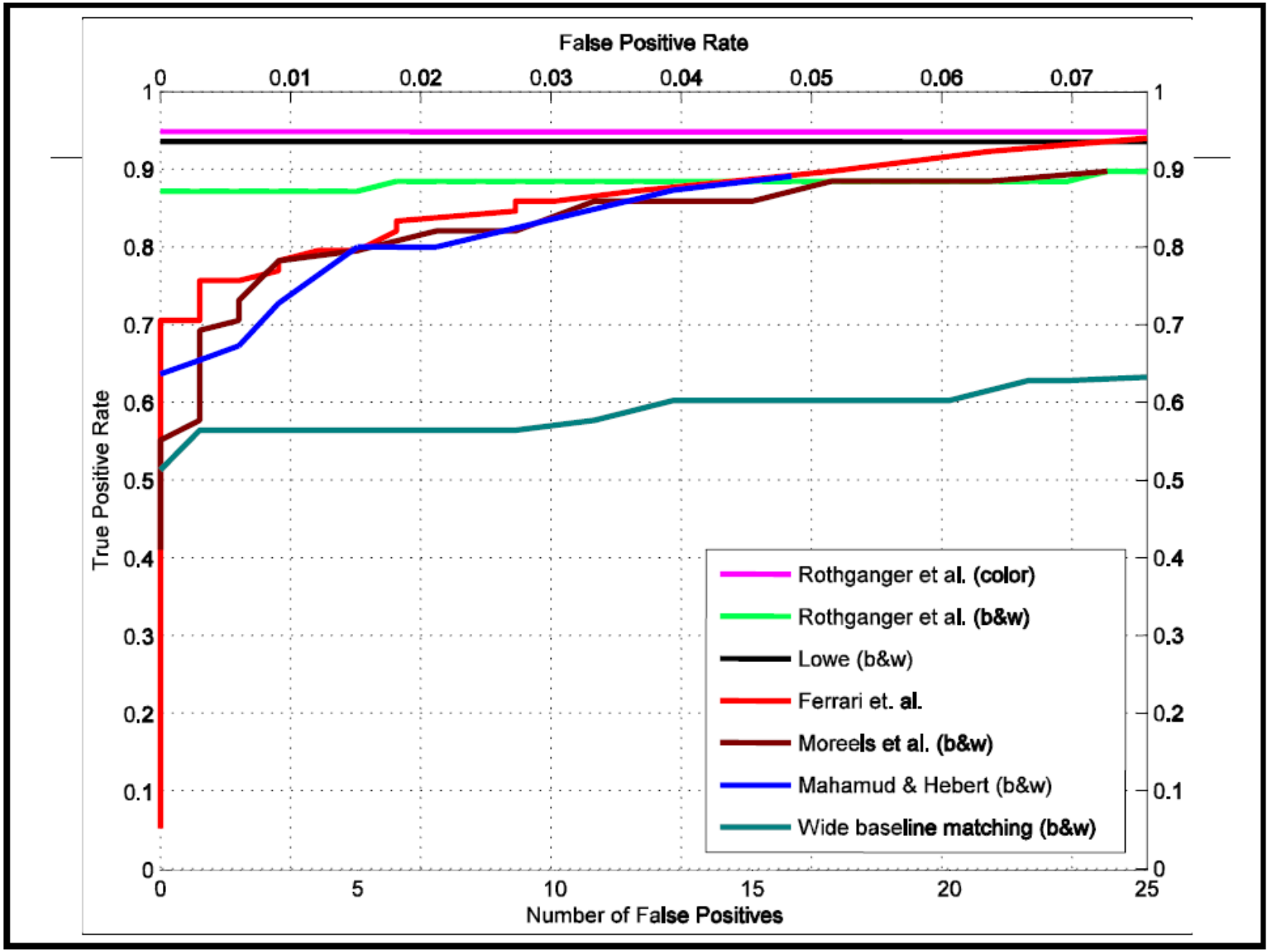
Courtesy of Rothganger et al

•Handle severe clutter



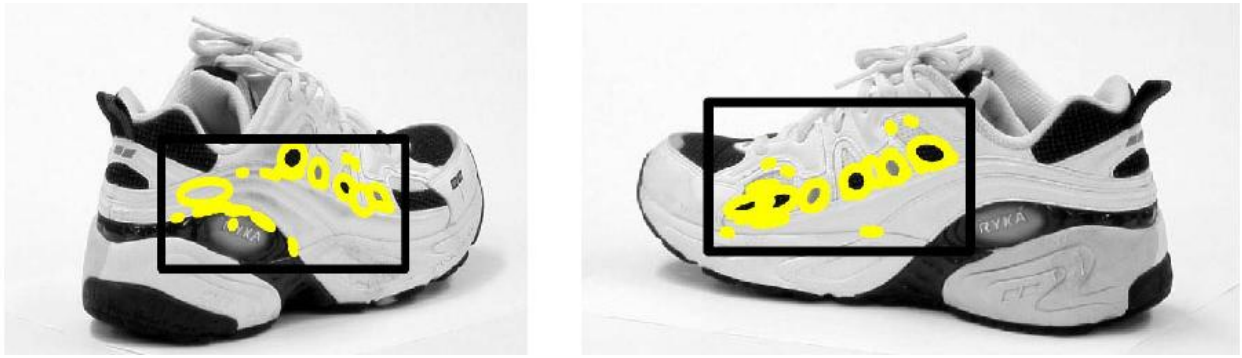
Test dataset

A comparative experiment



Other multi-view matching algorithms

-Ferrari et al. ’04, ‘06



- Lazebnick et al ’04

-Brown et al, ’05
-Toshev, Shi, Daniilidis, 07

Overview

- Single 3D object recognition
- Single view object categorization
- 3D object categorization

3D Object Categorization

Mixture of 2D single view models

- Weber et al. '00
- Schneiderman et al. '01
- Bart et al. '04

Full 3D models

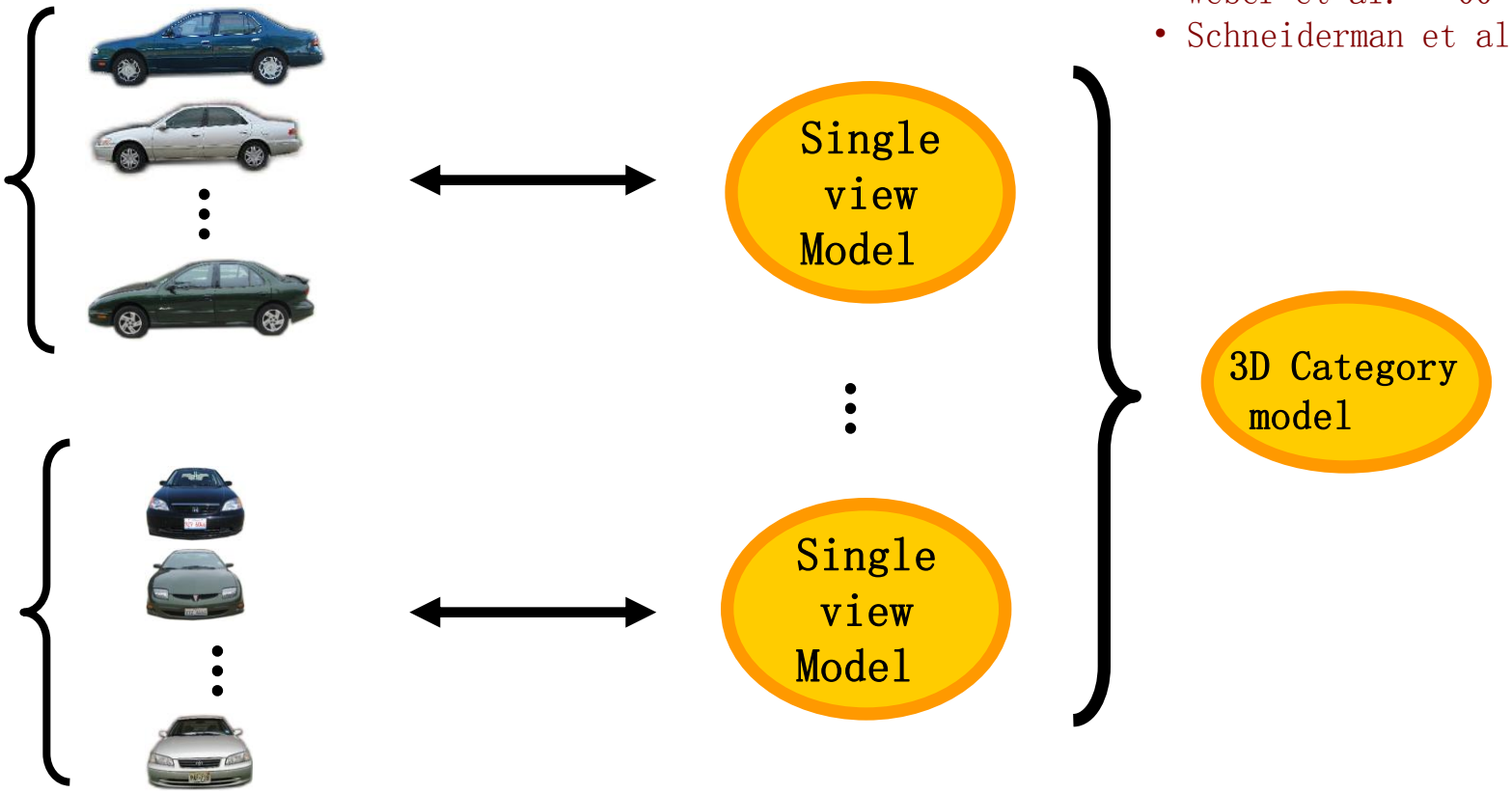
- Bronstein et al, '03
- Ruiz-Correa et al. '03,
- Funkhouser et al '03
- Capel et al '02
- Johnson & Herbert '99

Multi-view models

- Thomas et al. '06
- Savarese et al, 07, 08
- Chiu et al. '07
- Hoiem, et al., '07
- Yan, et al. '07
- Kushal, et al., '07
- Liebelt et al 08
- Sun et al 08

Mixture of single-view 2D models

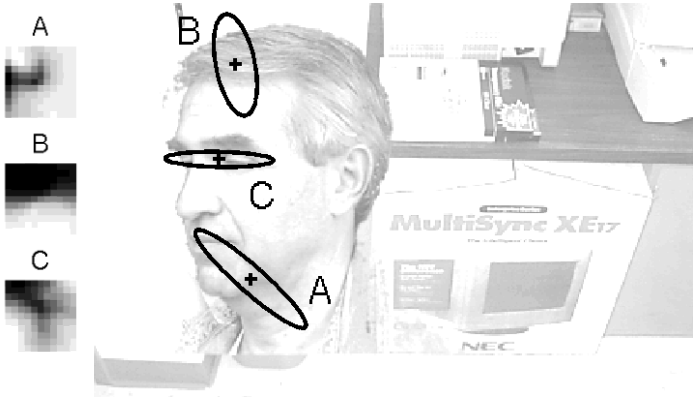
- Weber et al. ‘00
- Schneiderman et al. ’01



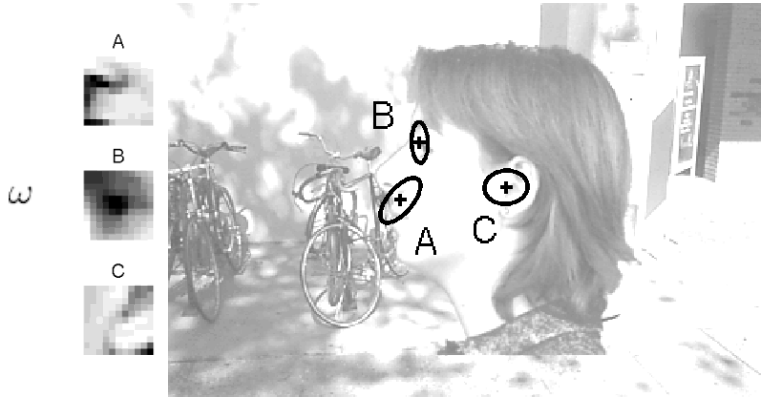
Mixture of single-view 2D models

- Mixture of 2-D models
  - Weber, Welling and Perona CVPR '00

Component 1



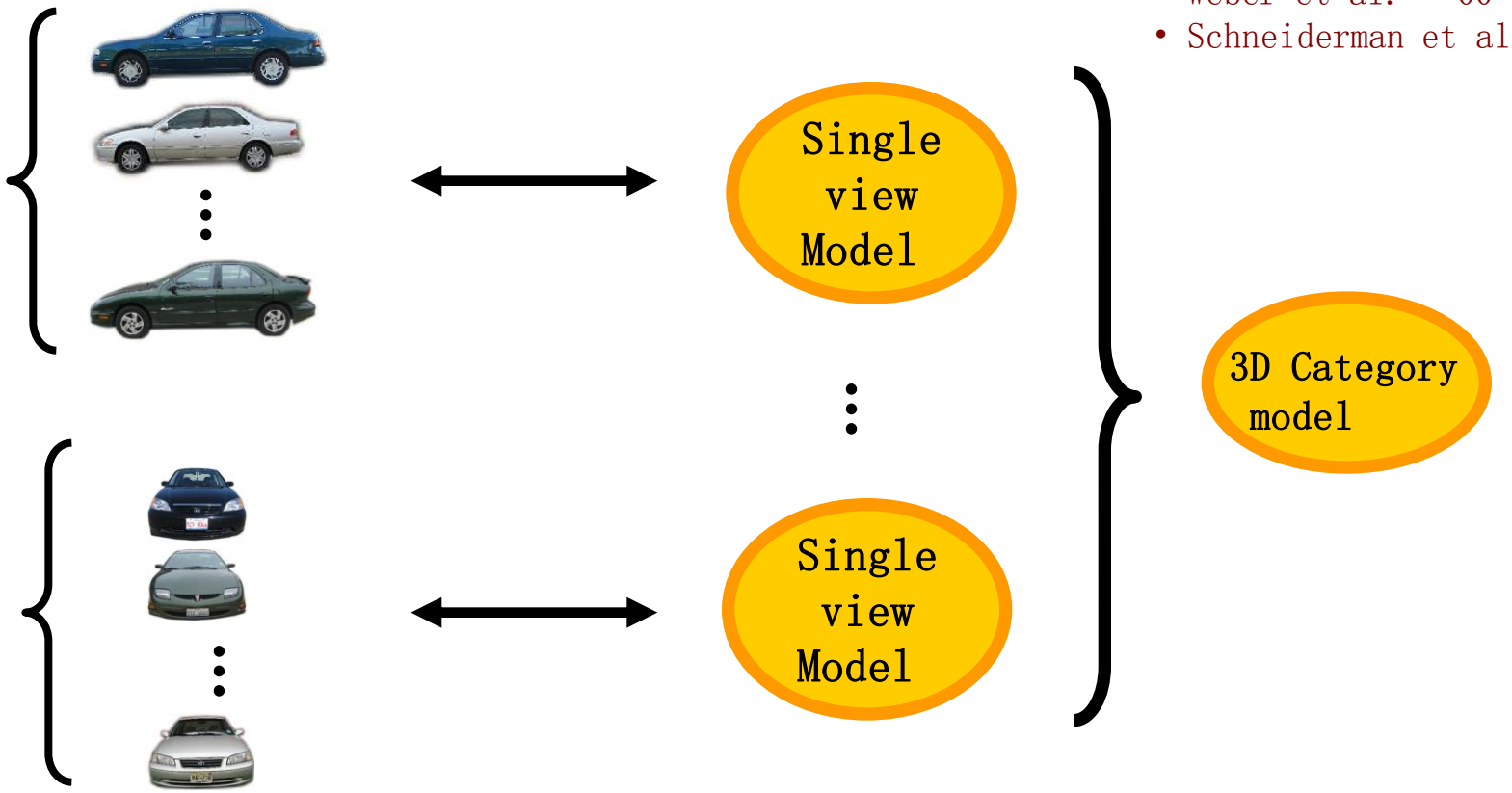
Component 2



$$p(X^o, \mathbf{x}^m, \mathbf{h}) = \sum_{\omega=1}^{\Omega} p(X^o, \mathbf{x}^m, \mathbf{h}|\omega)p(\omega).$$

Mixture of single-view 2D models

- Weber et al. '00
- Schneiderman et al. ' 01



Single view models are independent

- No information is shared
- No sense of correspondences of parts under 3D transformations

3D Object Categorization

Mixture of 2D single view models

- Weber et al. '00
- Schneiderman et al. '01
- Bart et al. '04

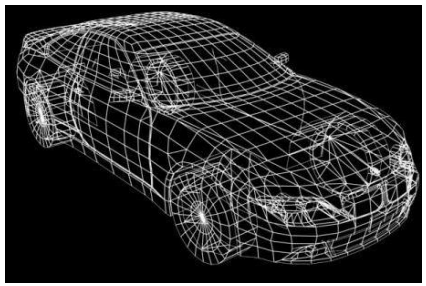
Full 3D models

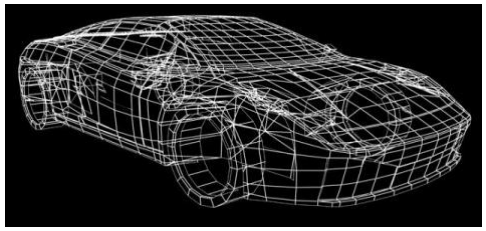
- Bronstein et al, '03
- Ruiz-Correa et al. '03,
- Funkhouser et al '03
- Capel et al '02
- Johnson & Herbert '99

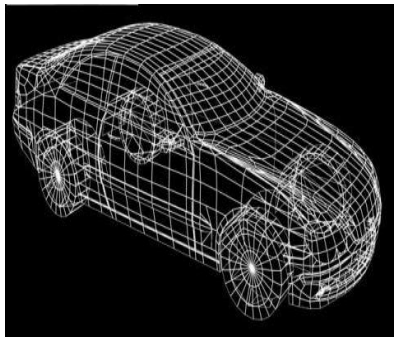
Multi-view models

- Thomas et al. '06
- Savarese et al, 07, 08
- Chiu et al. '07
- Hoiem, et al., '07
- Yan, et al. '07
- Kushal, et al., '07
- Liebelt et al 08
- Sun et al 08

Full 3D models







3D model instance

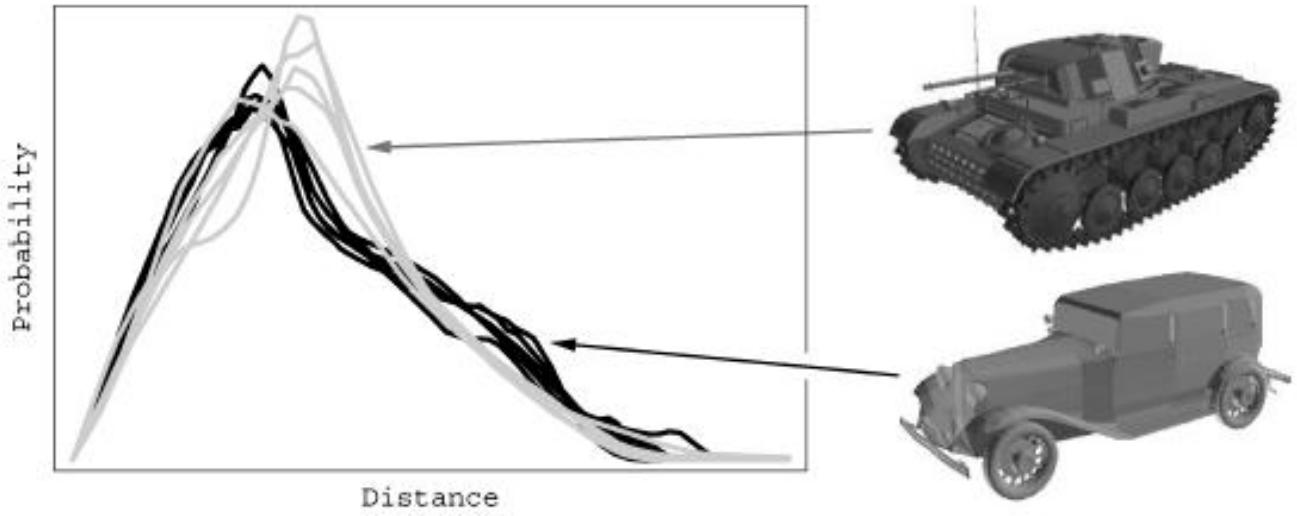
3D model instance

3D category model

- Bronstein et al, '03
- Ruiz-Correa et al. '03,
- Funkhouser et al '03
- Kazhdan et al.03
- Osada et al '02
- Capel et al '02
- Johnson & Herbert '99
- Amberg et al '08

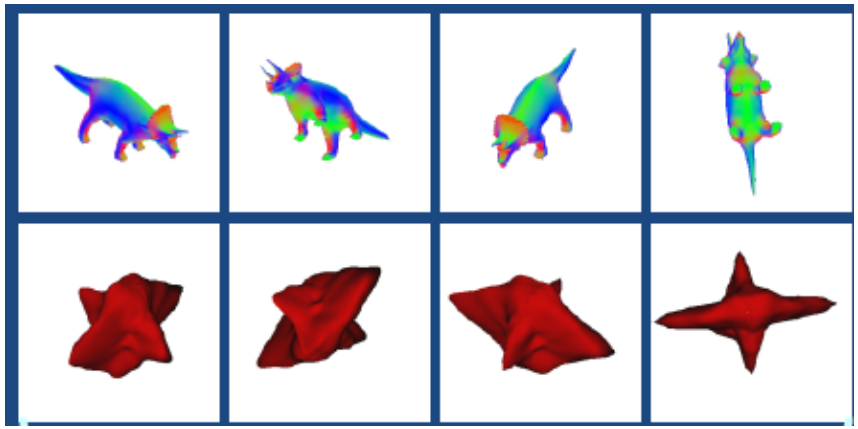
A 3D model category is built from a collection of 3D range data or CAD models

Shape distributions Osada et al 02

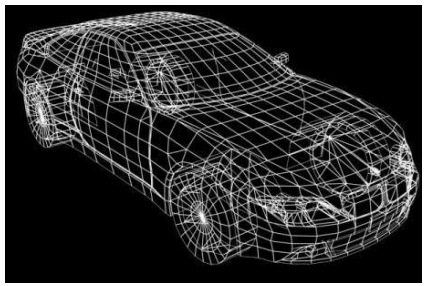


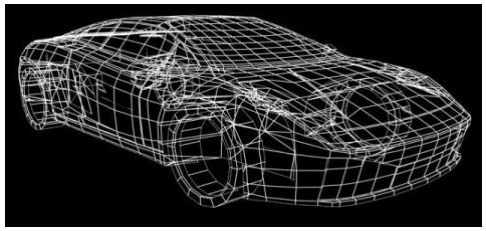
Spherical harmonics

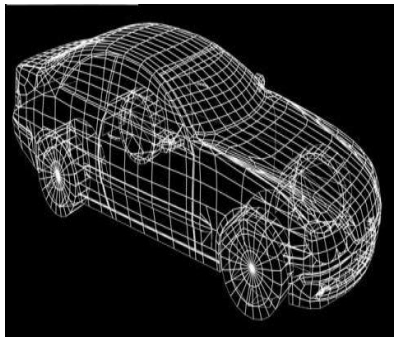
Kazhdan et al. 03



Full 3D models







3D model instance

3D model instance

3D category model

- Bronstein et al, '03
- Ruiz-Correa et al. '03,
- Funkhouser et al '03
- Kazhdan et al.03
- Osada et al '02
- Capel et al '02
- Johnson & Herbert '99
- Amberg et al '08

A 3D model category is built from a collection of 3D range data or CAD models

- Build a 3d model is expensive
- Difficult to incorporate appearance information
- Need to cope with 3D alignment (orientation, scale, etc…)

3D Object Categorization

Mixture of 2D single view models

- Weber et al. ‘00
- Schneiderman et al. ’ 01
- Bart et al. ‘04

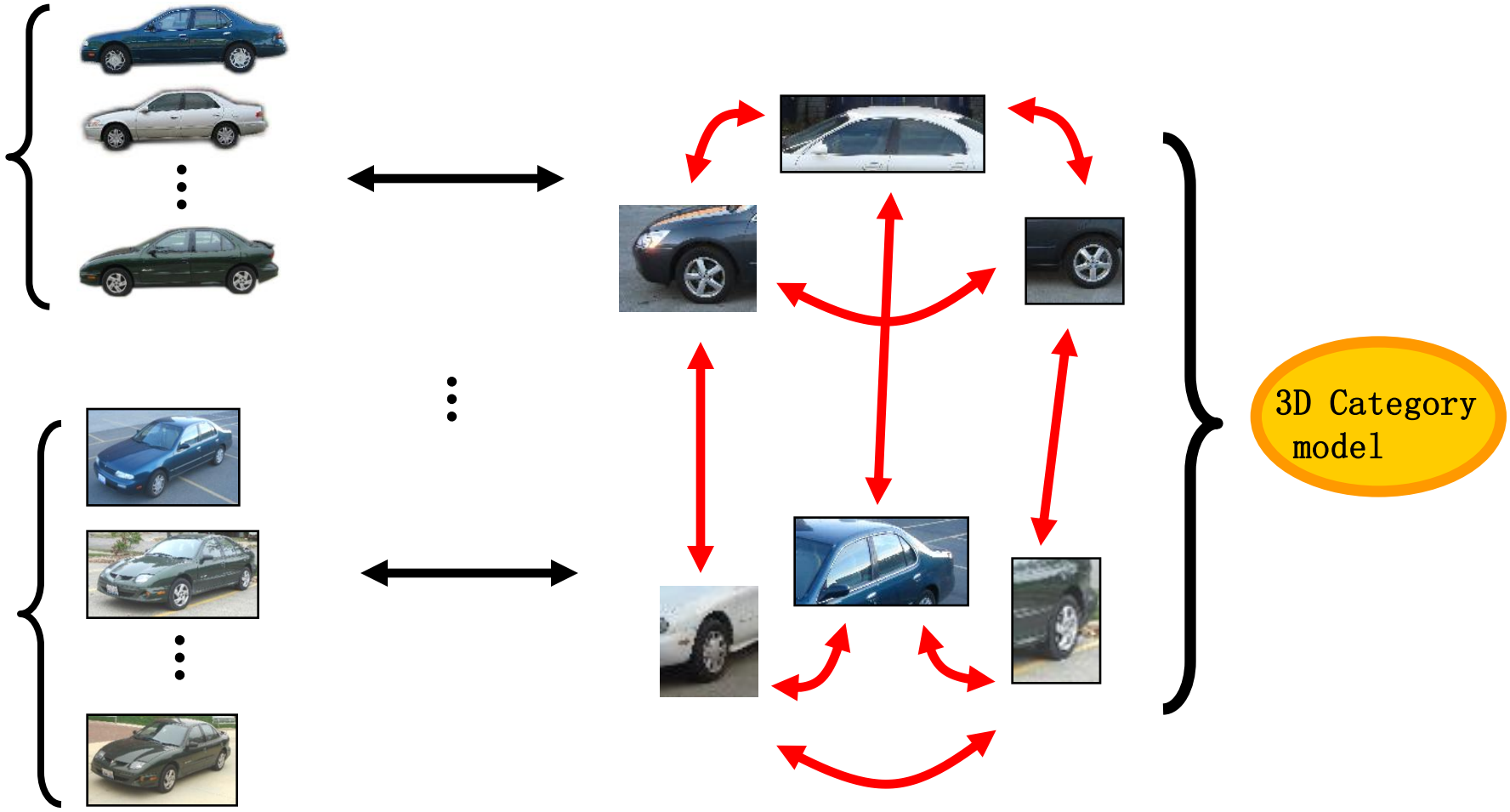
Full 3D models

- Bronstein et al, ‘03
- Ruiz-Correa et al. ’ 03,
- Funkhouser et al ’ 03
- Capel et al ’ 02
- Johnson & Herbert ‘99

Multi-view models

- Thomas et al. ‘06
- Savarese et al, 07, 08
- Chiu et al. ‘07
- Hoiem, et al., ’ 07
- Yan, et al. ’ 07
- Kushal, et al., ’ 07
- Liebelt et al 08
- Sun et al 08

Multi-view models



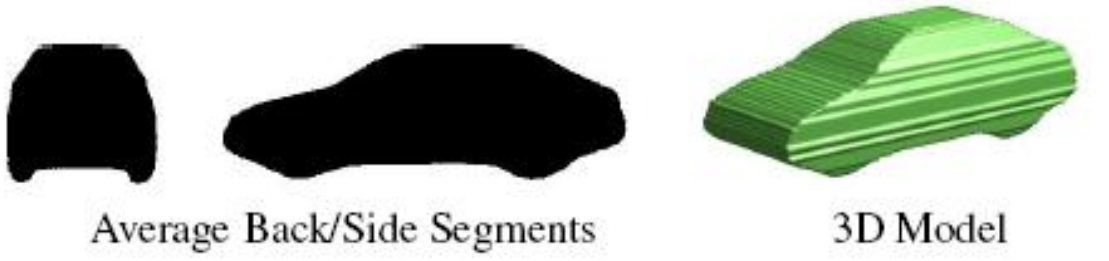
Sparse set of interest points or parts of the objects are linked across

Multi-view models by rough 3d shapes

Yan, et al. ’07

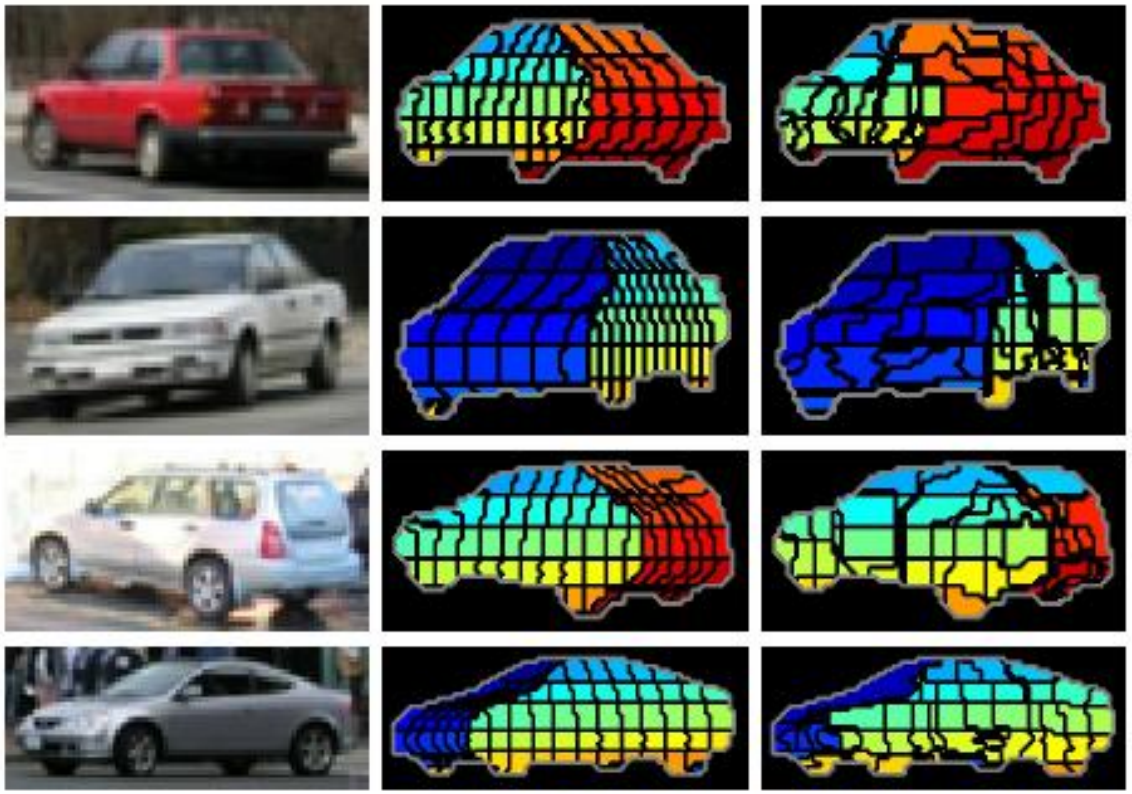
Multi-view models by rough 3d shapes

Hoiem, et al., ' C



Average Back/Side Segments

3D Model



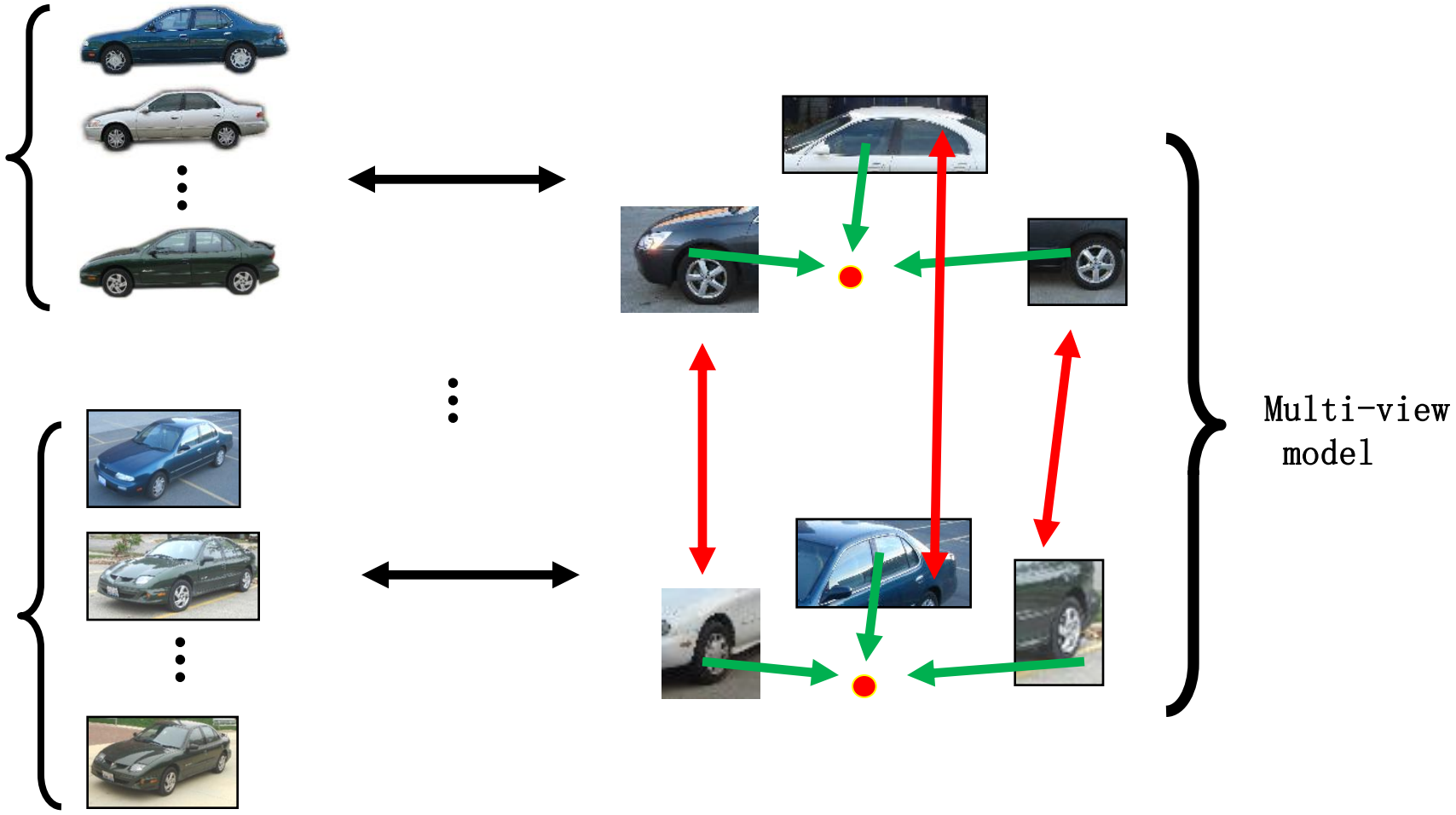
Image

Initial Labels

Deformed Labels

Multi-view models by ISM representations

[Thomas et al. ’06]



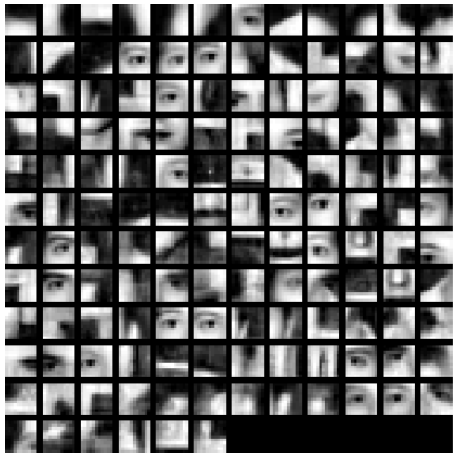
Sparse set of interest points or parts of the objects are linked across

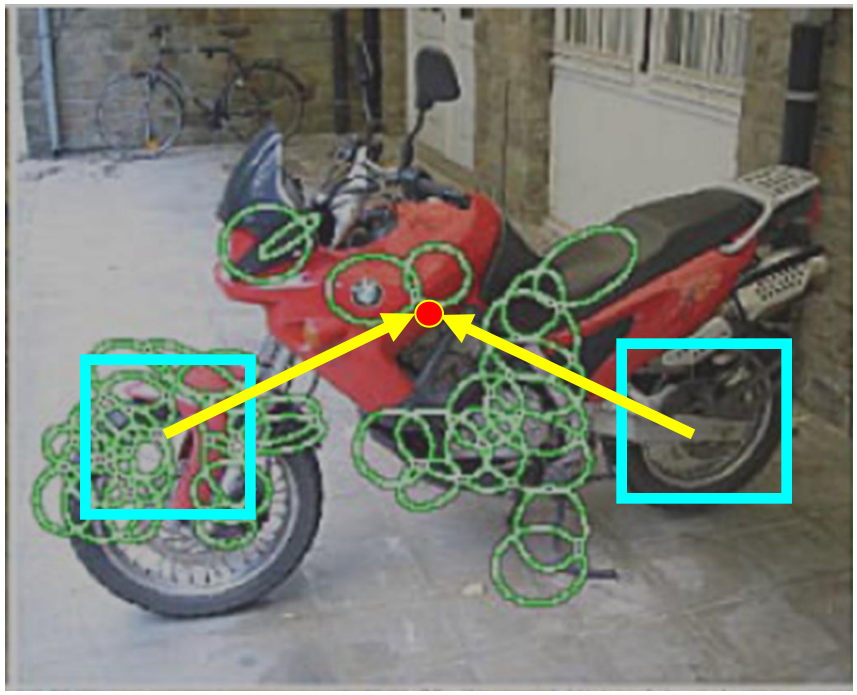
Multi-view models by ISM representations

[Thomas et al. ’06]

ISM representation

Leibe, Leonardis, and Schiele, ECCV Workshop on Statistical Learning in Computer Vision 2004





visual codeword with displacement vectors

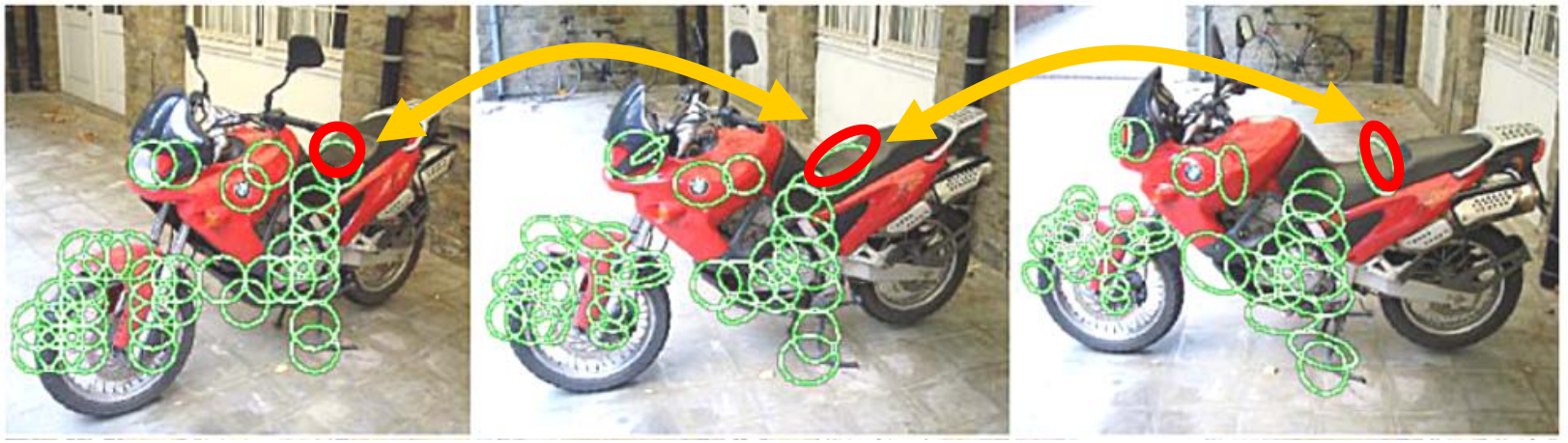
- Visual codebook is used to index votes for object position
- Generalized Hough transform

Combining multi-views and ISM models

[Thomas et al. ’06]

Region tracks

[Ferrari et al. ’04, ‘06]



Courtesy of Thomas et al. 06

Set of *region-tracks* connecting model views
Each track is composed of image regions of a single physical surface patch
along the model views in which it is visible.

Properties

	Single view/ Mixture	Multi-view
View point invariant	X	√
No supervision	√	X Category View point
# Categories	~300	2
Share information across views	X	√
View synthesis	X	X
Pose estimation	X → √	X