Lecture 14
Announcements

- Midterm review on Friday
- Midterm next Monday!
  - 1:30-3:50pm in this class room
  - previous years midterms are available!
  - Open notes/books but no electronics!

Lecture 14
Visual recognition

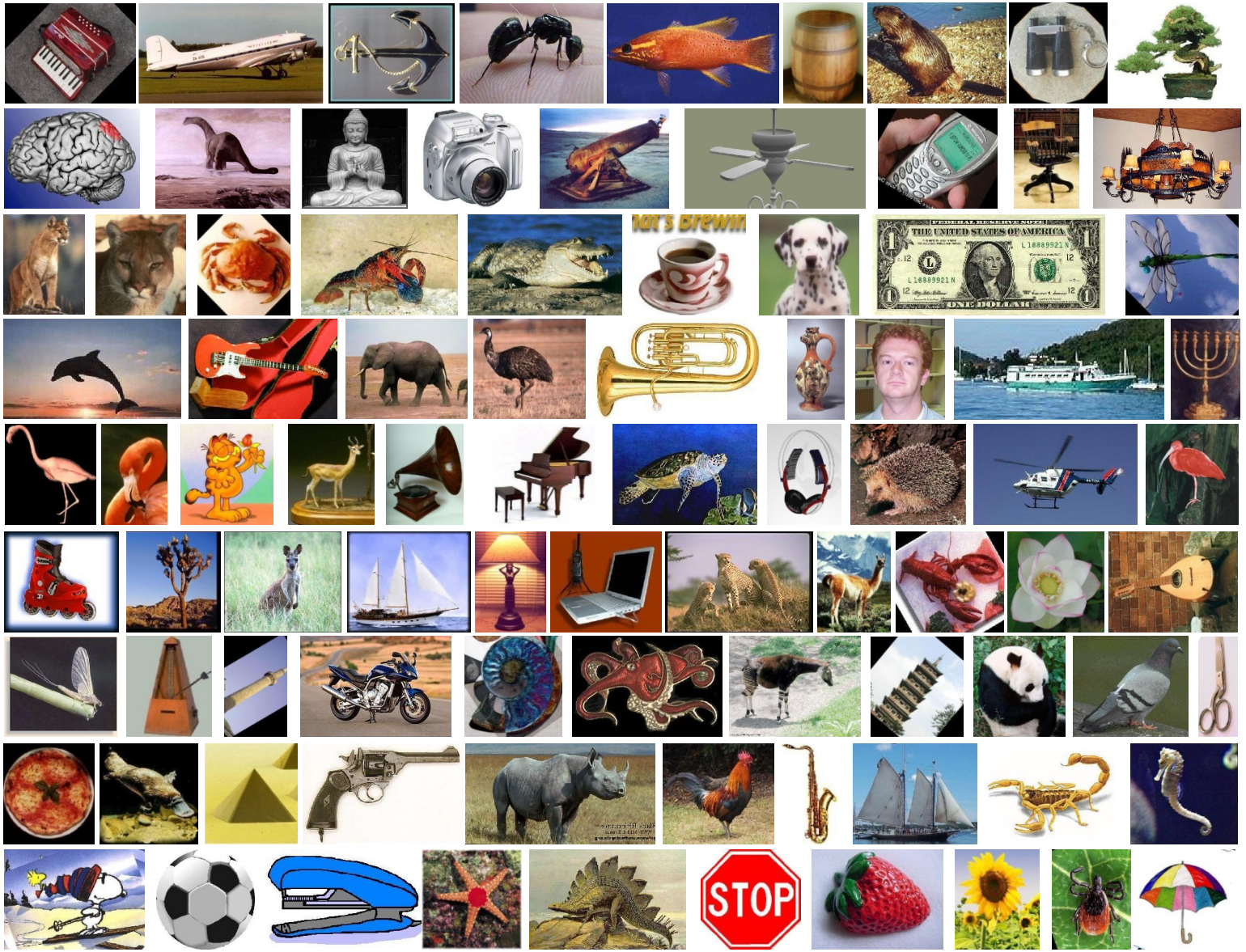
- Datasets
- 3D object detection

Caltech 101

Learning generative visual models from few training examples: an incremental Bayesian approach tested on 101 object categories. L. Fei-Fei, R. Fergus, and P. Perona. CVPR 2004, Workshop on Generative-Model Based Vision. 2004

- Pictures of objects belonging to 101 categories.
- About 40 to 800 images per category. Most categories have about 50 images.
- The size of each image is roughly 300 x 200 pixels.

Caltech 101 images



Caltech-101: Drawbacks

- Smallest category size is 31 images: $N_{train} \leq 30$

- Too easy?
  - left-right aligned
  - Rotation artifacts
  - Saturated performance





Caltech-256

Griffin, Gregory and Holub, Alex and Perona, Pietro (2007) *Caltech-256 Object Category Dataset*. California

- Smallest category size now 80 images
- About 30K images

- Harder
  - Not left-right aligned
  - No artifacts
  - More categories

- New and larger clutter category




Caltech 256 images

baseball-bat

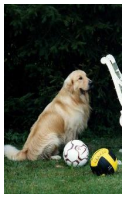




dog







basketball-hoop







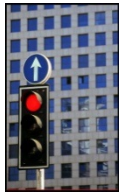
kayac



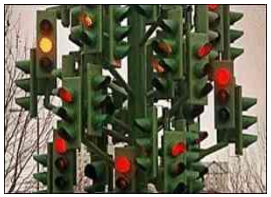




traffic light





J. Deng, et al. (2009-2016)

From prof. Li's group @ Stanford!

Largest dataset for object categories up to date

- ~20K categories;
- 14 million images;
- ~700im/categ;
- free to public at **www.image-net.org**

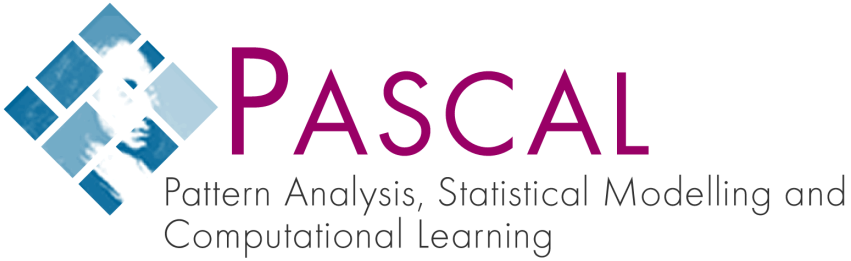
IMAGENET is a knowledge ontology

- Taxonomy

- S: (n) Eskimo dog, **husky** (breed of heavy-coated Arctic sled dog)
  - *direct hypernym* / ***inherited hypernym*** / *sister term*
    - S: (n) working dog (any of several breeds of usually large powerful dogs bred to work as draft animals and guard and guide dogs)
      - S: (n) dog, domestic dog, Canis familiaris (a member of the genus Canis (probably descended from the common wolf) that has been domesticated by man since prehistoric times; occurs in many breeds) *"the dog barked all night"*
        - S: (n) canine, canid (any of various fissiped mammals with nonretractile claws and typically long muzzles)
          - S: (n) carnivore (a terrestrial or aquatic flesh-eating mammal) *"terrestrial carnivores have four or five clawed digits on each limb"*
            - S: (n) placental, placental mammal, eutherian, eutherian mammal (mammals having a placenta; all mammals except monotremes and marsupials)
              - S: (n) mammal, mammalian (any warm-blooded vertebrate having the skin more or less covered with hair; young are born alive except for the small subclass of monotremes and nourished with milk)
                - S: (n) vertebrate, craniate (animals having a bony or cartilaginous skeleton with a segmented spinal column and a large brain enclosed in a skull or cranium)
                  - S: (n) chordate (any animal of the phylum Chordata having a notochord or spinal column)
                    - S: (n) animal, animate being, beast, brute, creature, fauna (a living organism characterized by voluntary movement)
                      - S: (n) organism, being (a living thing that has (or can develop) the ability to act or function independently)
                        - S: (n) living thing, animate thing (a living (or once living) entity)
                          - S: (n) whole, unit (an assemblage of parts that is regarded as a single entity) *"how big is that part compared to the whole?"; "the team is a unit"*
                            - S: (n) object, physical object (a tangible and visible entity; an entity that can cast a shadow) *"it was full of rackets, balls and other objects"*
                              - S: (n) physical entity (an entity that has physical existence)
                                - S: (n) entity (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))

The PASCAL Visual Object Classes (VOC) Dataset and Challenge (2005-2012)

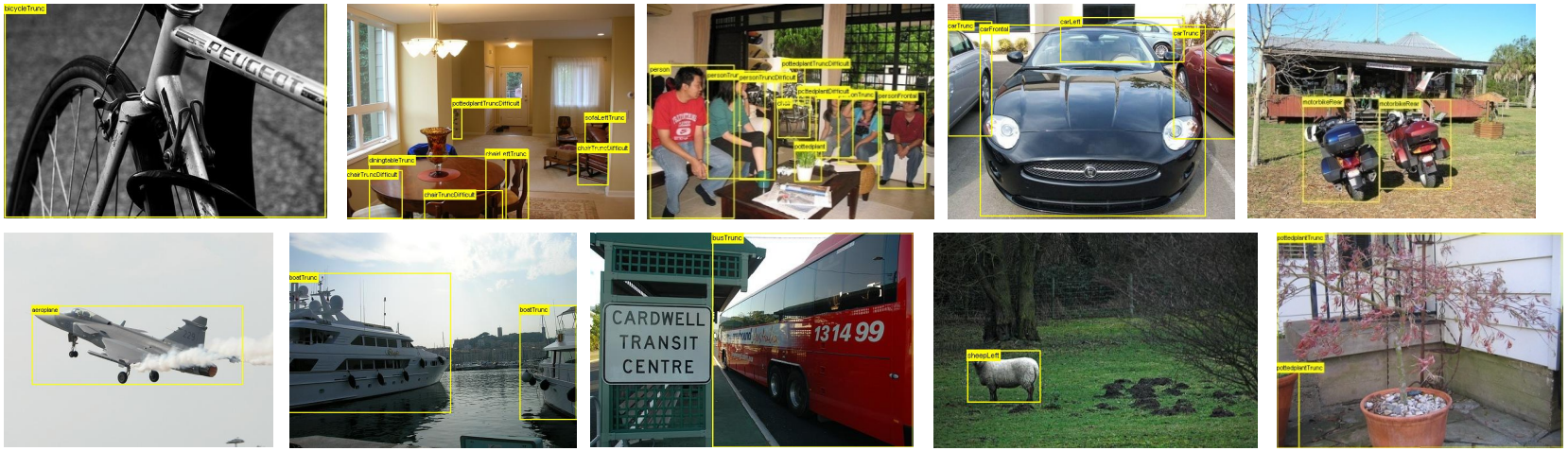
Mark Everingham
Luc Van Gool
Chris Williams
John Winn
Andrew Zisserman



Dataset Content

- 20 classes: aeroplane, bicycle, boat, bottle, bus, car, cat, chair, cow, dining table, dog, horse, motorbike, person, potted plant, sheep, train, TV

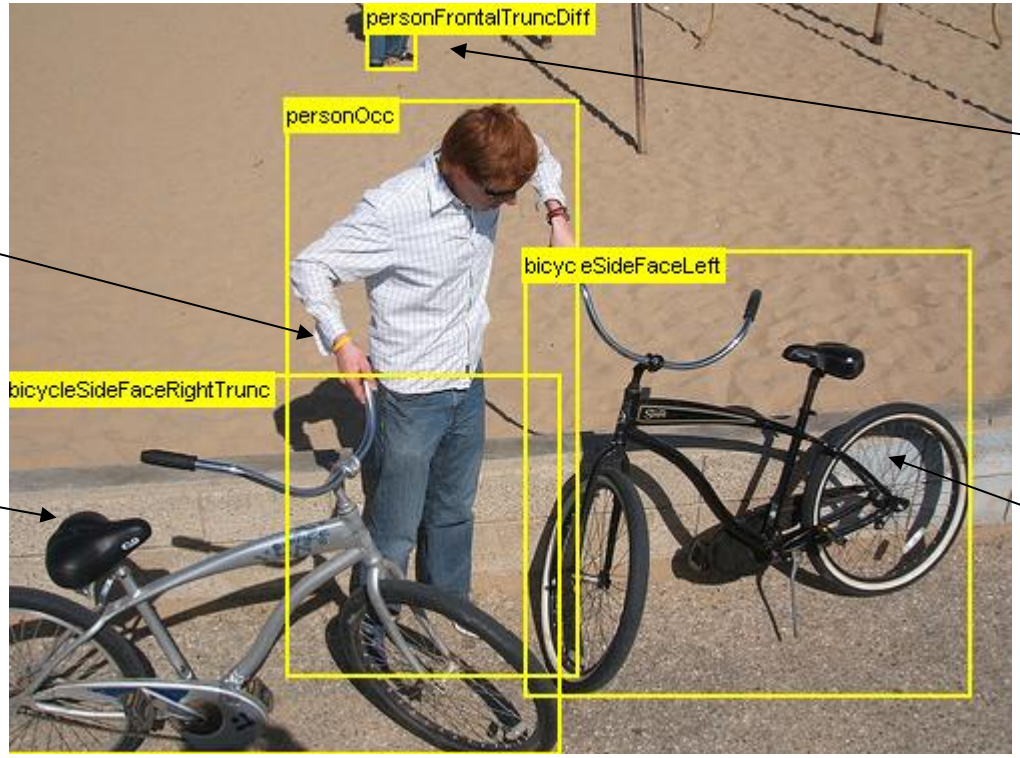
- Real images downloaded from flickr, not filtered for “quality”



- Complex scenes, scale, pose, lighting, occlusion, ...

Annotations

- Complete annotation of all objects
- Annotated in one session with written guidelines



Occluded
Object is significantly occluded within BB

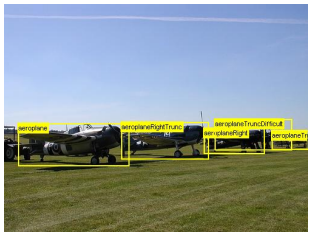
Truncated
Object extends beyond BB

Difficult
Not scored in evaluation

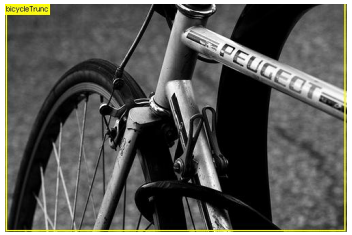
Pose
Facing left

Examples

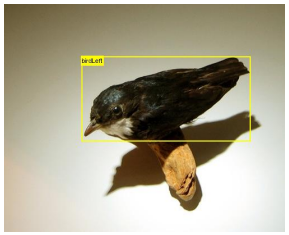
Aeroplane



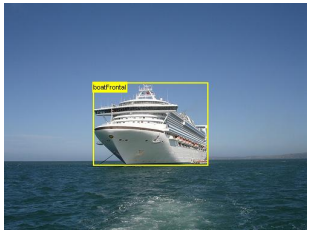
Bicycle

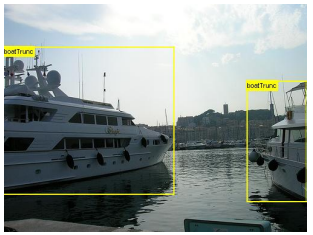


Bird

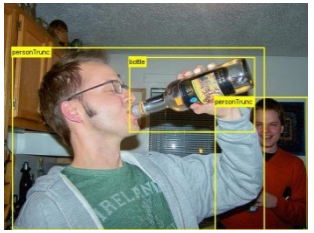


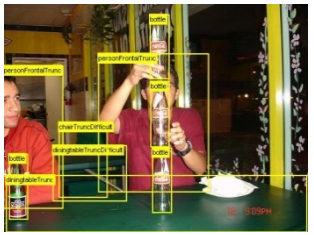
Boat



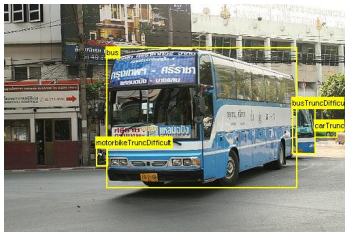


Bottle

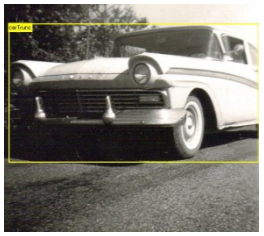




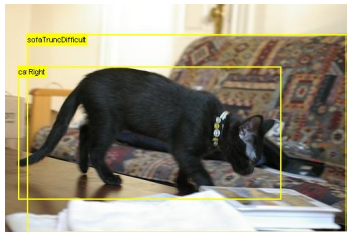
Bus



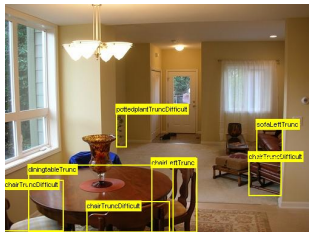
Car

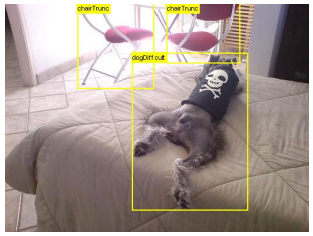


Cat

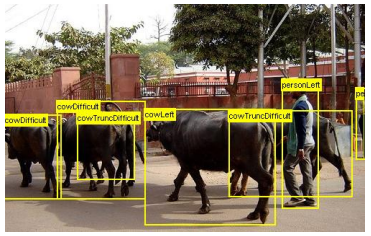


Chair





Cow



History

	Images	Objects	Classes	Notes
2005	2,232	2,871	4	*Collection of existing and some new data.*
2006	5,304	9,507	10	*Completely new dataset from flickr (+MSRC)*
2007	9,963	24,640	20	*Increased classes to 20. Introduced tasters.*
2008	8,776	20,739	20	*Added "occlusion" flag."*
⇩				
2012	11,530	27,450	20	*Added segmentation masks*

- Challenge: annotation of test set is withheld until after challenge

PASCAL 3D+

Xiang, Mottaghi, Savarese (2014)

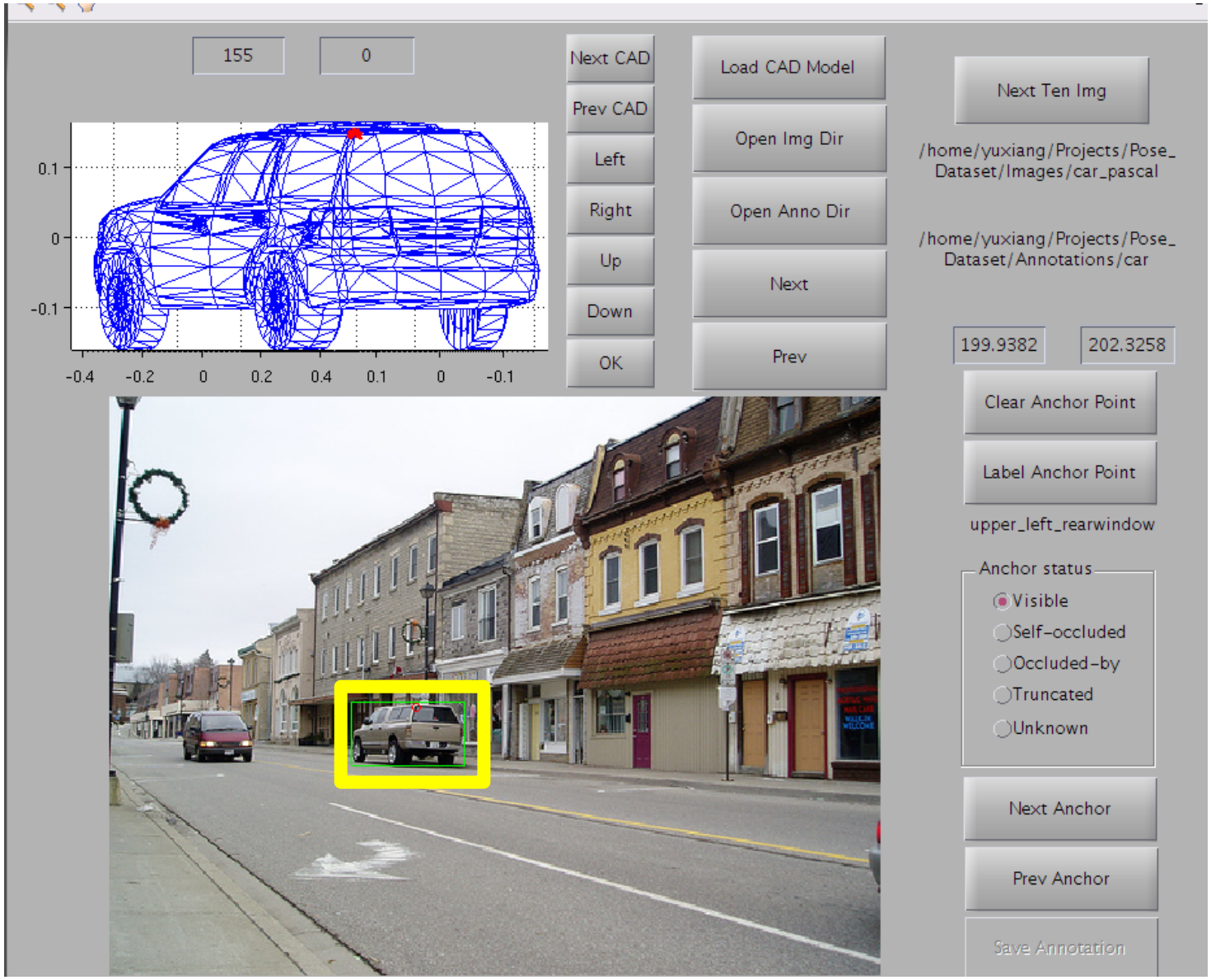
- 12 rigid categories from PASCAL VOC are annotated with 3D pose and aligned with 3D cad models
- Benchmark for continuous 3D pose estimation and shape recovery of object categories








PASCAL 3D+


155
0
Next CAD
Prev CAD
Left
Right
Up
Down
OK
Load CAD Model
Open Img Dir
Open Anno Dir
Next
Prev
Next Ten Img
/home/yuxiang/Projects/Pose_Dataset/Images/car_pascal
/home/yuxiang/Projects/Pose_Dataset/Annotations/car
199.9382
202.3258
Clear Anchor Point
Label Anchor Point
upper_left_rearwindow
Anchor status
Visible
Self-occluded
Occluded-by
Truncated
Unknown
Next Anchor
Prev Anchor
Save Annotation
0.1
0
-0.1
-0.4
-0.2
0
0.2
0.4
0.1
0
-0.1

ShapeNet3D

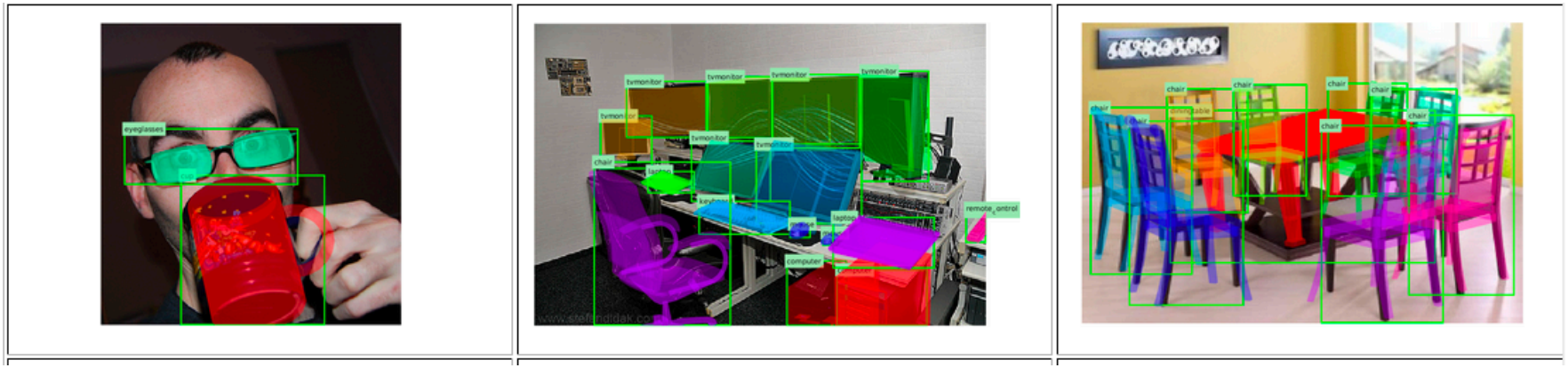
Chang, et al., (2015)

- Covers 55 common object categories with about 51,300 unique 3D models.
- 12 object categories of PASCAL 3D+

ObjectNet3D

Xiang, et al., (2016)

- ~100 rigid categories from ImageNet and PASCAL
- Annotated with 3D pose and aligned with 3D cad models from ShapeNet



OpenSurfaces Bell et al. , 2014-2015

Materials

Reflectances

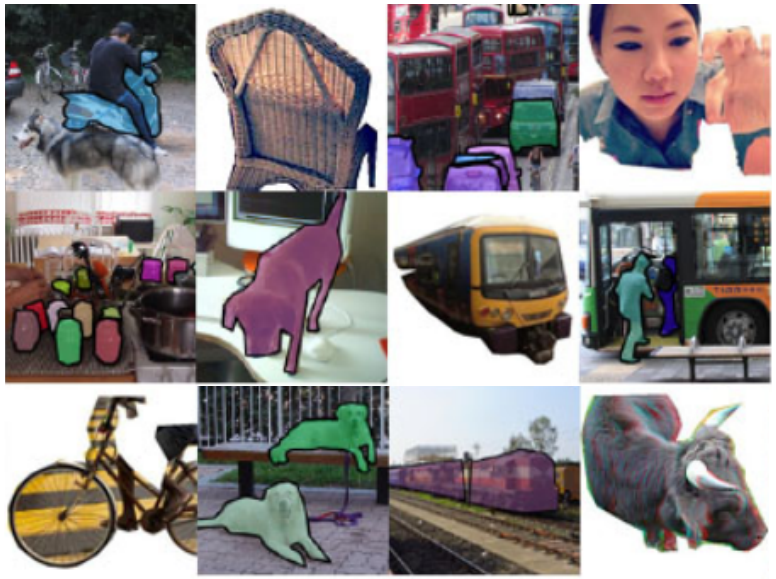
Textures

+80K annotated images of materials

MSR COCO dataset

Lin et al 2014-2016

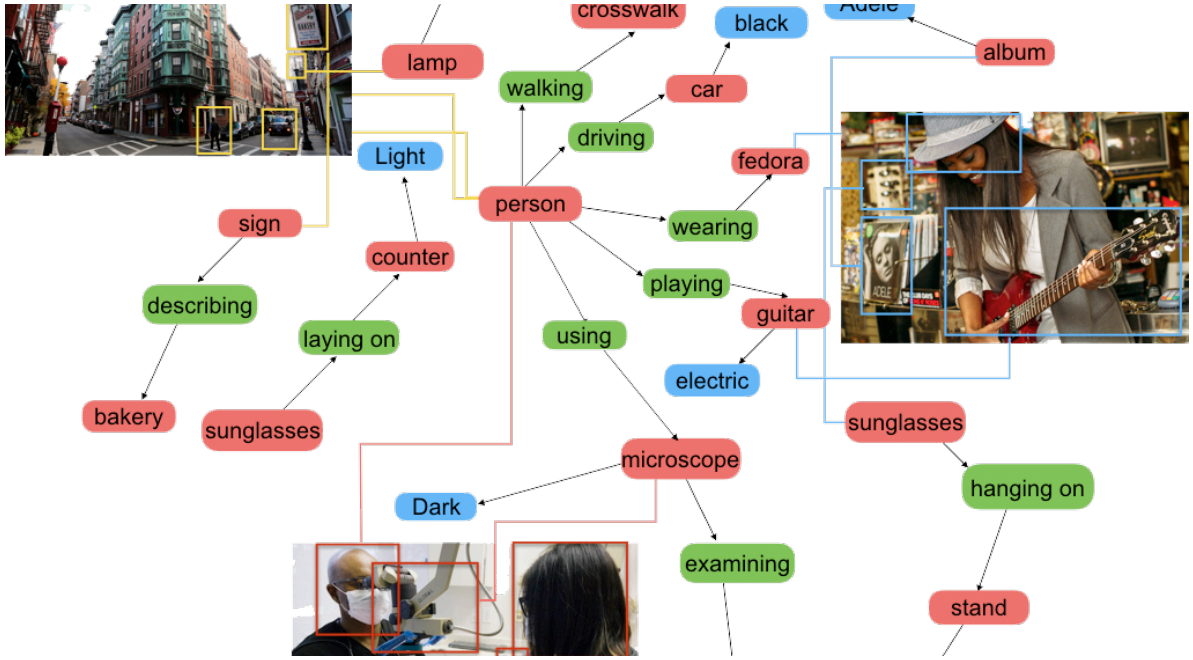
- Dataset for image recognition, segmentation, and captioning
- Features:
  - Object segmentation
  - Recognition in Context
  - Multiple objects per image
  - More than 300,000 images
  - More than 2 Million instances
  - 80 object categories
  - 5 captions per image



Visual genome

Krishna et al.2015- 2016
From prof. Li's group @ Stanford!

Knowledge-based dataset that connects structured image concepts to language



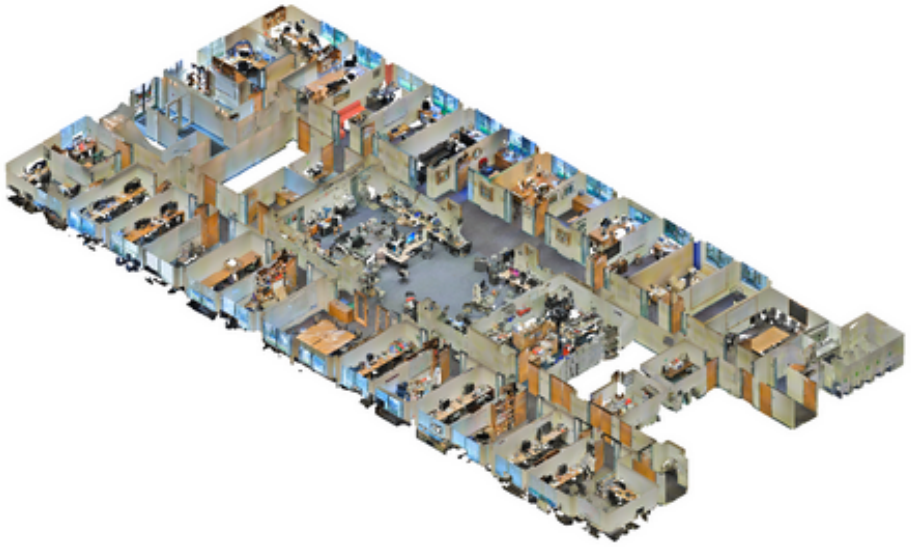
108,249 Images
4.2 Million Region Descriptions
1.7 Million Visual Question Answers
2.1 Million Object Instances
1.8 Million Attributes
1.8 Million Relationships

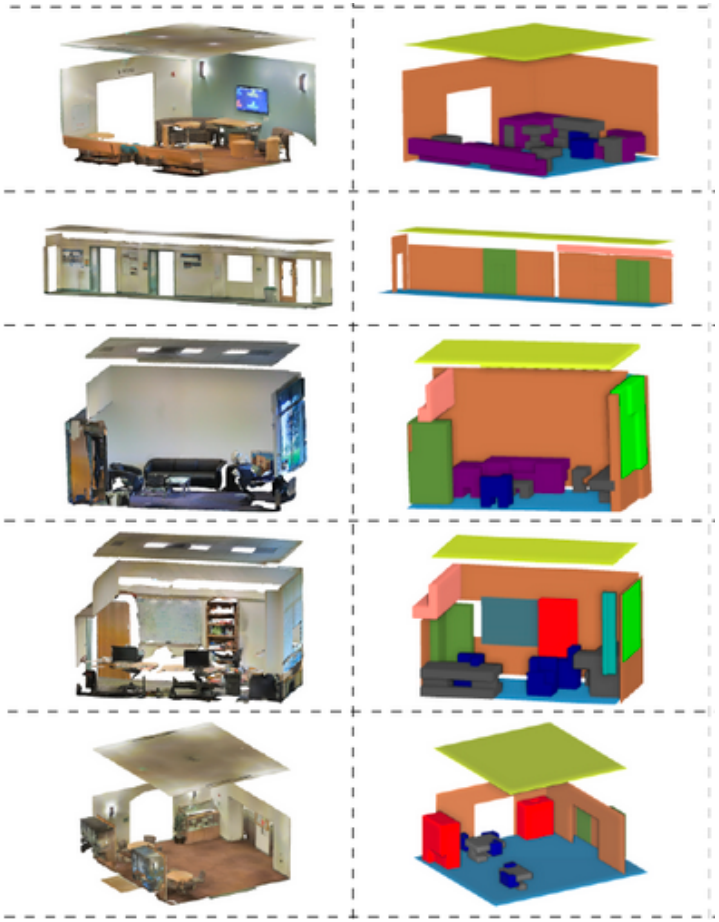
Scene understanding

LabelMe, Russell et al., 2005

3D scene understanding

Stanford 2D-3D-Semantics Dataset, Armeni et al., 2016



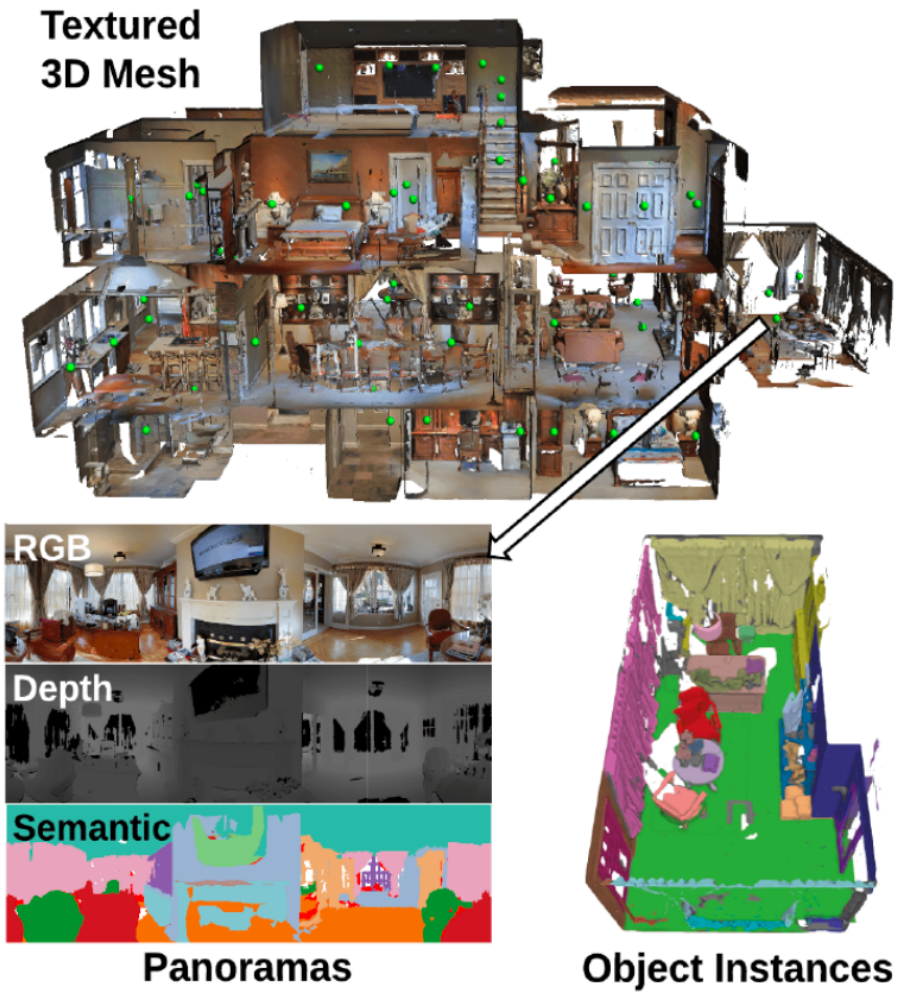


6 buildings ~500 rooms ~6000m^2
area ~6000 Building Elements

http://buildingparser.stanford.edu/dataset.html

3D scene understanding

Stanford, Princeton, MatterPort 2017



- 10,800 aligned 3D panoramic views (RGB + depth per pixel)
- 194,400 RGB + depth images
- 90 -scale scenes.
- Scenes were captured with Matterport's Pro 3D Camera

More Datasets….

UIUC Cars (2004)
S. Agarwal, A. Awan, D. Roth

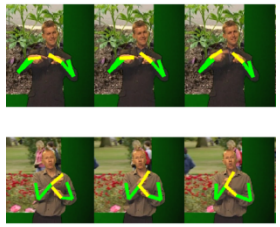
CMU/VASC Faces (1998)
H. Rowley, S. Baluja, T. Kanade

FERET Faces (1998)
P. Phillips, H. Wechsler, J. Huang, P. Raus

COIL Objects (1996)
S. Nene, S. Nayar, H. Murase

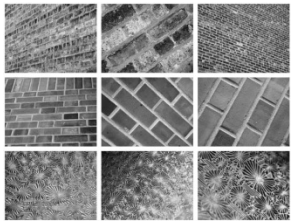
MNIST digits (1998-10)
Y LeCun & C. Cortes

KTH human action (2004)
I. Leptev & B. Caputo

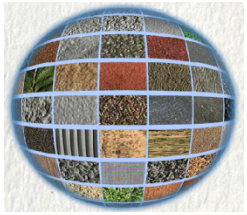


Sign Language (2008)
P. Buehler, M. Everingham, A. Zisserman

Segmentation (2001)
D. Martin, C. Fowlkes, D. Tal, J. Malik.

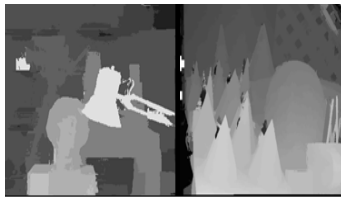


3D Textures (2005)
S. Lazebnik, C. Schmid, J. Ponce



CuRRET Textures (1999)
K. Dana B. Van Ginneken S. Nayar J. Koenderink

CAVIAR Tracking (2005)
R. Fisher, J. Santos-Victor J. Crowley



Middlebury Stereo (2002)
D. Scharstein R. Szeliski

Lecture 14
Visual recognition

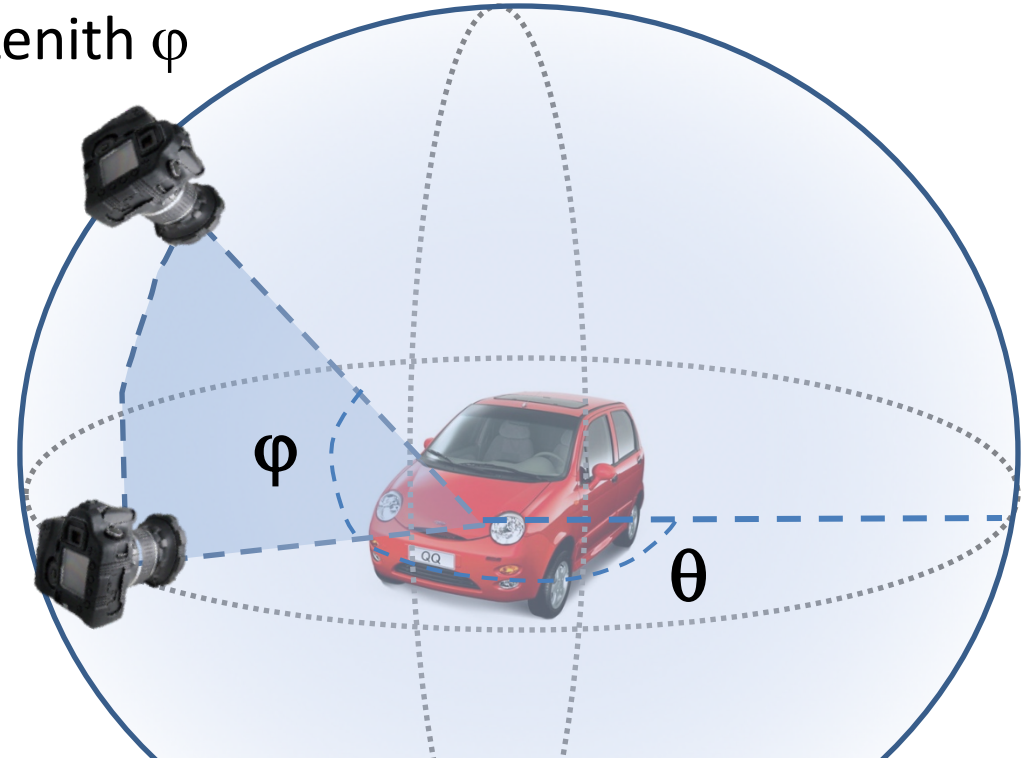
- 3D object detection
  - Introduction
  - Single instance 3D object detectors
  - Generic 3D object detectors

3D object detection

Properties of a 3D object detector

Azimuth θ, Zenith φ

Viewing sphere

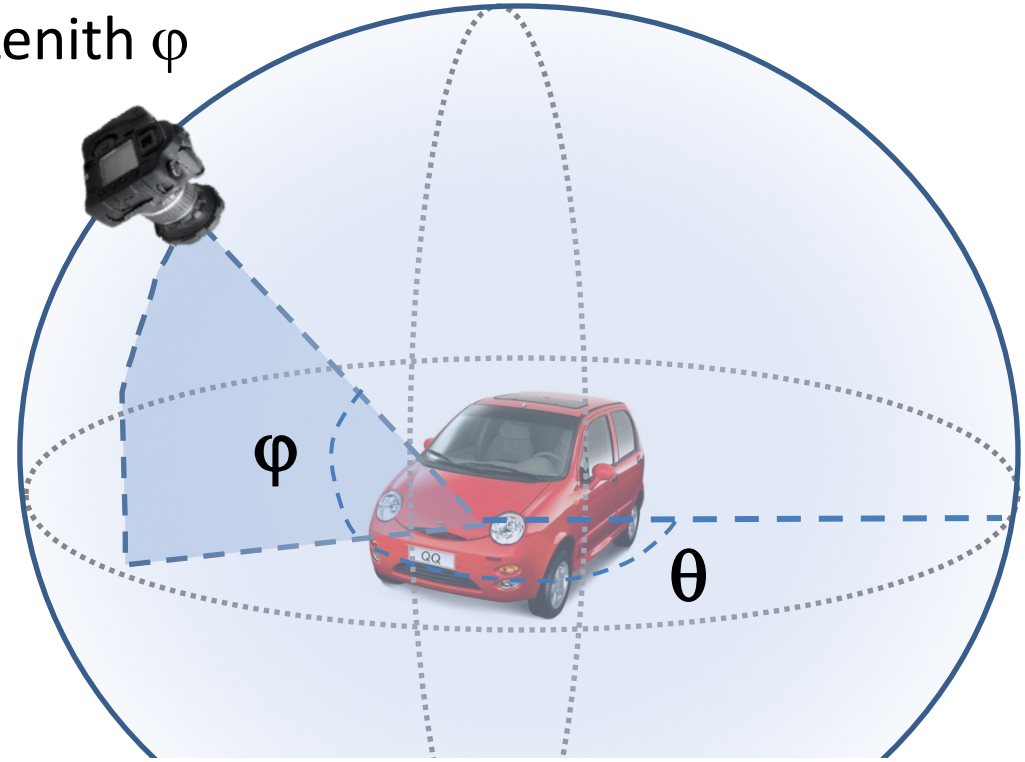


- Detect objects under generic view points
- Estimate object pose & 3D shape

Properties of a 3D object detector

Viewing sphere

Azimuth θ, Zenith φ

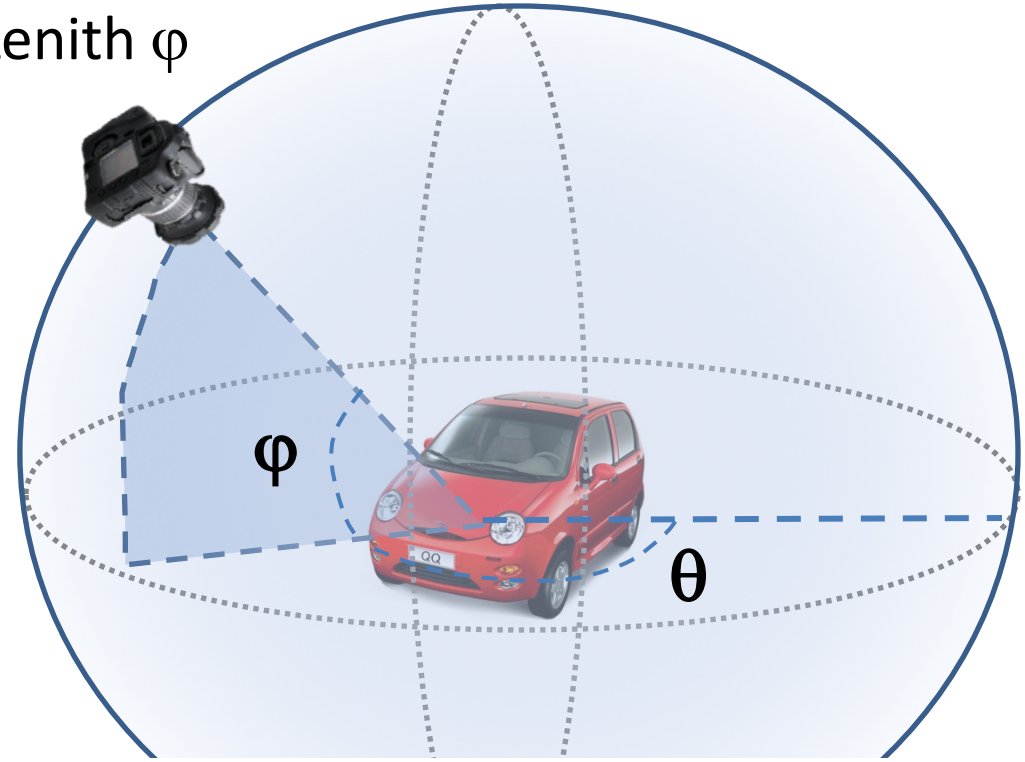


- Detect objects under generic view points
- Estimate object pose & 3D shape
- Work at different levels of specificity

Properties of a 3D object detector

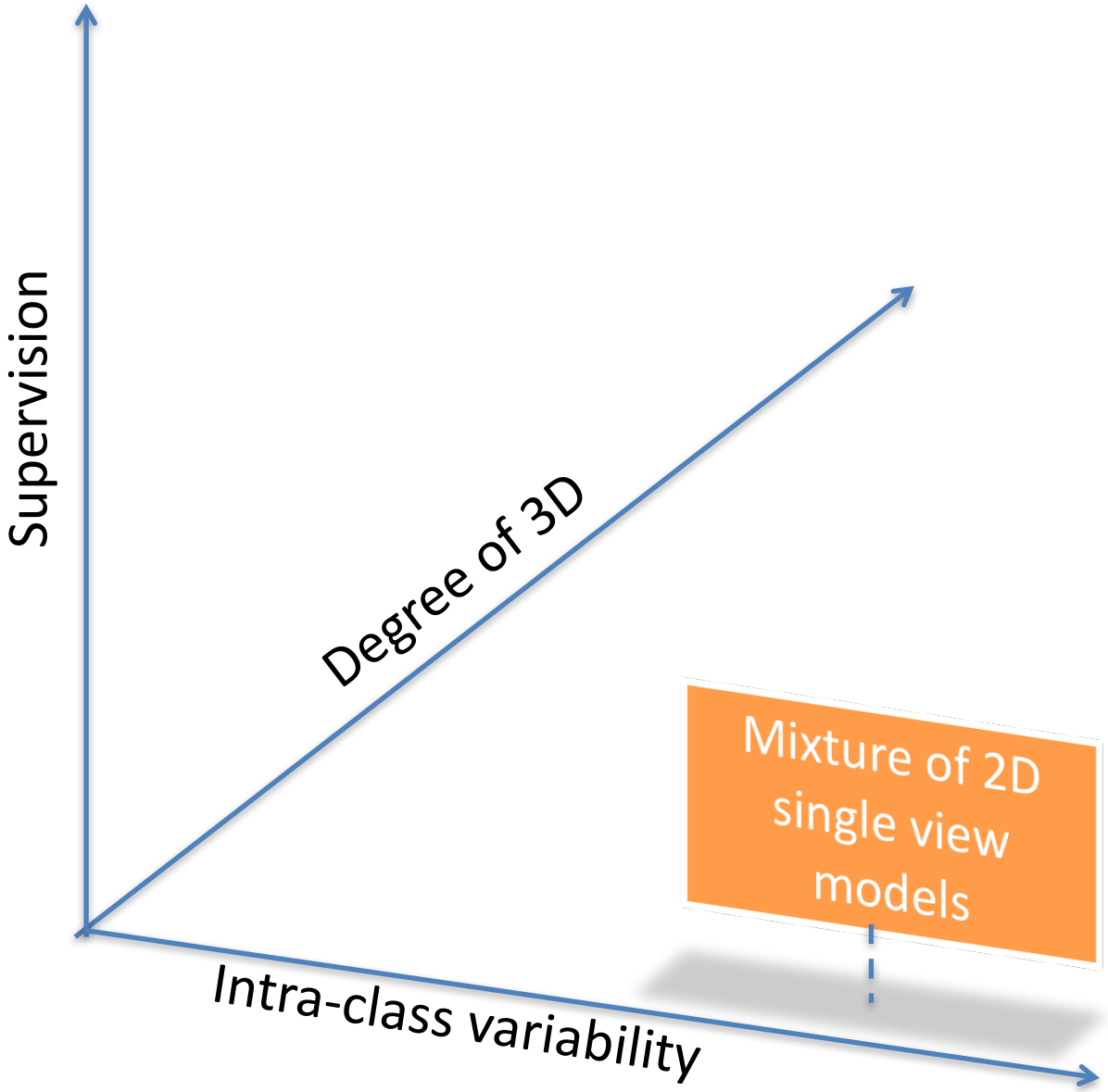
Azimuth θ, Zenith φ

Viewing sphere

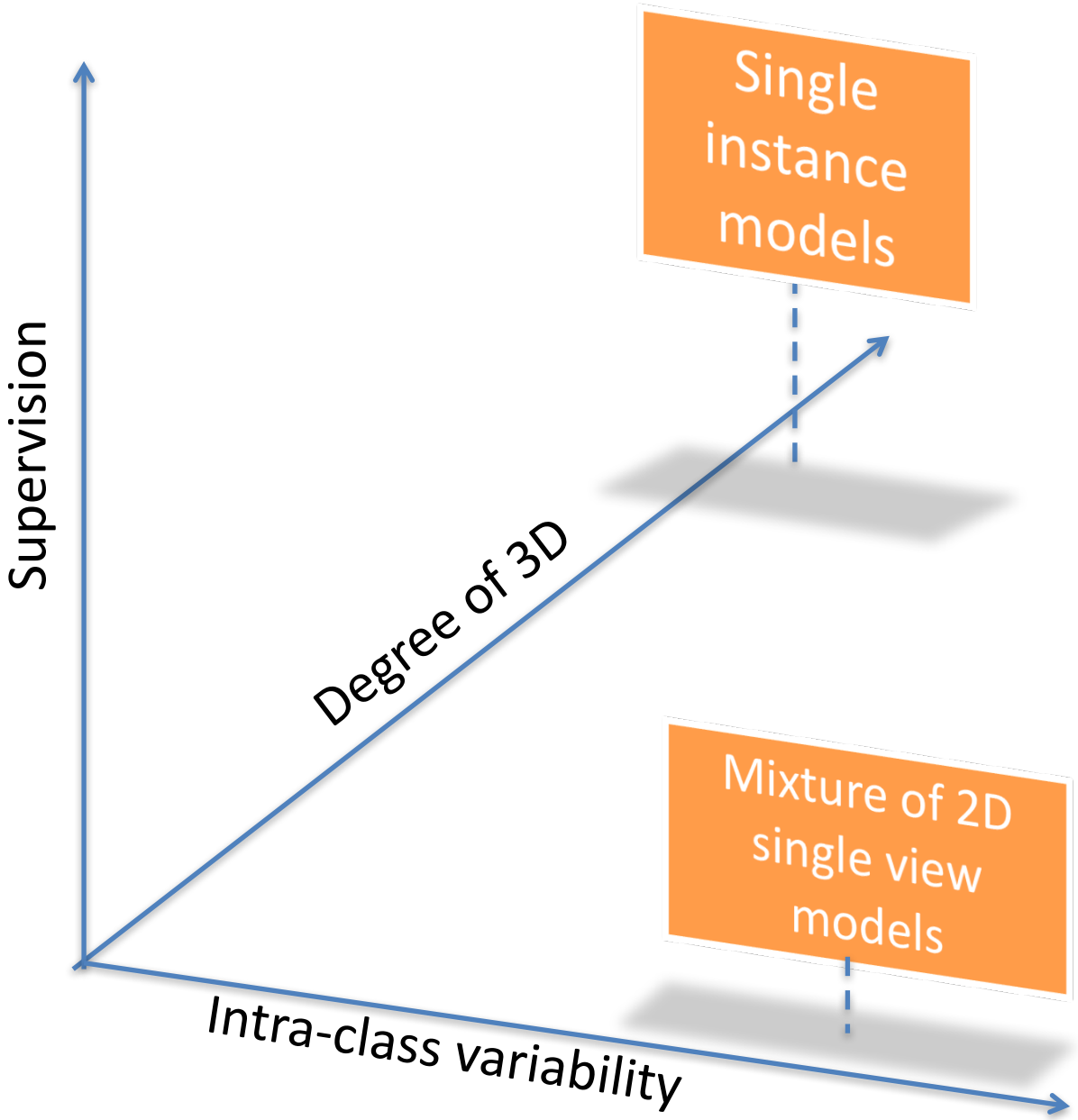


- Detect objects under generic view points
- Estimate object pose & 3D shape
- Work at different levels of specificity
- Limited amount of supervision

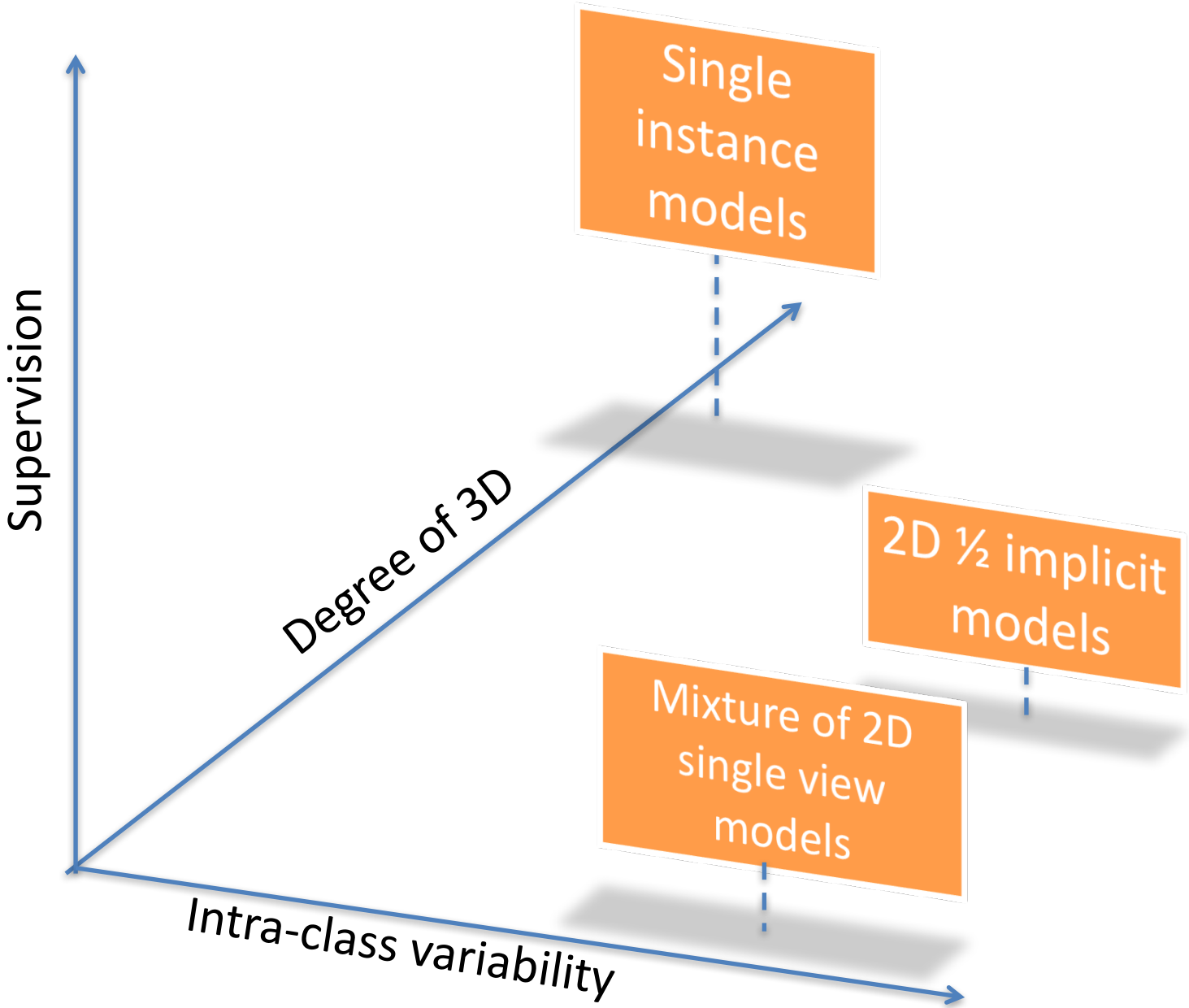
Models for 3d Object detection



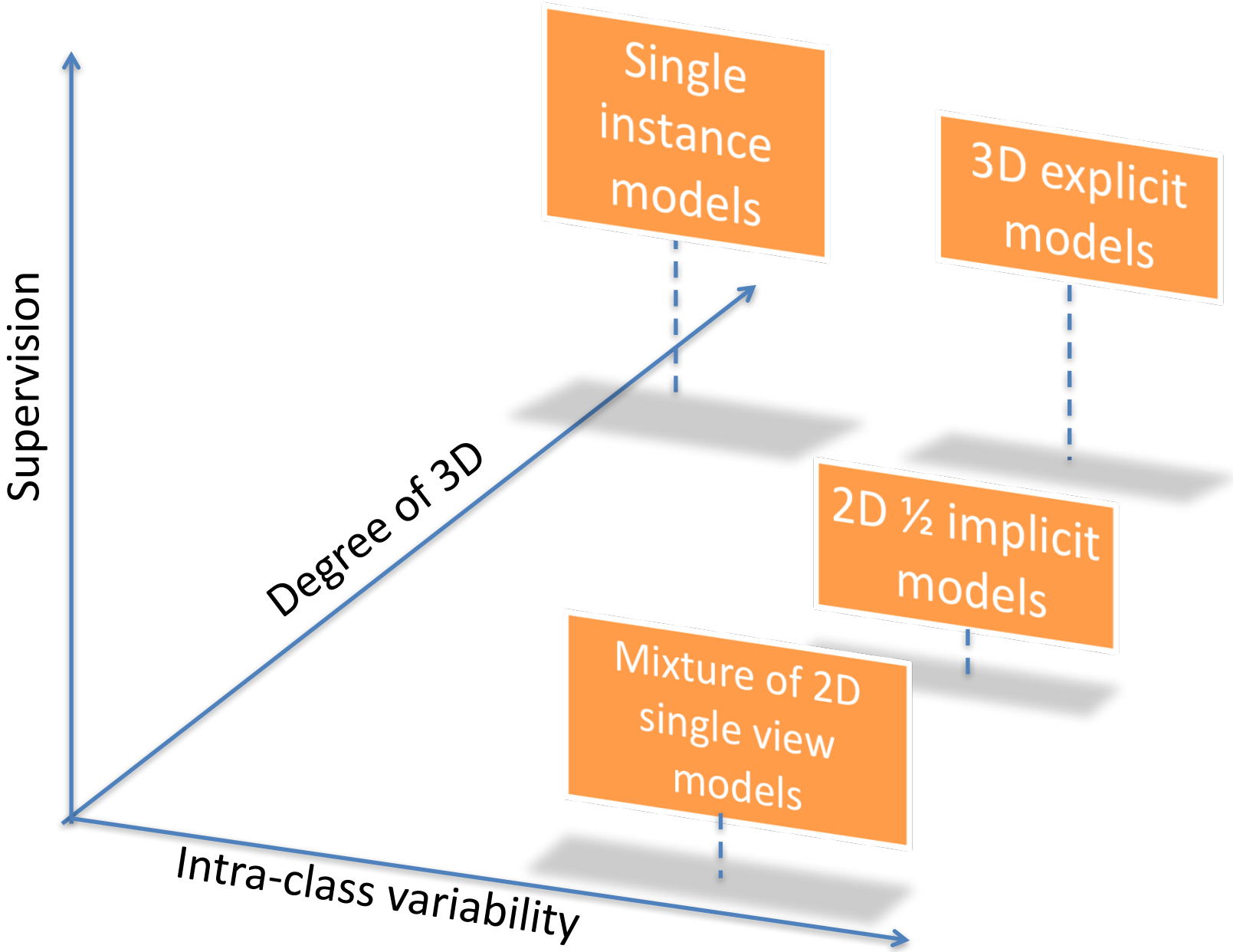
Models for 3d Object detection



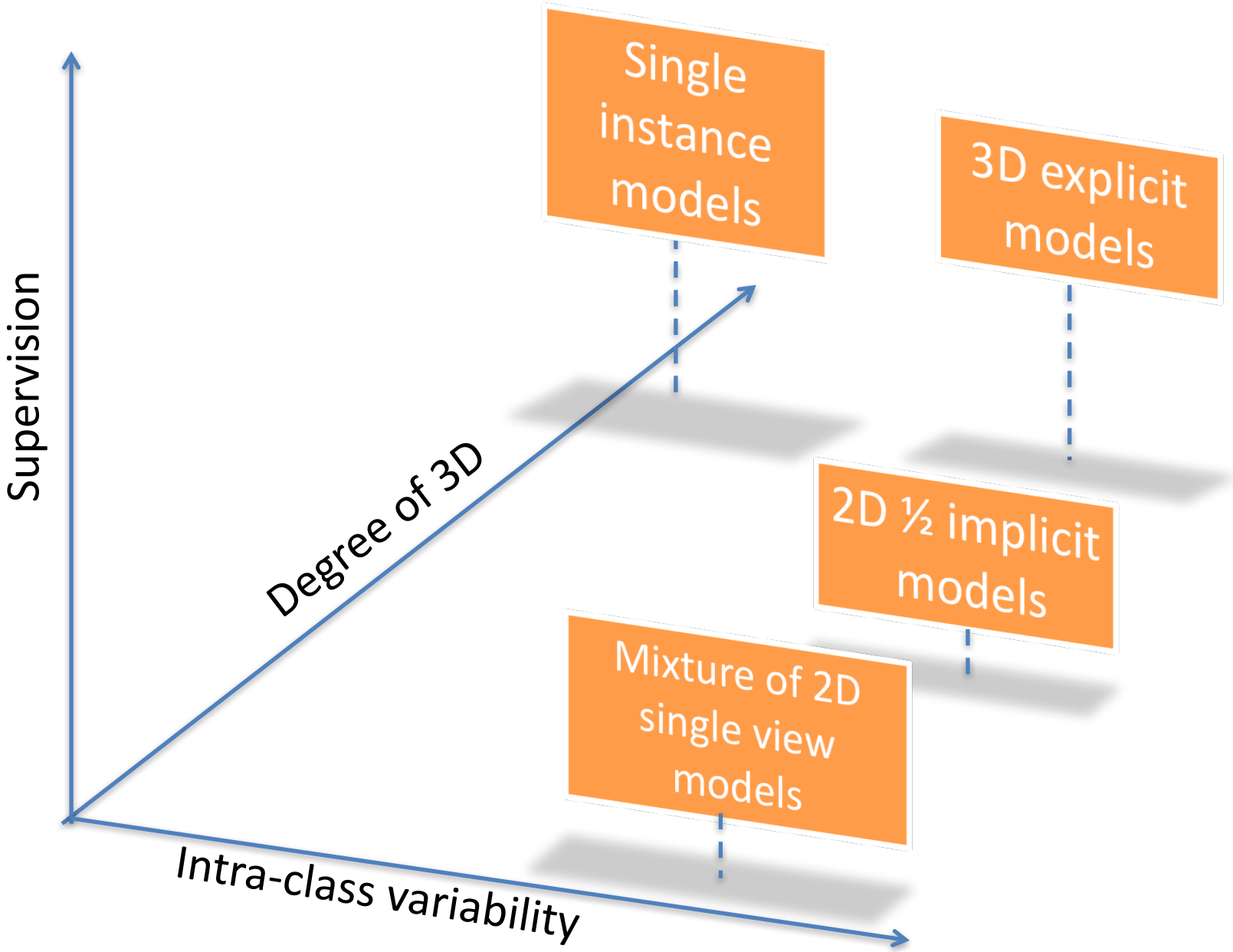
Models for 3d Object detection



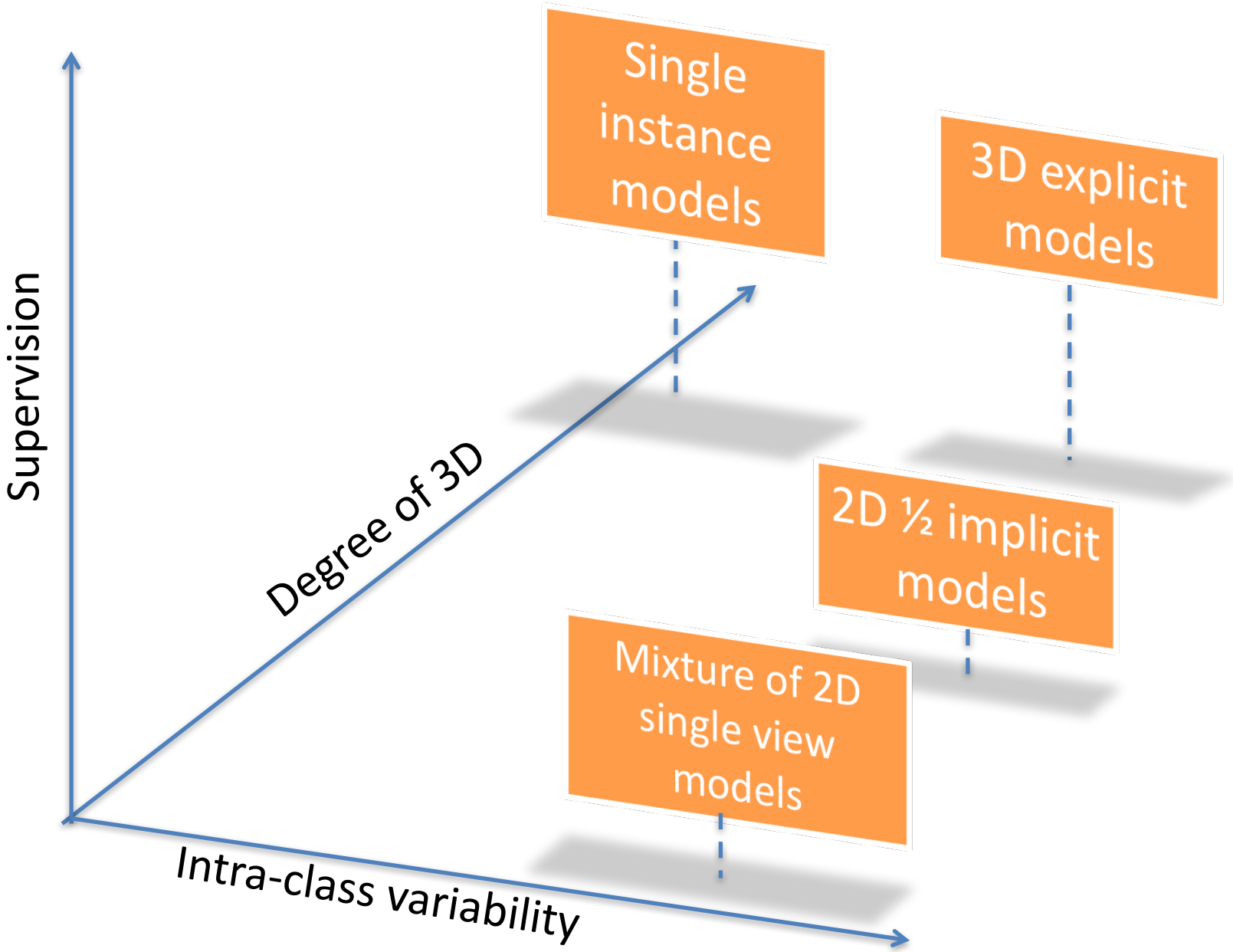
Models for 3d Object detection



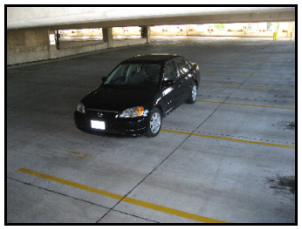
Models for 3d Object detection



Models for 3d Object detection



Single 3D object recognition
















Single 3D object recognition











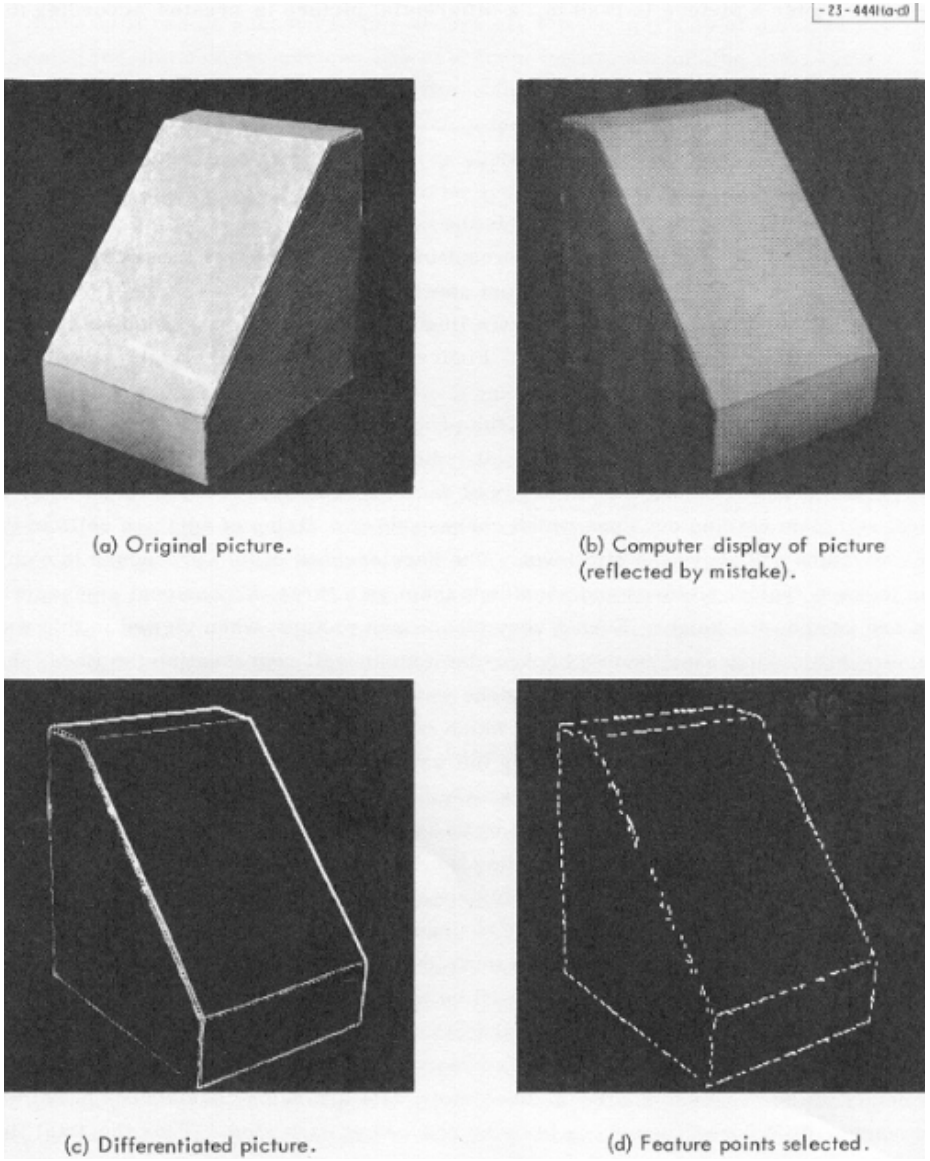






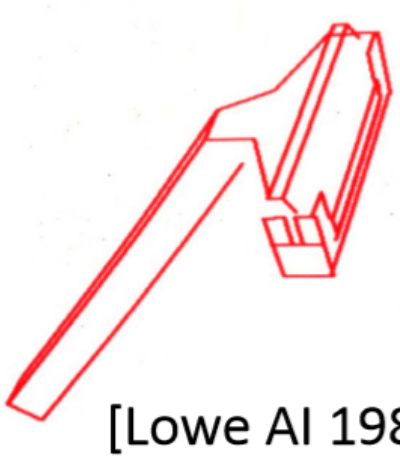


1963: Block world

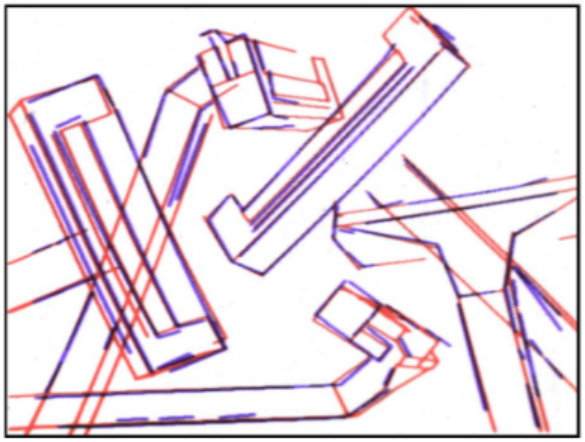


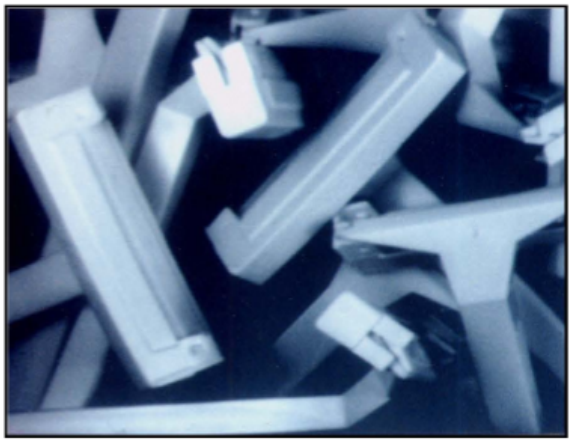
Larry Roberts

80s: First 3D object detectors



[Lowe AI 1987]





- Marr ’78, ‘82
- Ballard, ‘81
- Grimson & L.-Perez, ‘87

- Lowe, ’87
- Forsyth et al. ‘91
- Edelman et al. ’91

- Ullman & Barsi, ’91
- Rothwell ‘92
- Linderberg, ’94
- Murase & Nayar ‘94

Key Challenges

Variability due to:

- View point
- Illumination
- Occlusions
- Arbitrary texture

NOTE: intra-class variability doesn't need to be modeled

Modern 3D object recognition

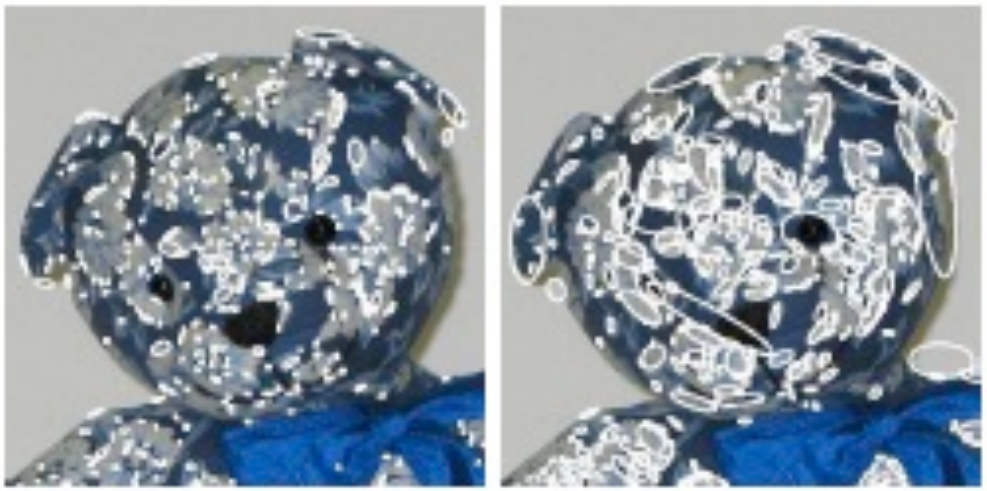
- Rothganger et al. ’04, ’06
- Brown et al, ’05
- Lowe ’99, ’04
- Ferrari et al. ’04, ’06
- Lazebnick et al ’04
- Hsiao et al., ’11-14
- Lim et al., ’13-16

Recognition paradigm:

Hypothesis generation & validation

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

Affine Harris-Laplace detector



Courtesy of Rothganger et al.

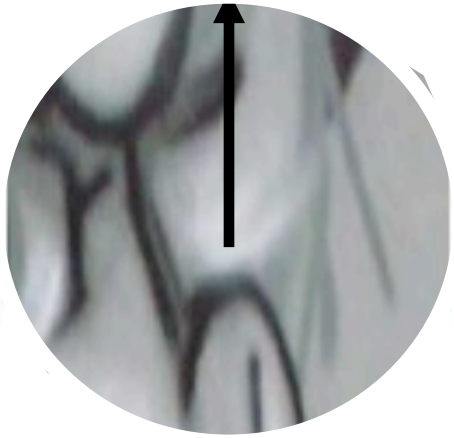
- x,y
- Scale
- Orientation

Key idea: use scale and orientation to normalize descriptors

View invariant descriptors

View 1

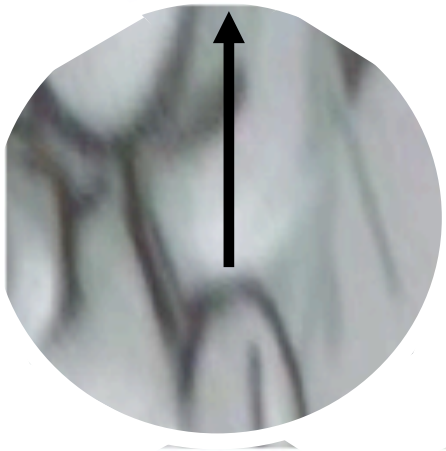
Rectification



SIFT

View 2

Rectification



SIFT

Basic scheme

-Representation

-Features

-2D/3D Geometrical constraints

-Model learning

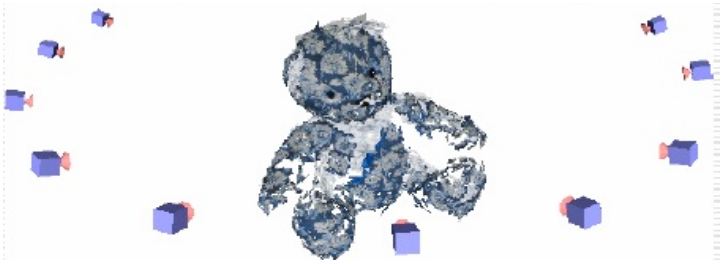
-Recognition

-hypothesis generation

-validation

Model learning

Build a 3D model:

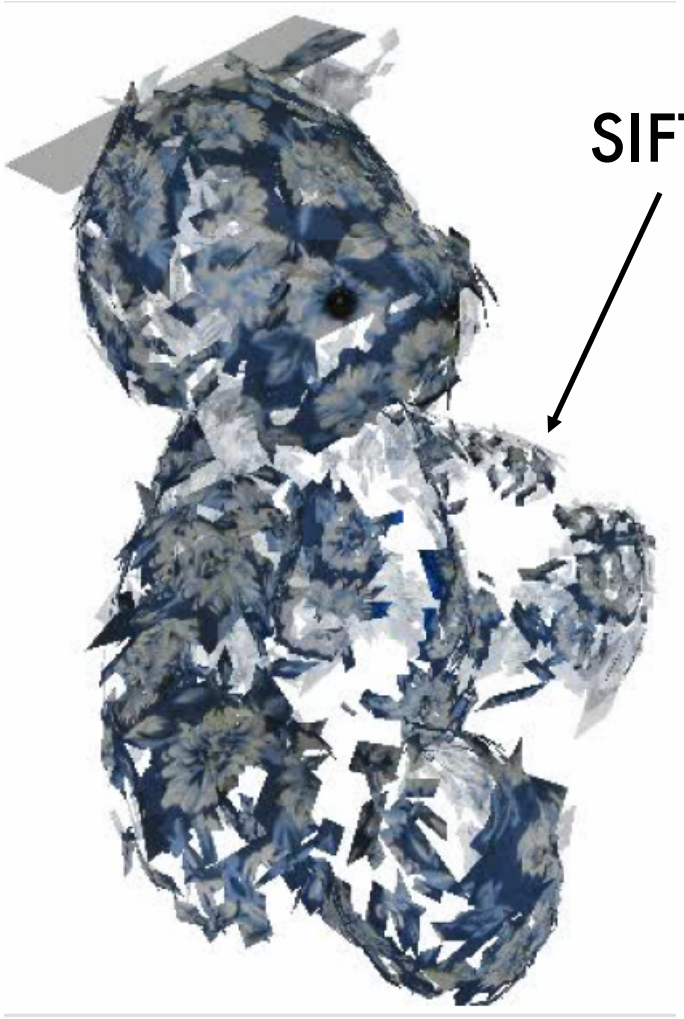


- N images of object from N different views

- Extract key points from each view
- Match key points between 2 views
- Use affine structure from motion to compute:
  - Keypoints 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

Learnt models



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

Courtesy of Rothganger et al

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, detect object instance and estimate its pose

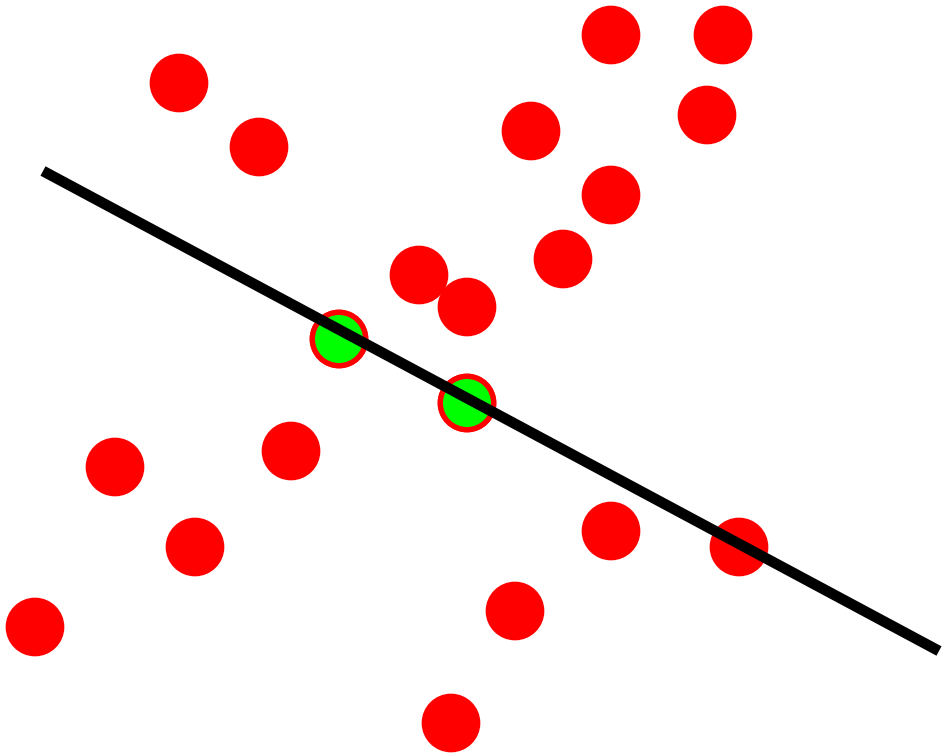
Equivalent to: from a collection of learnt object models, find object model that fits object in image

Equivalent to a fitting problem!

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

RANSAC!

RANSAC!



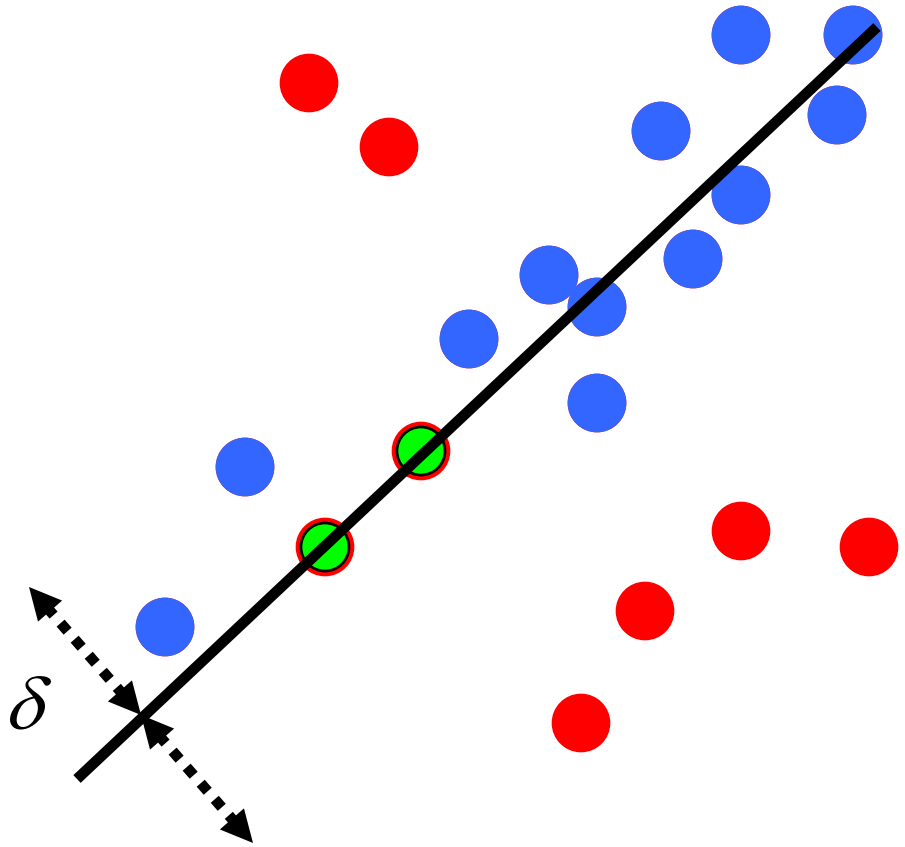
Sample set = set of points in 2D

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!



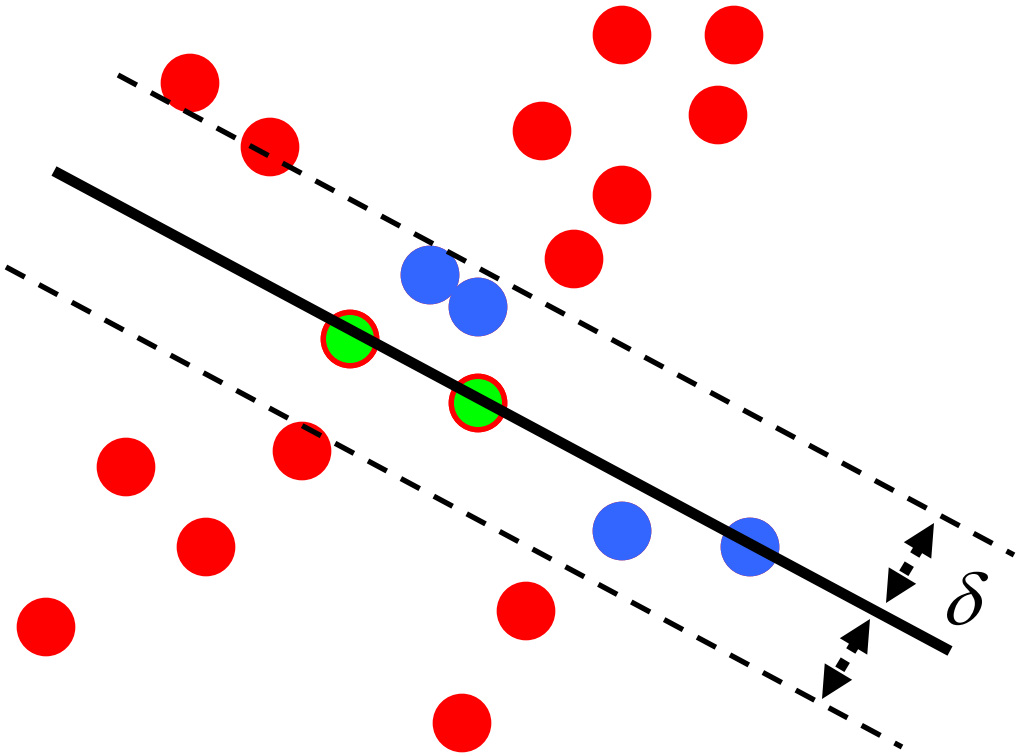
$|O| = 6$

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

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

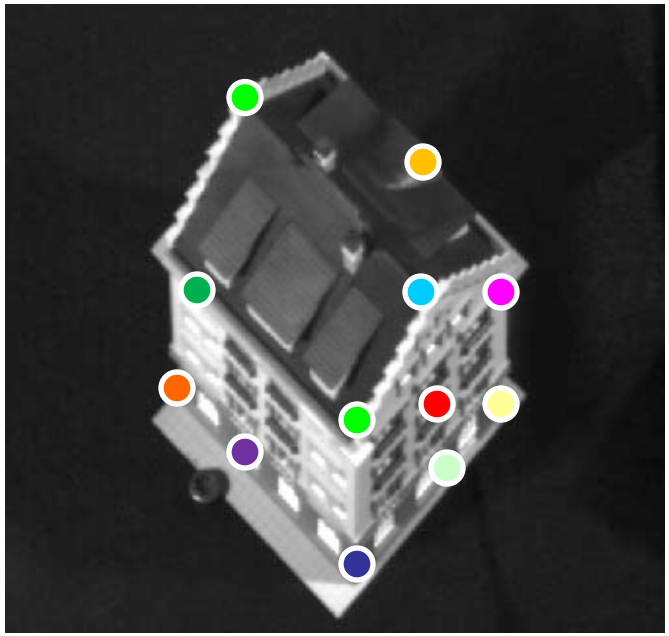
Repeat 1-3 until model with the most inliers over all samples is found

2D model detection

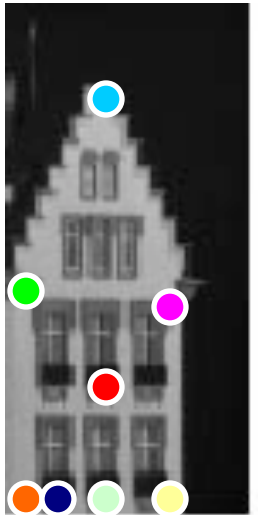
Goal: given a query image I, find object model that matches with I

Model: collection of points on planar surface

query



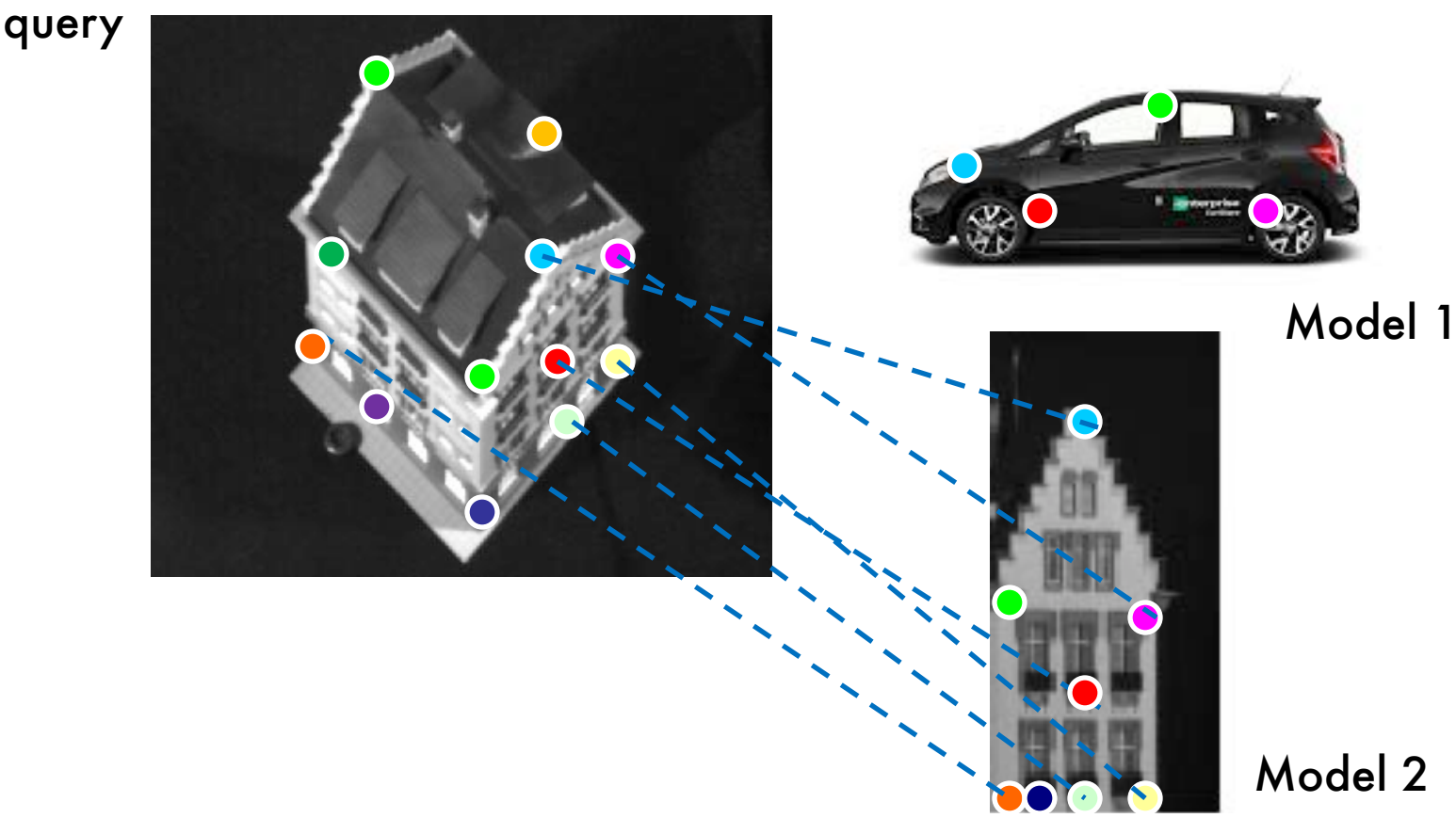
Model 1



Model 2

2D model detection

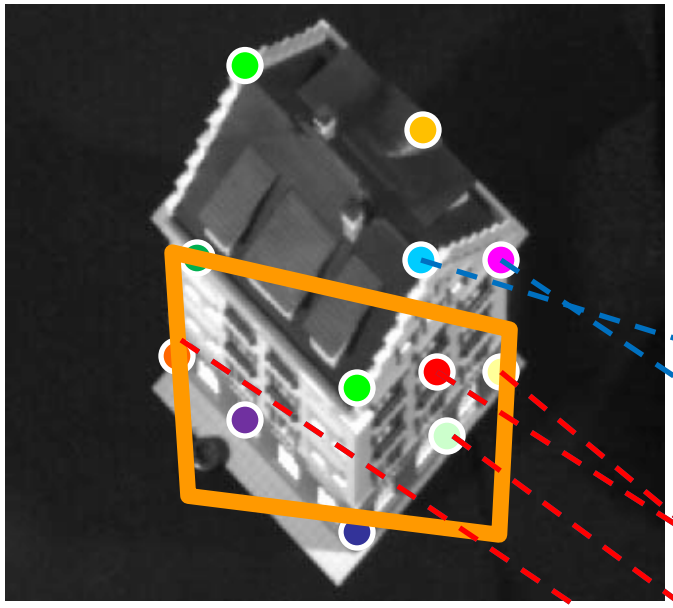
- Find matches between “model” points and “query” points
- Using N matches to fit homographic transformation
- If matches and selected model are correct, the fitting error is small



2D model detection

- Find matches between "model" points and "query" points
- Using N matches to fit homographic transformation
- If matches and selected model are correct, the fitting error is small

query



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

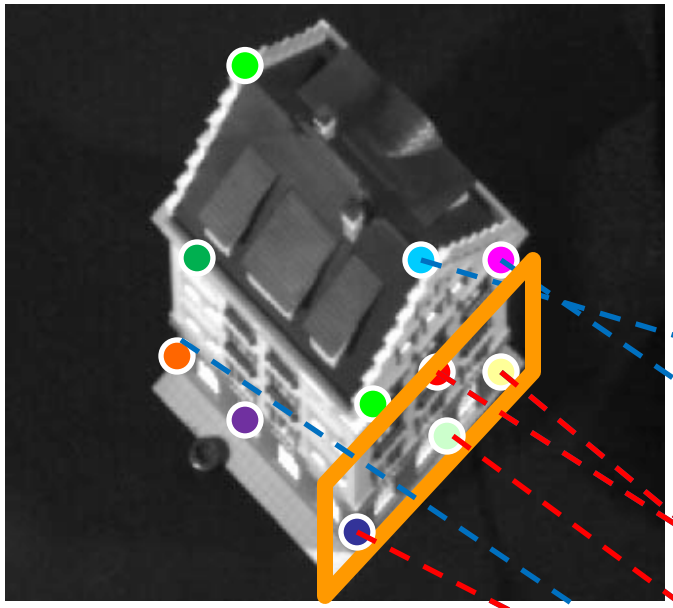
Model 2

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

2D model detection

- Find matches between "model" points and "query" points
- Using N matches to fit homographic transformation
- If matches and selected model are correct, the fitting error is small

query



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

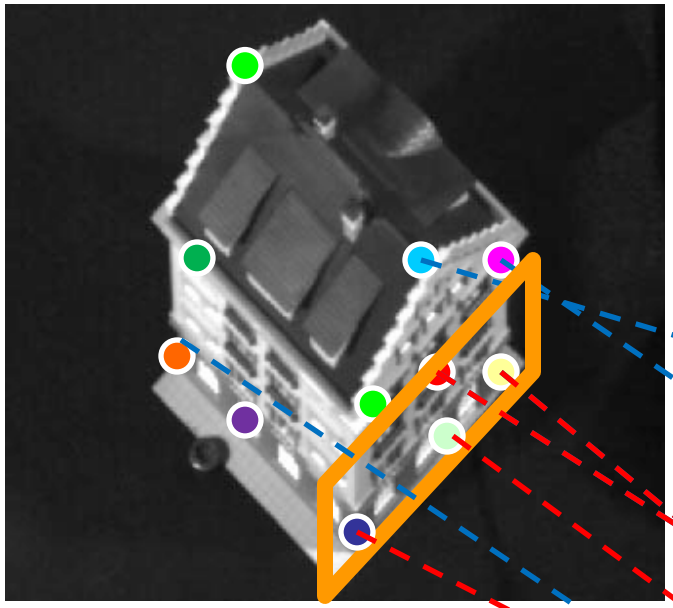
Model 2

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

2D model detection

- Find matches between "model" points and "query" points
- Using N matches to fit homographic transformation
- If matches and selected model are correct, the fitting error is small

query



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

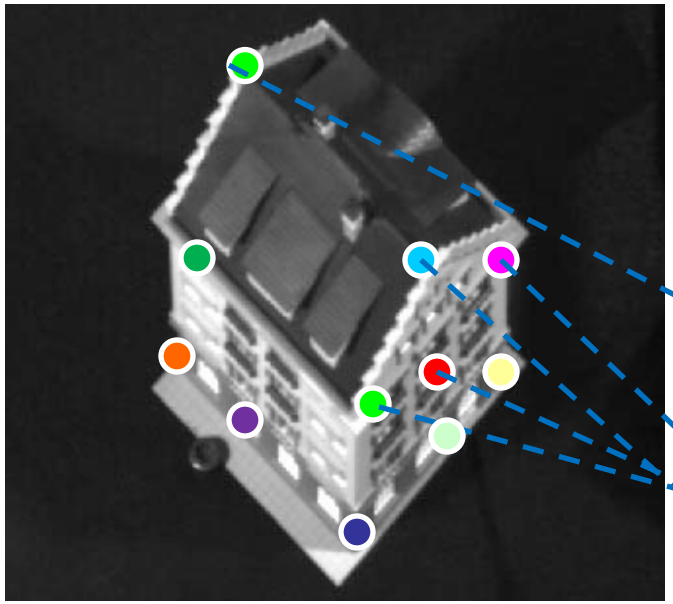
Model 2

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

2D model detection

- Find matches between “model” points and “query” points
- Using N matches to fit homographic transformation
- If matches and selected model are correct, the fitting error is small

query



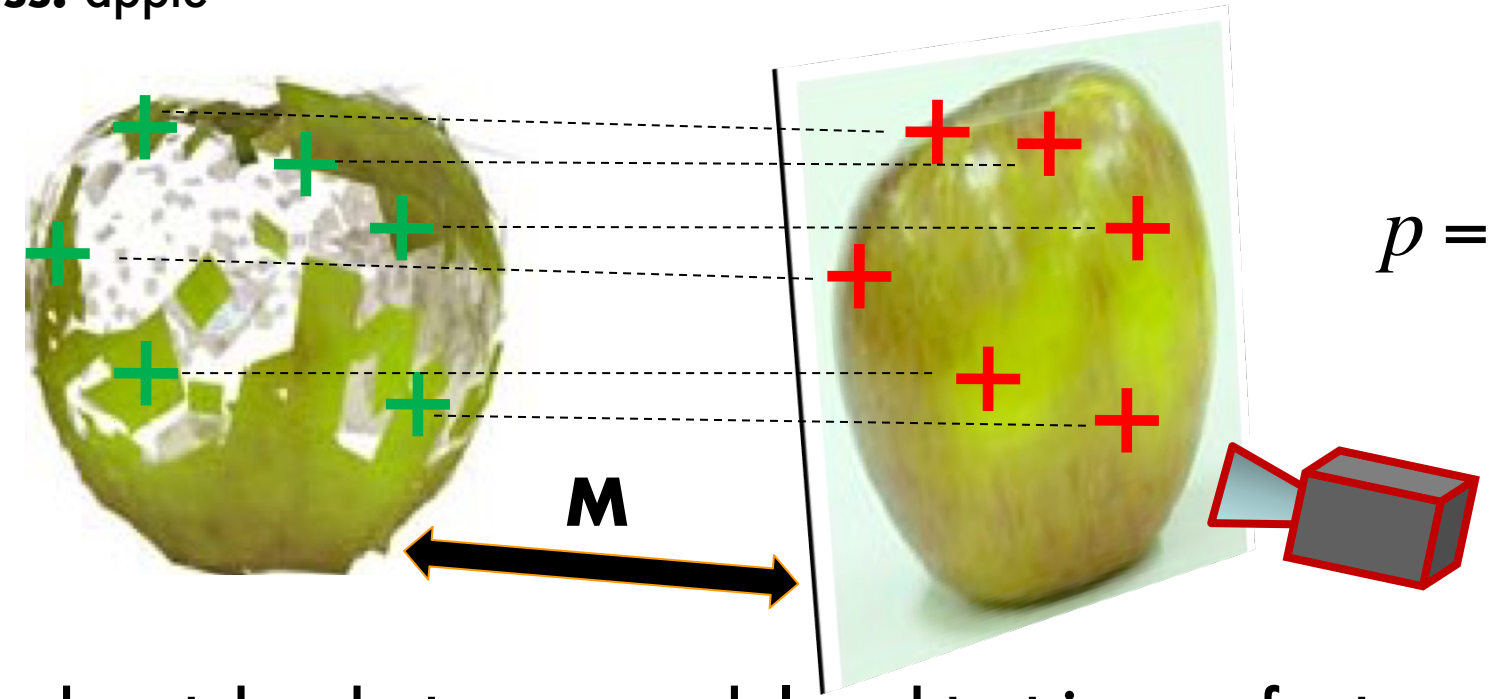
- Generate hypothesis
- Verify hypothesis
- Select hypothesis with lowest fitting error
- Generate recognition results

Verification: if the model is wrong, each hypothesis generates **high** fitting errors!

Model 1

Recognition

Class: apple

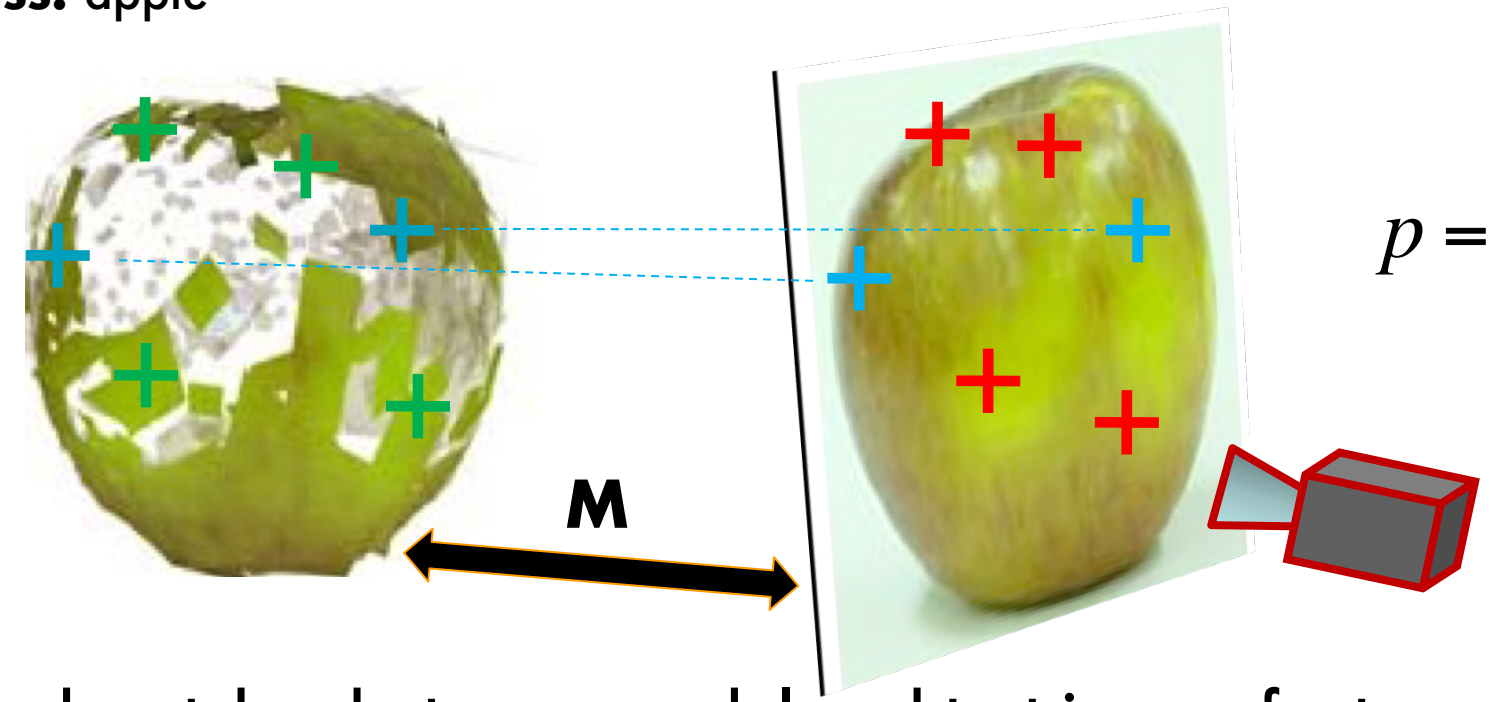


$$p = M\ P$$

1. Find matches between model and test image features
2. Generate hypothesis:
   - Compute transformation p = **M** P, from N matches

Recognition

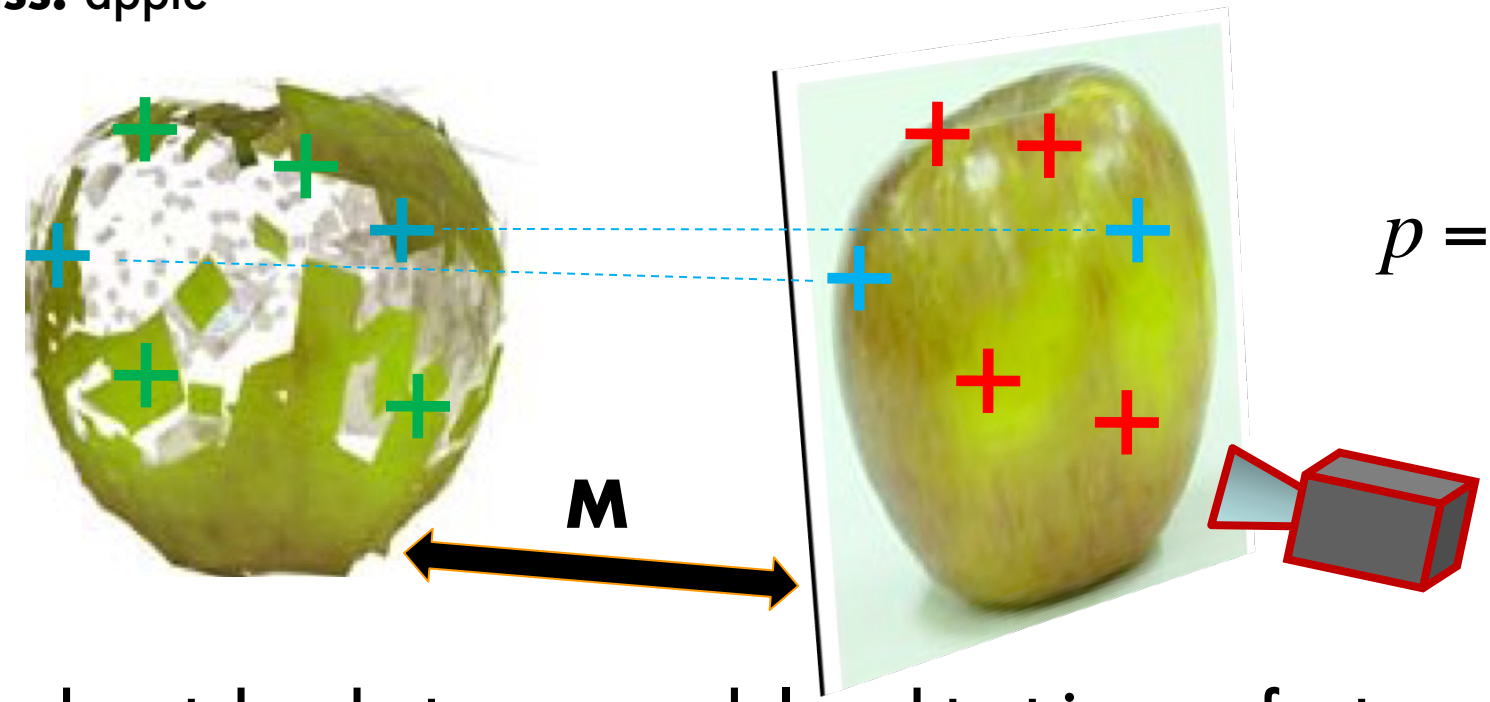
Class: apple



1. Find matches between model and test image features
2. Generate hypothesis:
   - Compute transformation p = **M** P, from N matches (N=2, if affine camera & affine key points

Recognition

Class: apple



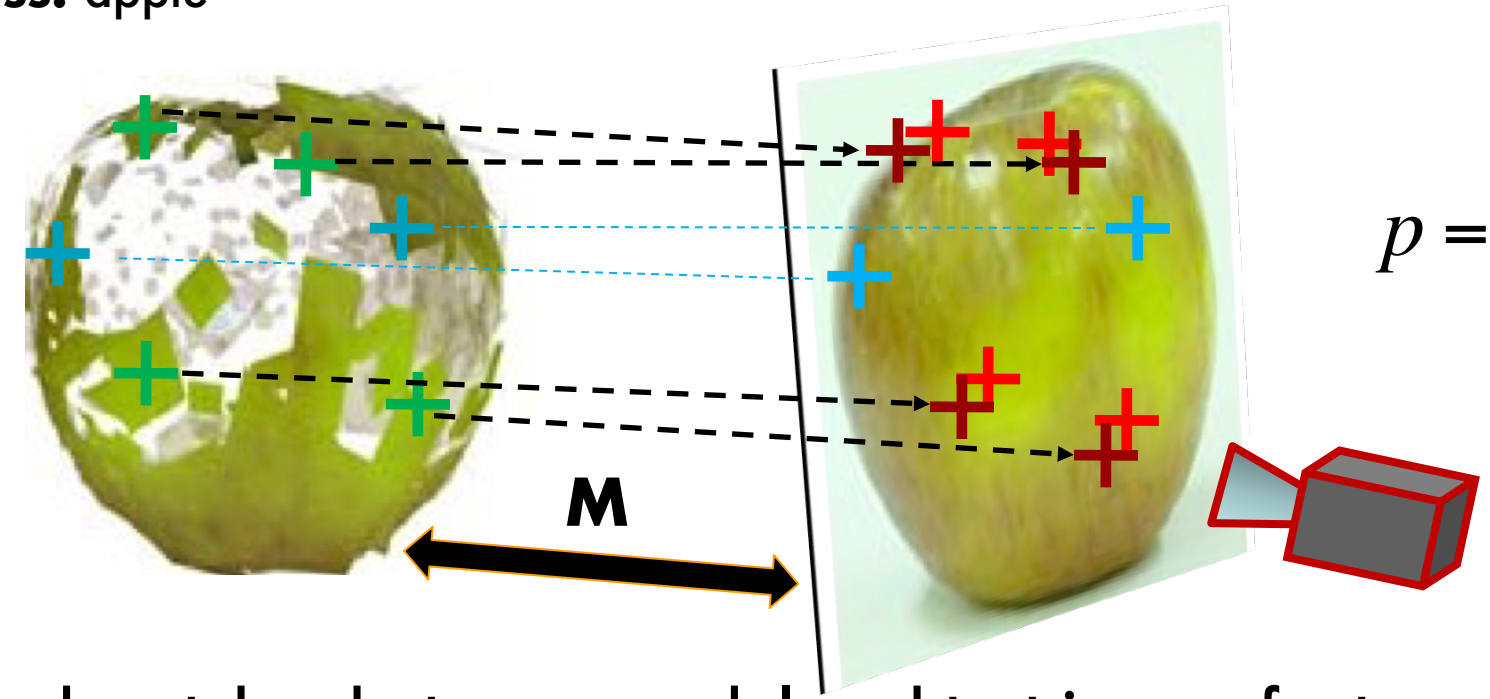
$$p = M\ P$$

1. Find matches between model and test image features
2. Generate hypothesis:
   - Compute transformation p = **M** P, from N matches (N=2, if affine camera & affine key points)

   → Generate hypothesis of object location and pose w.r.t. camera

Recognition

Class: apple



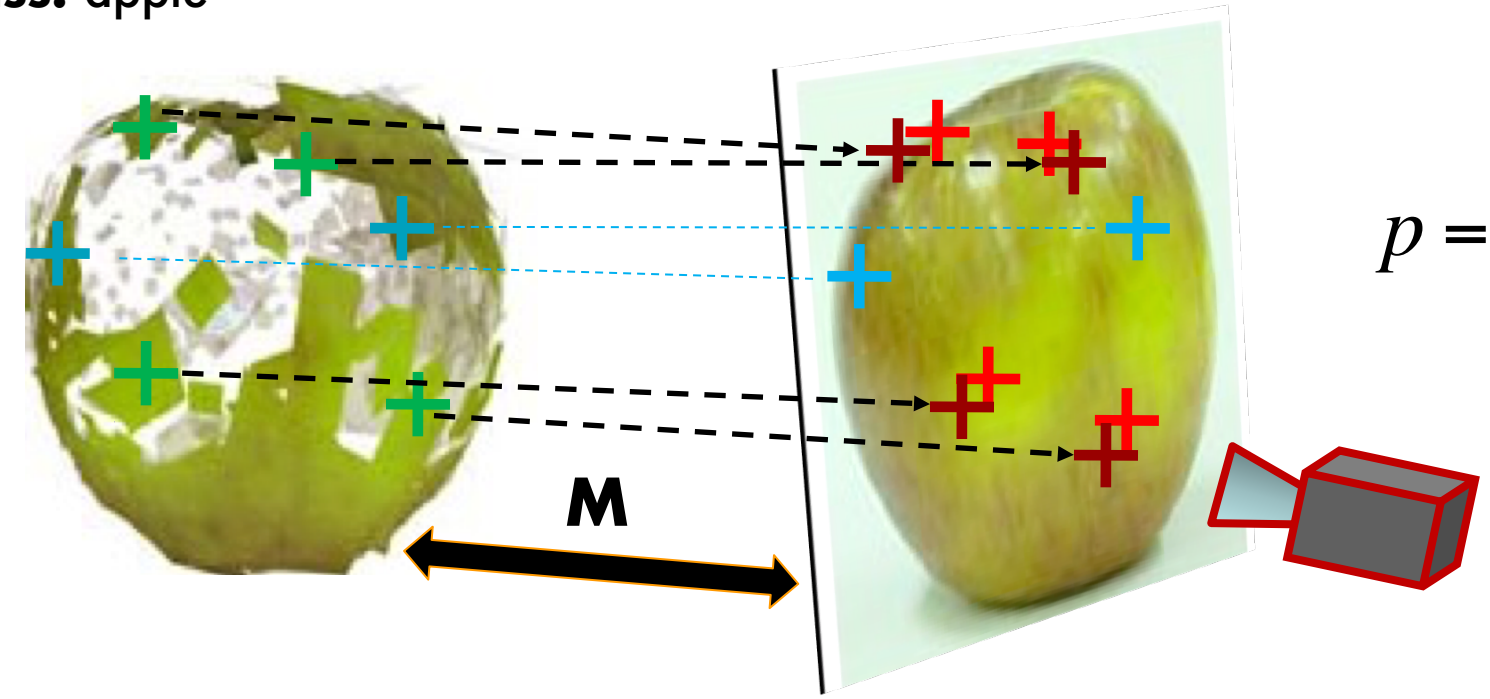
$$p = M\ P$$

1. Find matches between model and test image features
2. Generate hypothesis:
   - Compute transformation p = **M** P, from N matches (N=2, if affine camera & affine key points)

   → Generate hypothesis of object location and pose w.r.t. camera
3. Model verification
   - Use **M** to project other 3D model features into test image
   - Compute residual = D(projections, measurements)

Recognition

Class: apple



$$p = M\ P$$

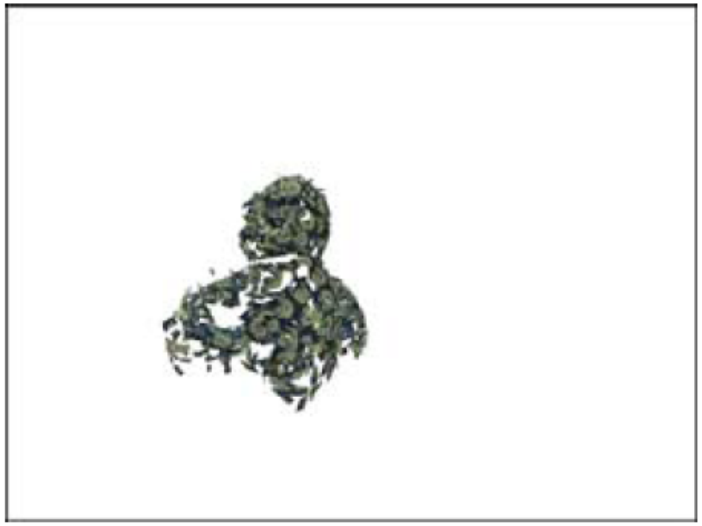
4. Repeat steps 2 and 3 until residual doesn't decrease anymore
5. Repeat steps 1-4 for different object instance C (apple, teddy bear, etc…)
6. M and C corresponding to min residual return the estimated object pose and object instance

Object to recognize

Initial matches based on appearance

Matches verified with geometrical constraints

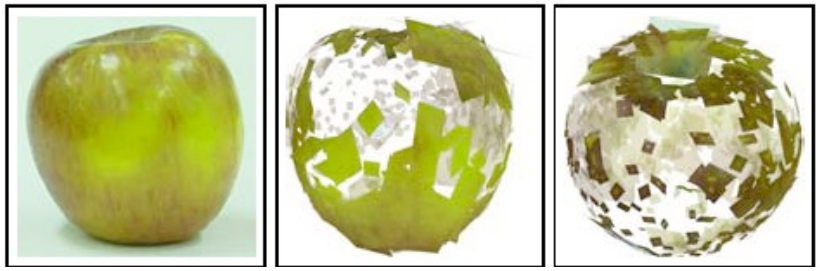
Recovered pose



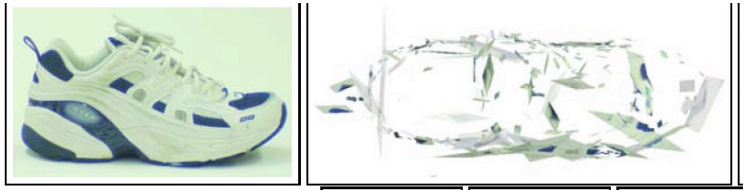
Courtesy of Rothganger et al

Detection and pose estimations results

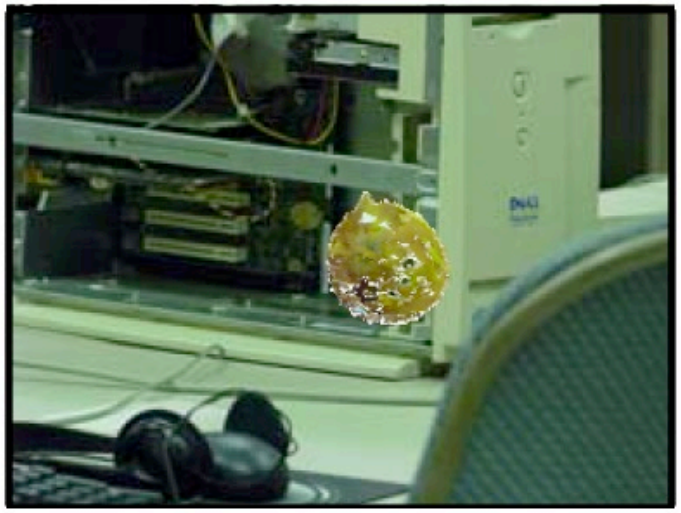
"apple" model



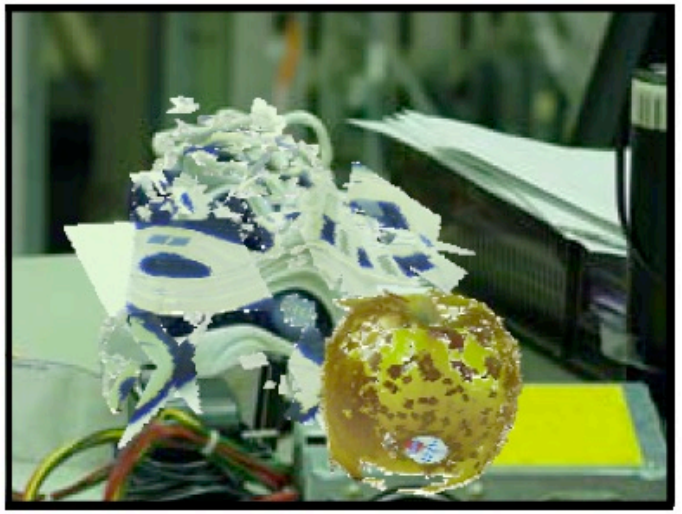
"shoe" model



Test image



Test image



Courtesy of Rothganger et al

- Handle severe clutter

3D object detectors

Lowe. '99, '04

Courtesy of D. Lowe

- Handle severe occlusions
- Fast!

Detecting food in your kitchen!

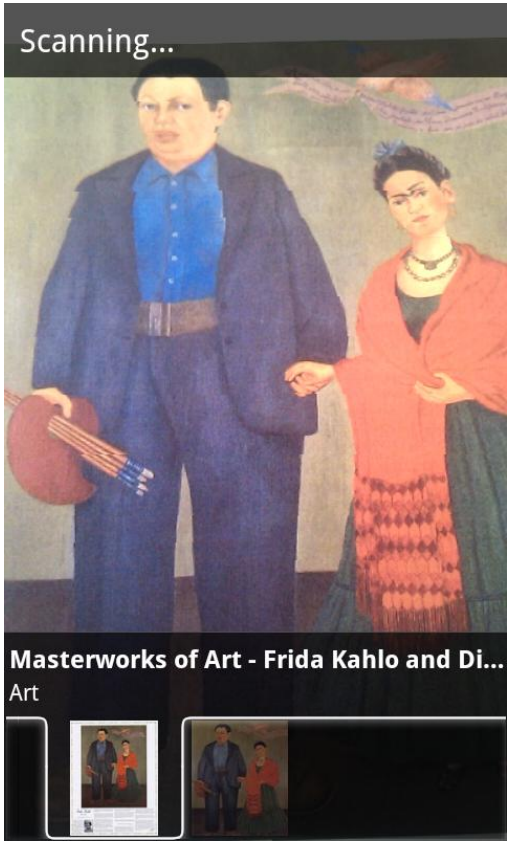
Edward Hsiao, Alvaro Collet and Martial Hebert. **Making specific features less discriminative to improve point-based 3D object recognition**. *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, June, 2010.

Hsiao, Alvaro Collet and Martial Hebert, **Occlusion Reasoning for Object Detectionunder Arbitrary Viewpoint**, PAMI 2014

Detecting IKEA furniture!

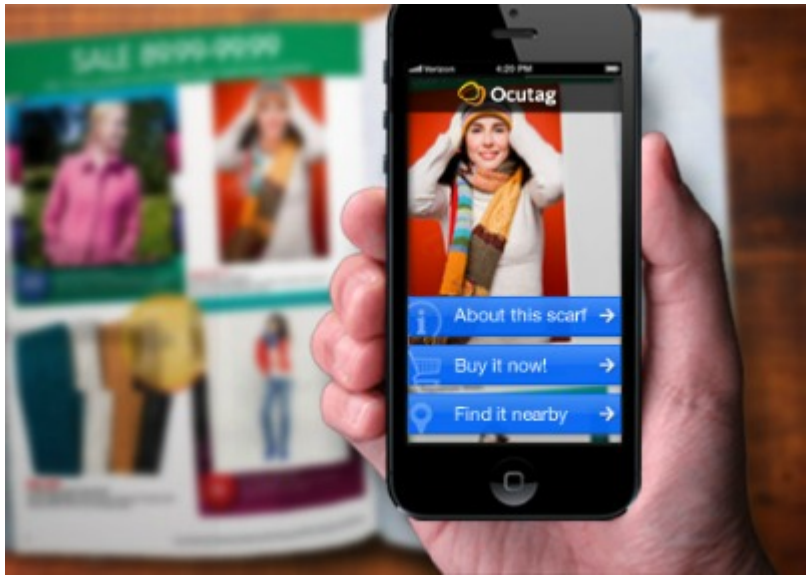
Parsing IKEA Objects: Fine Pose Estimation. Joseph Lim, Hamed Pirsiavash, and Antonio Torralba. International Conference on Computer Vision (ICCV), 2013.

Visual search and landmarks recognition



Google Goggles

Visual search and landmarks recognition



RICOH

A9

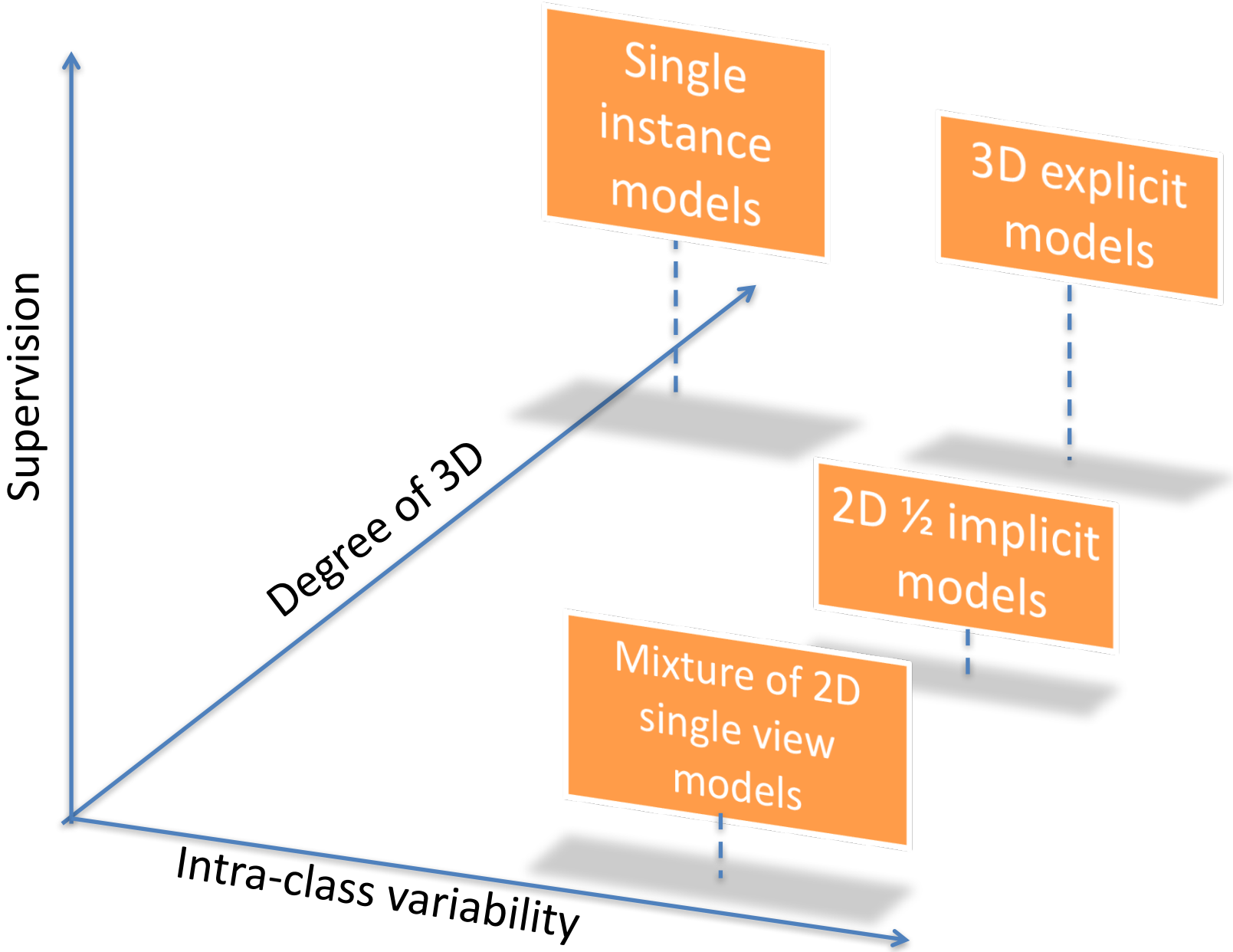
Limitations of single instance 3D object detectors

- Cannot handle intra-class variability.

Why?

- Models capture fine-grained details of the object instance which are not shared across instances in the same class
- Hypothesis-generation and verification scheme is not designed to maximize discrimination power

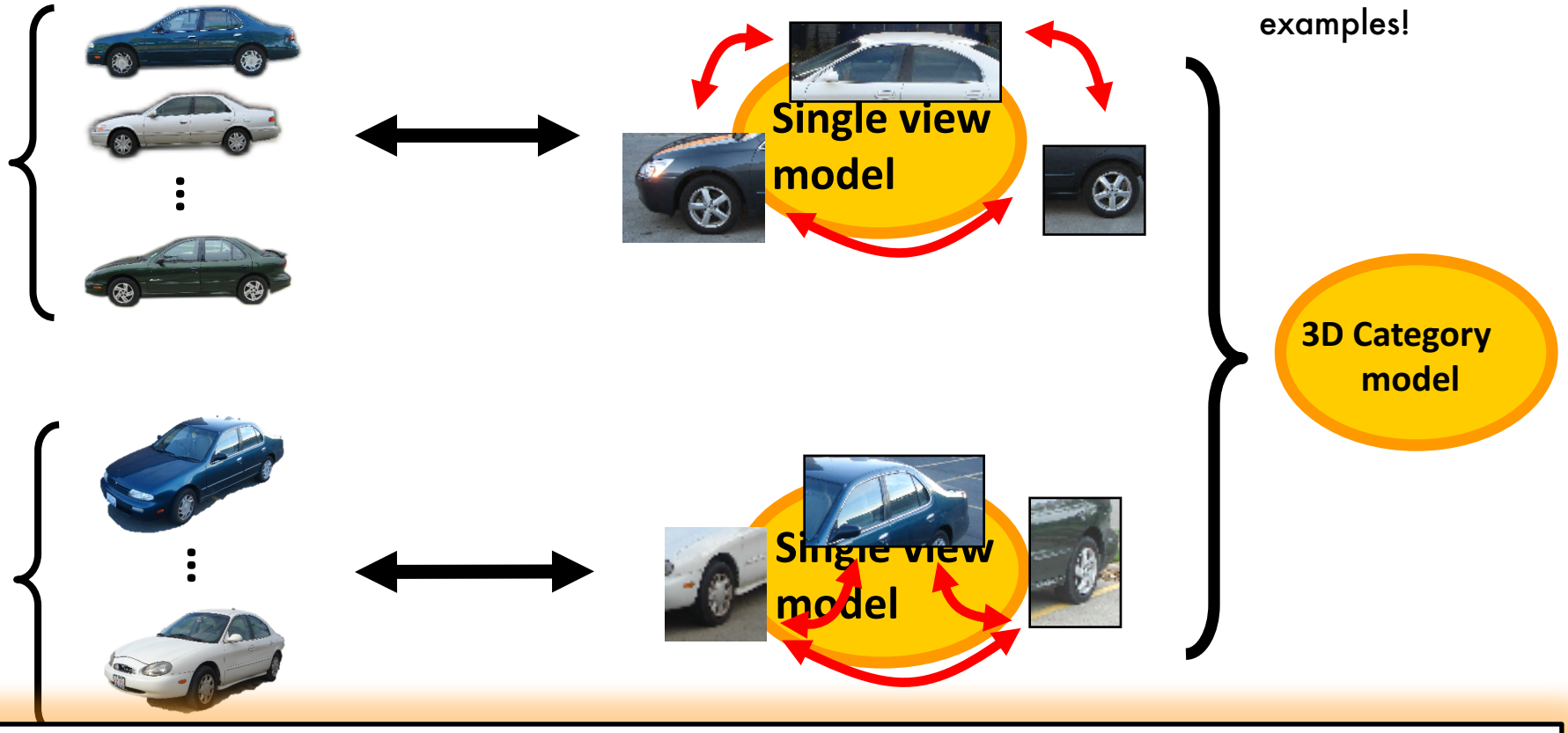
Models for 3d Object detection



Mixture of 2D models

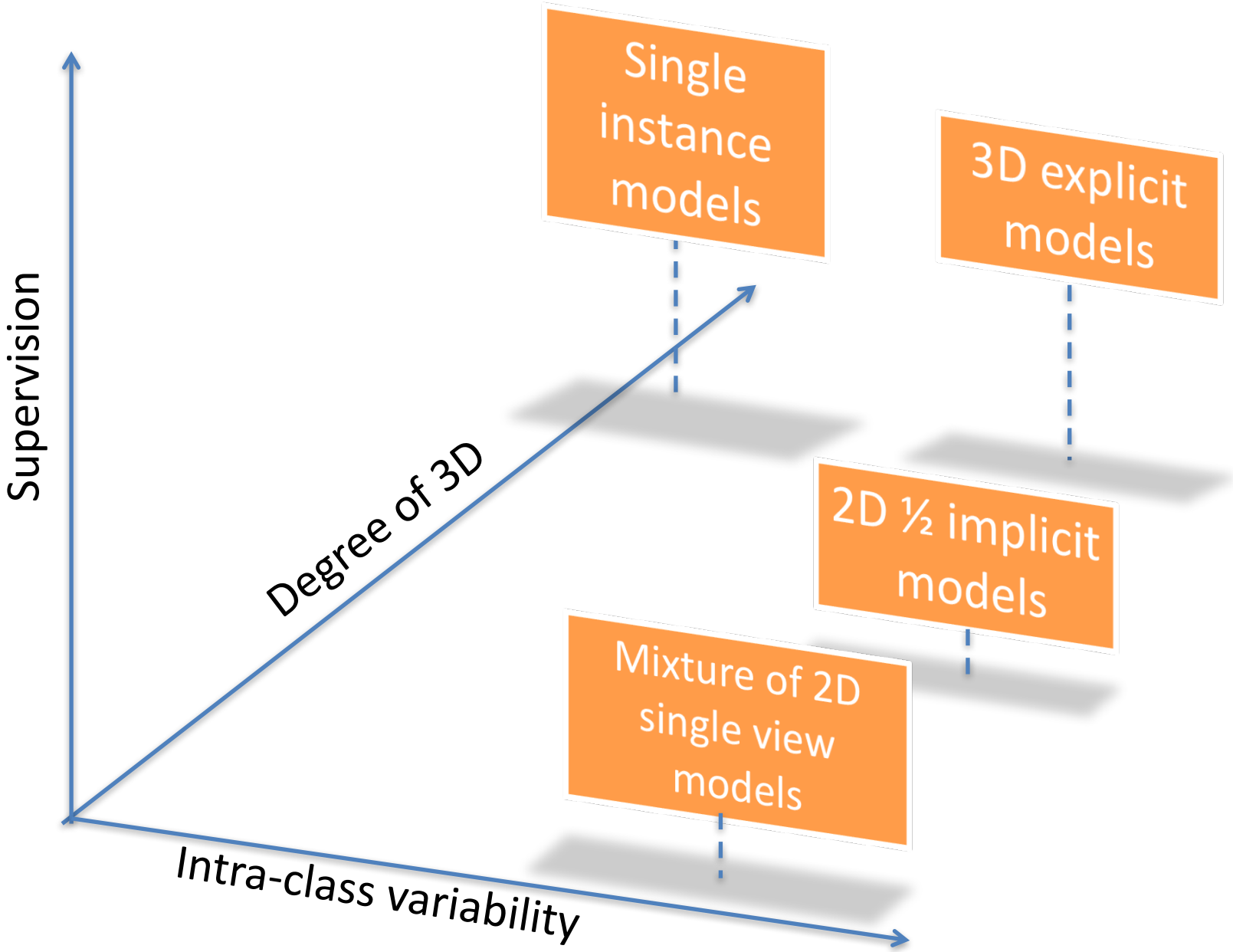
- Weber et al. '00
- Schneiderman et al. '01
- Ullman et al. 02
- Fergus et al. '03
- Torralba et al. '03
- Felzenszwalb & Huttenlocher '03
- Leibe et al. '04
- Shotton et al. '05
- Grauman et al. '05
- Savarese et al, '06
- Todorovic et al. '06
- Vedaldi & Soatto '08
- Zhu et al 08
- Gu & Ren, '10

DPM and ISM are examples!



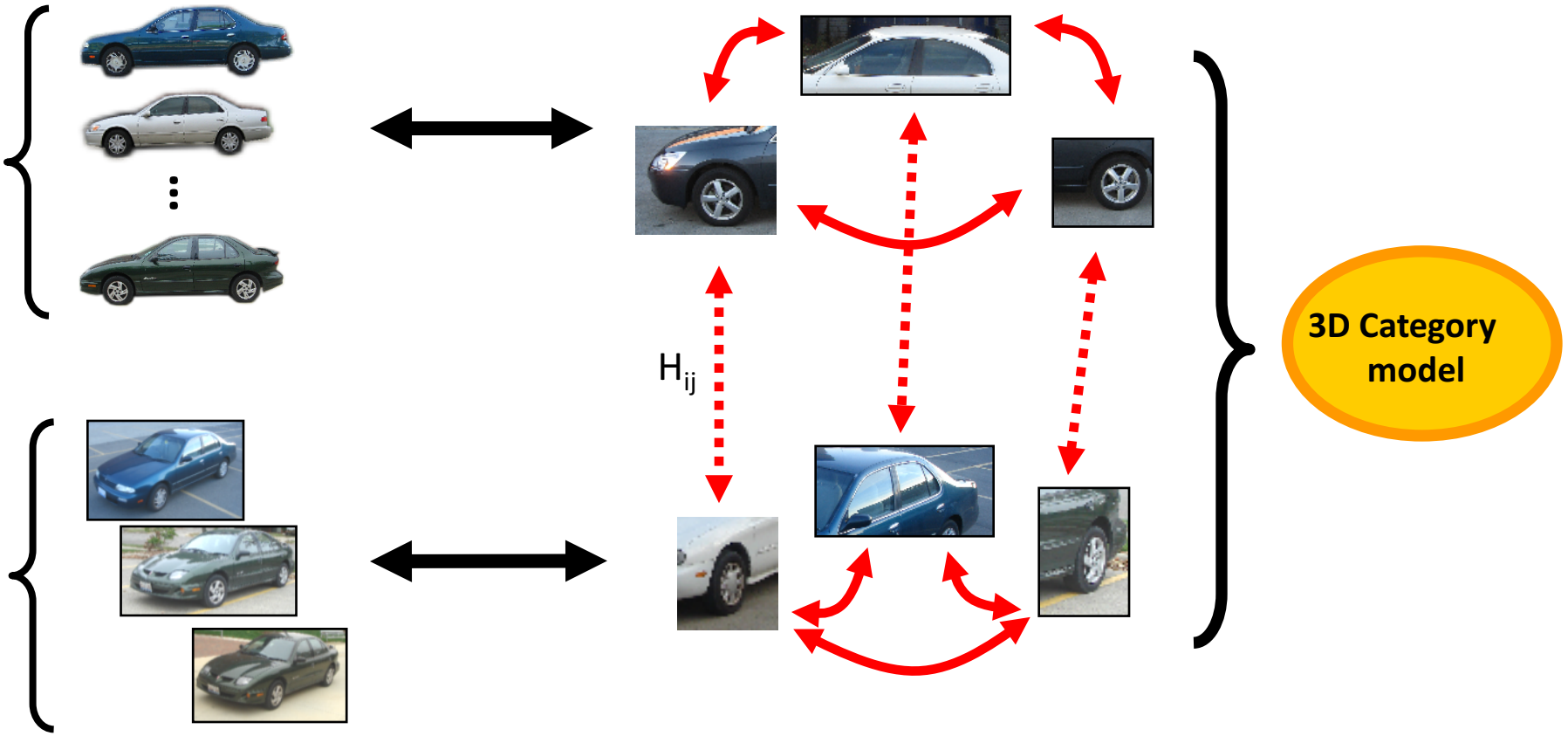
CONS: Single view models are independent • Non scalable to large number of categories/view-points • Just b. boxes • Cannot estimate 3D pose or 3D layout

Models for 3d Object detection



2D ½ implicit models

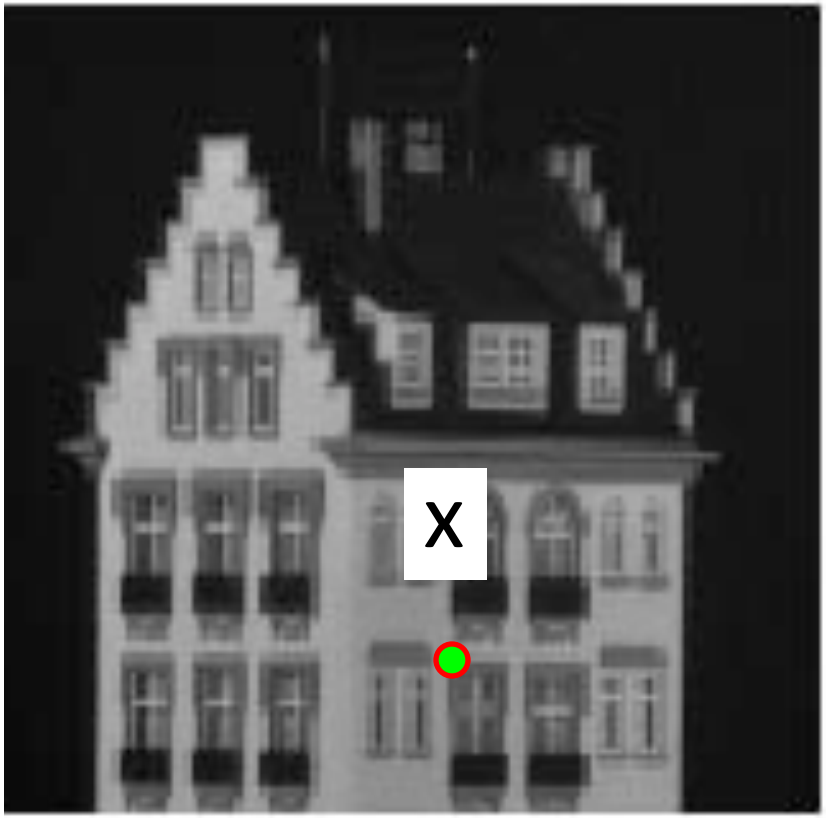
- Savarese & Fei-Fei, ICCV 07
- Savarese & Fei-Fei, ICCV 07
- Su, Sun, Fei-Fei, Savarese., CVPR 2009
- Sun, Su, Fei-Fei, Savarese, ICCV 2009
- Thomas et al. '06-09
- Kushal, et al., '07
- Farhadi '09
- Zhu et al. '09
- Ozuysal et al. '10
- Stark et al.'10
- Payet & Todorovic, 11
- Glasner et al., '11

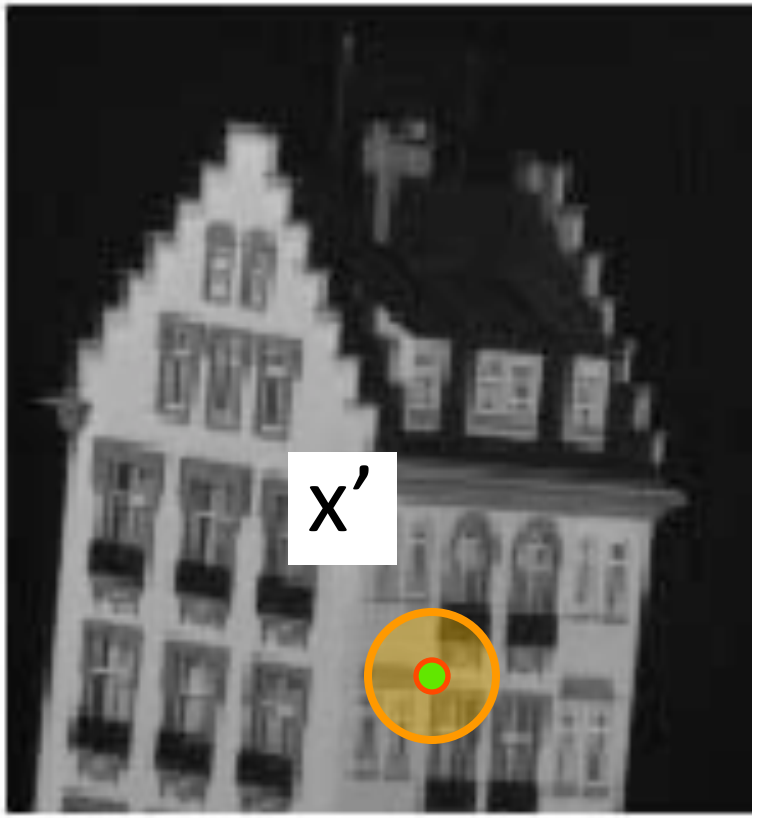


- Parts relationship can be probabilistic and learnt automatically

Linking features or parts across views:

Perspective or affine transformation constraints

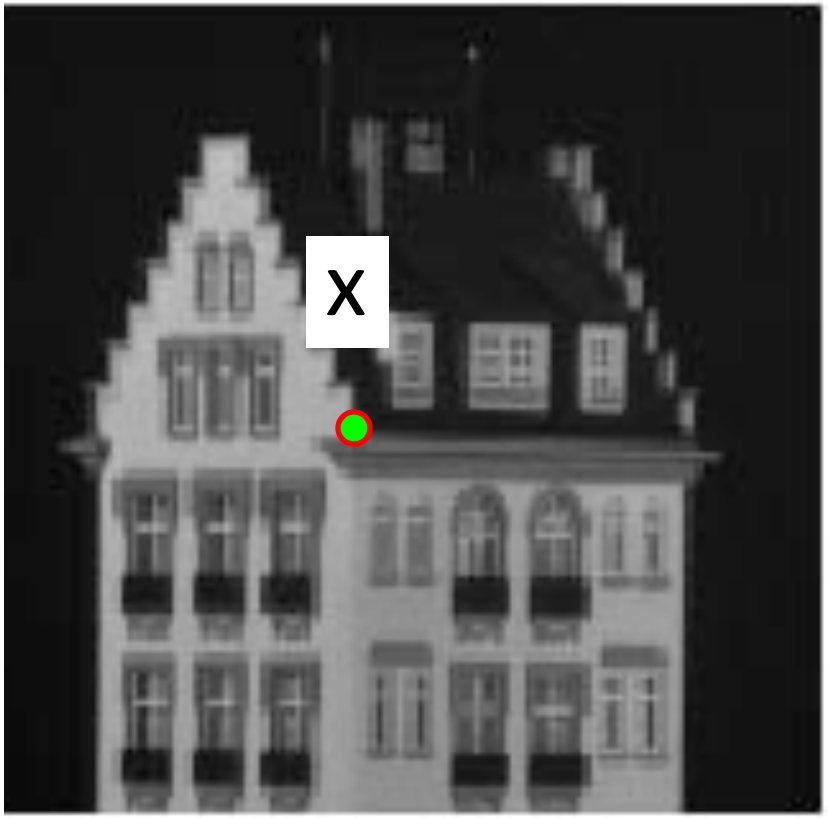


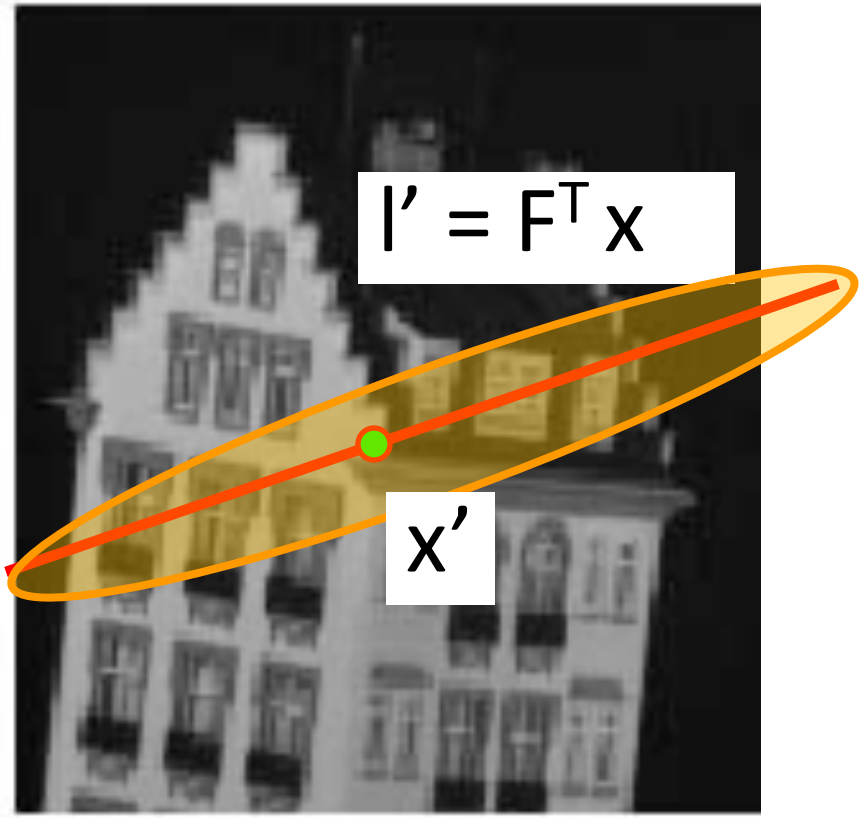


$$x' = H\,x$$

Linking features or parts across views:

Epipolar Transformation Constraints



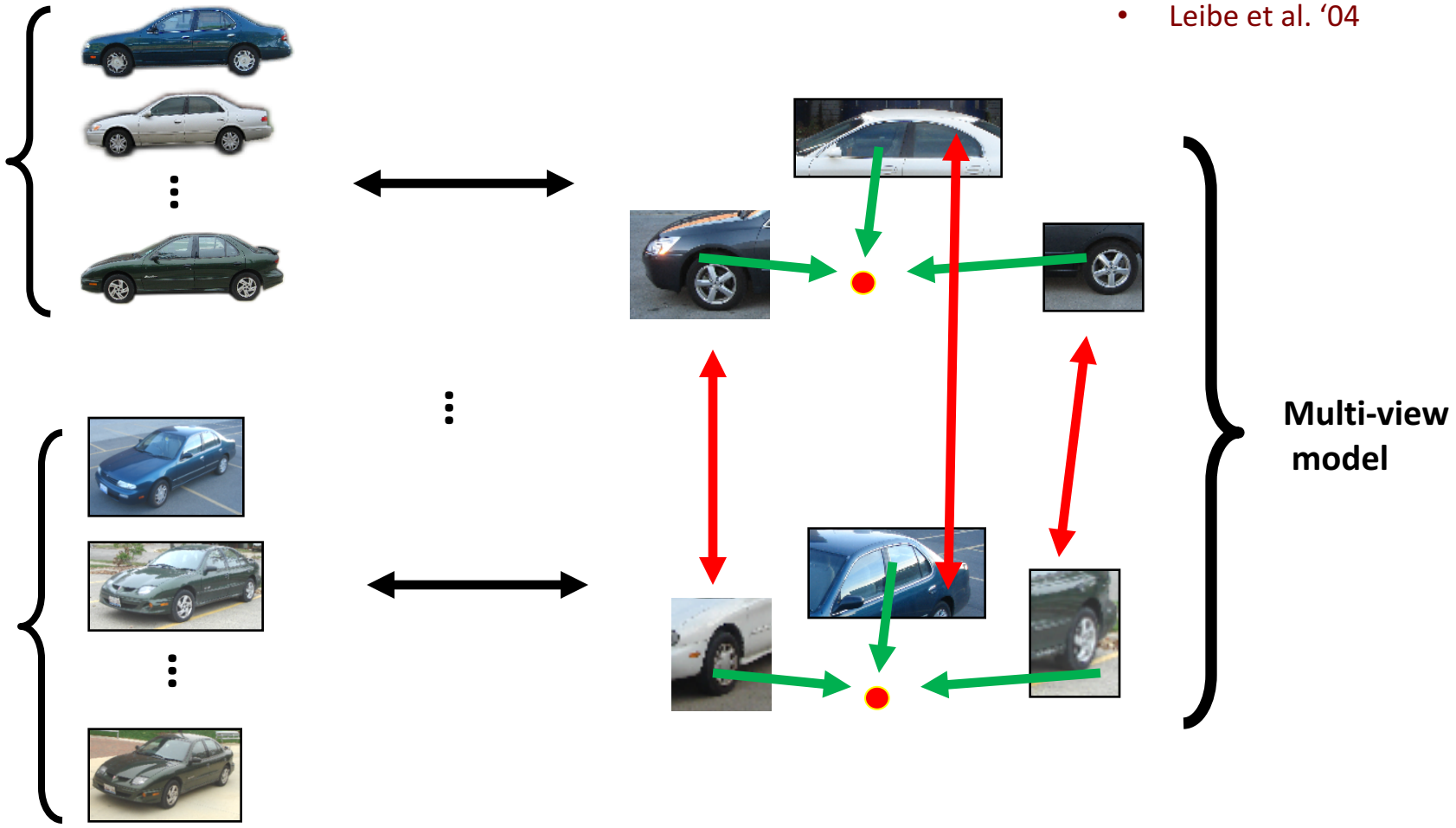


$$l' = F^T x$$

$$x' \in l'$$

Implicit 3D models – built upon the ISM

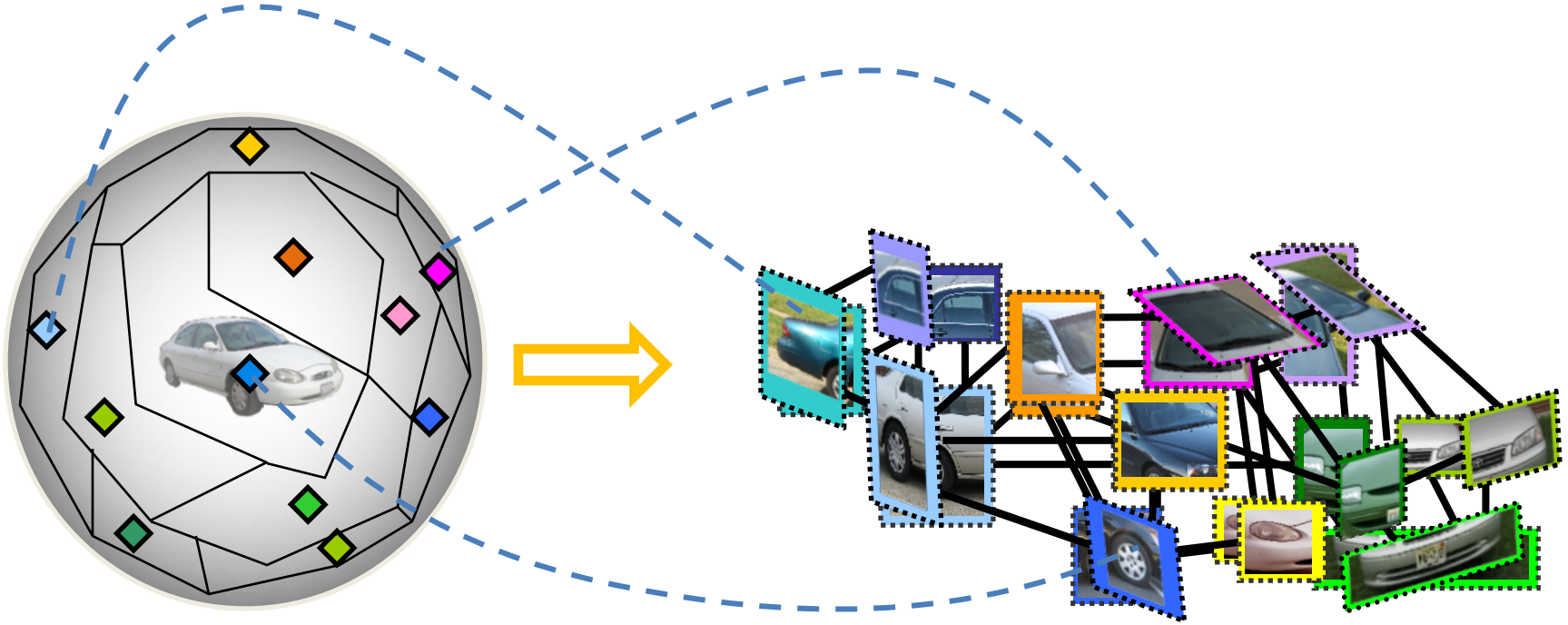
- Thomas et al. ’06
- Leibe et al. ‘04



- Sparse set of interest points or parts of the objects are linked across views.
- These links are used to transfer votes across views
- Each detected codeword votes for the object centroid within nearby views

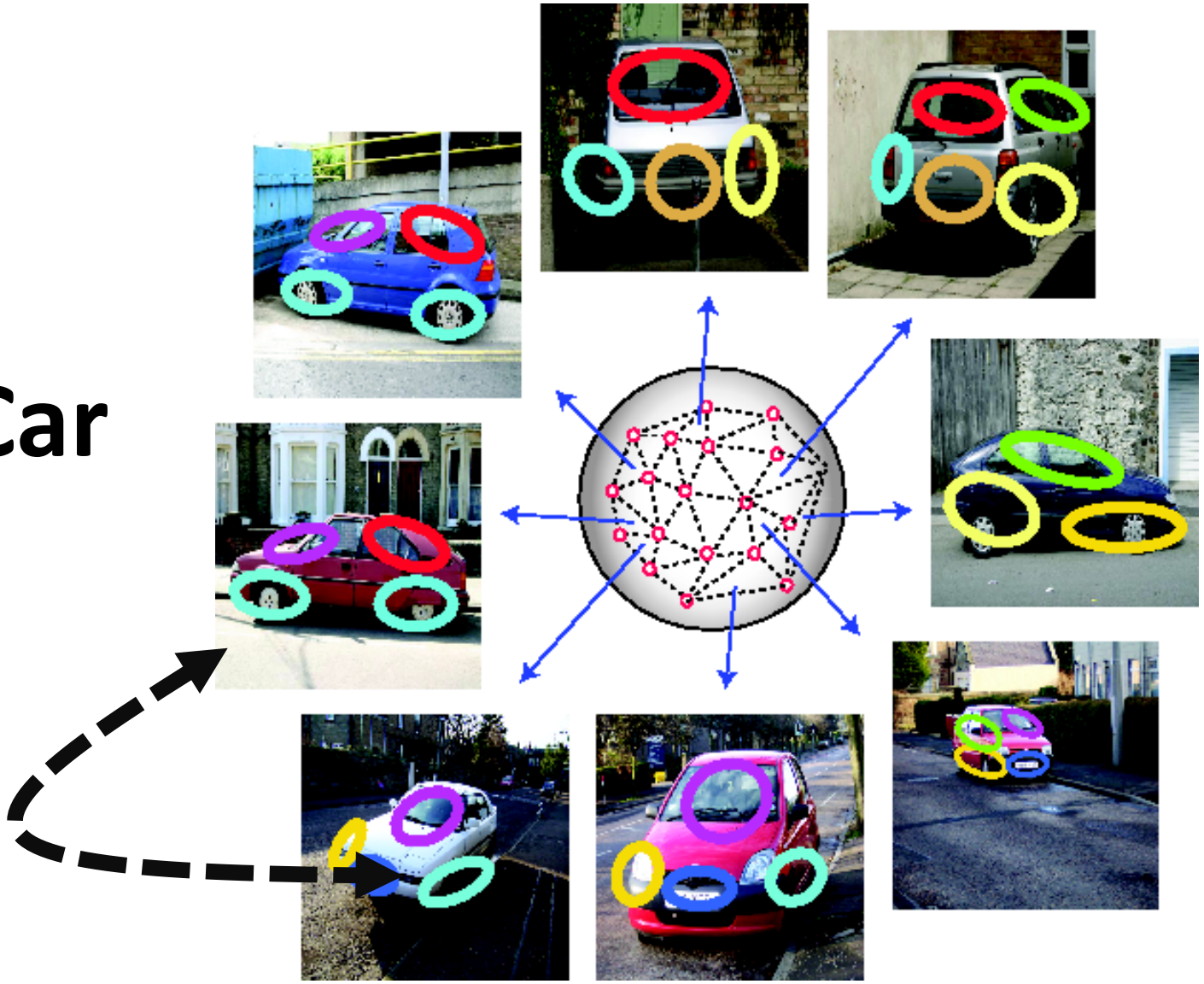
Implicit 3D models – graph-based representations

Savarese, Fei-Fei, ICCV 07
Sun, et al, CVPR 2009, ICCV 09



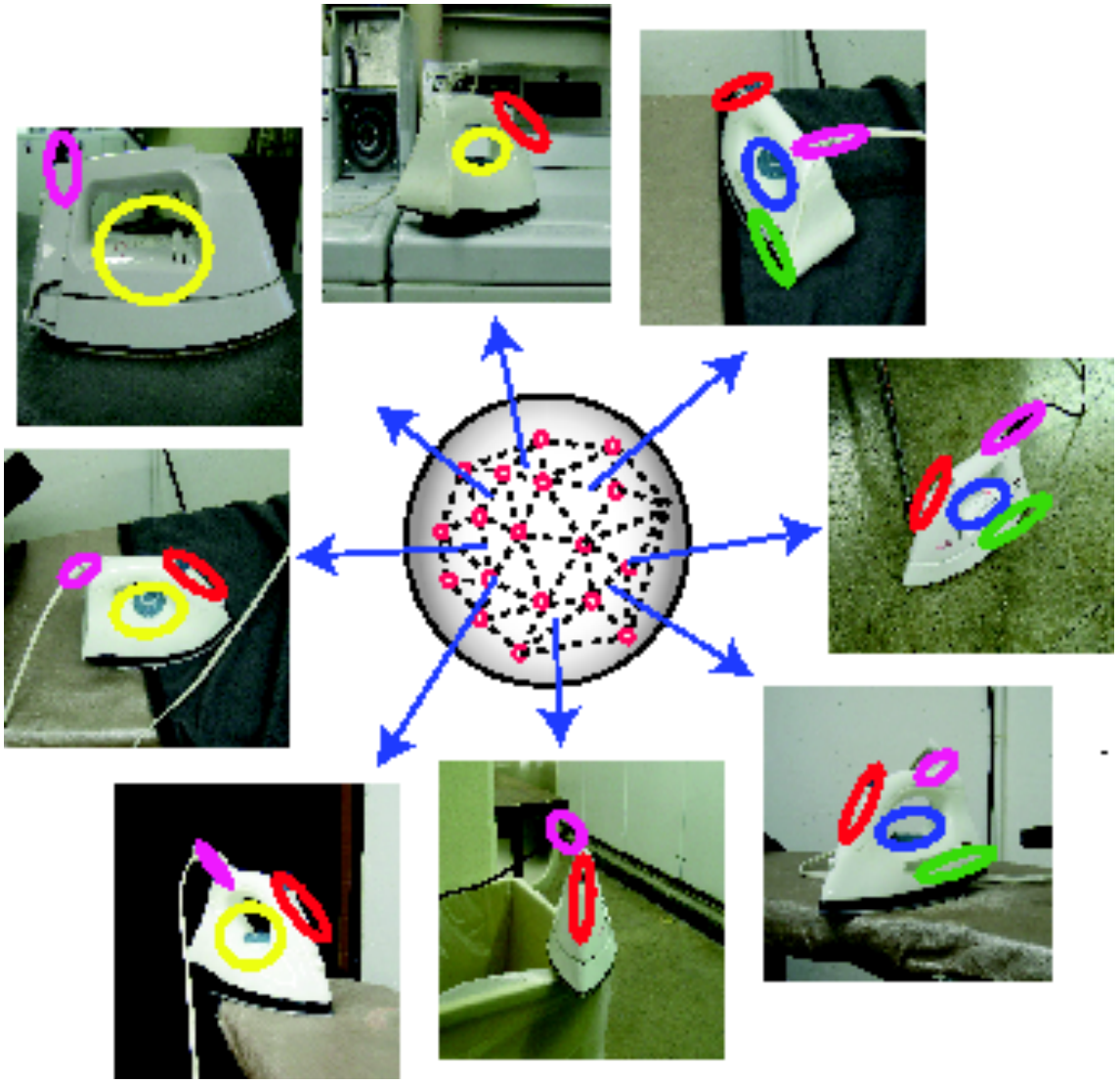
- Canonical parts captures view invariant diagnostic appearance information
- 2d ½ structure linking parts via weak geometry
- Parts and relationship are modeled in a probabilistic fashion
- Semi-supervised: only class labels, not view point or part annotations

Examples of learnt part-based models



Examples of learnt part-based models

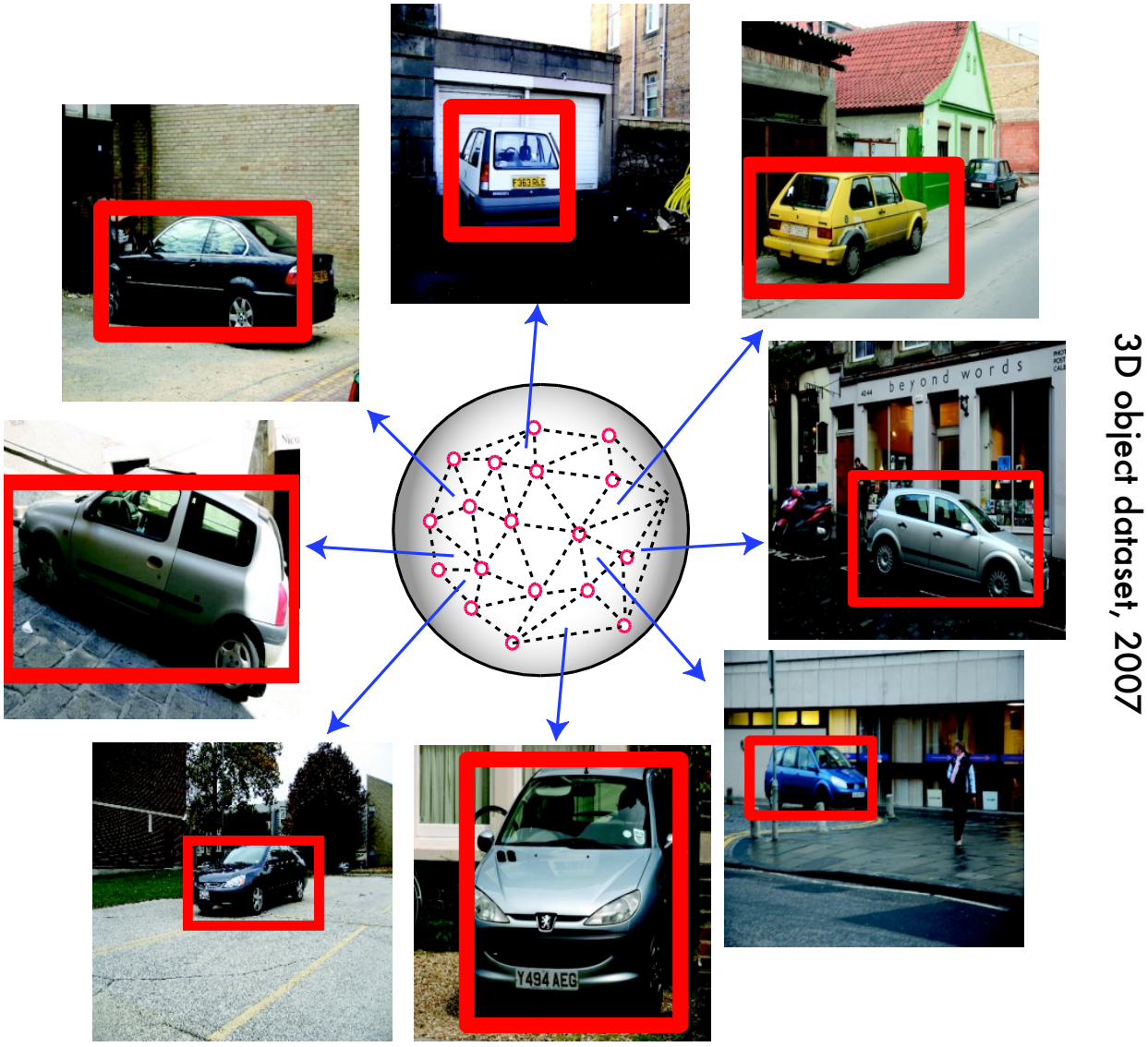
Travel iron



Experimental results

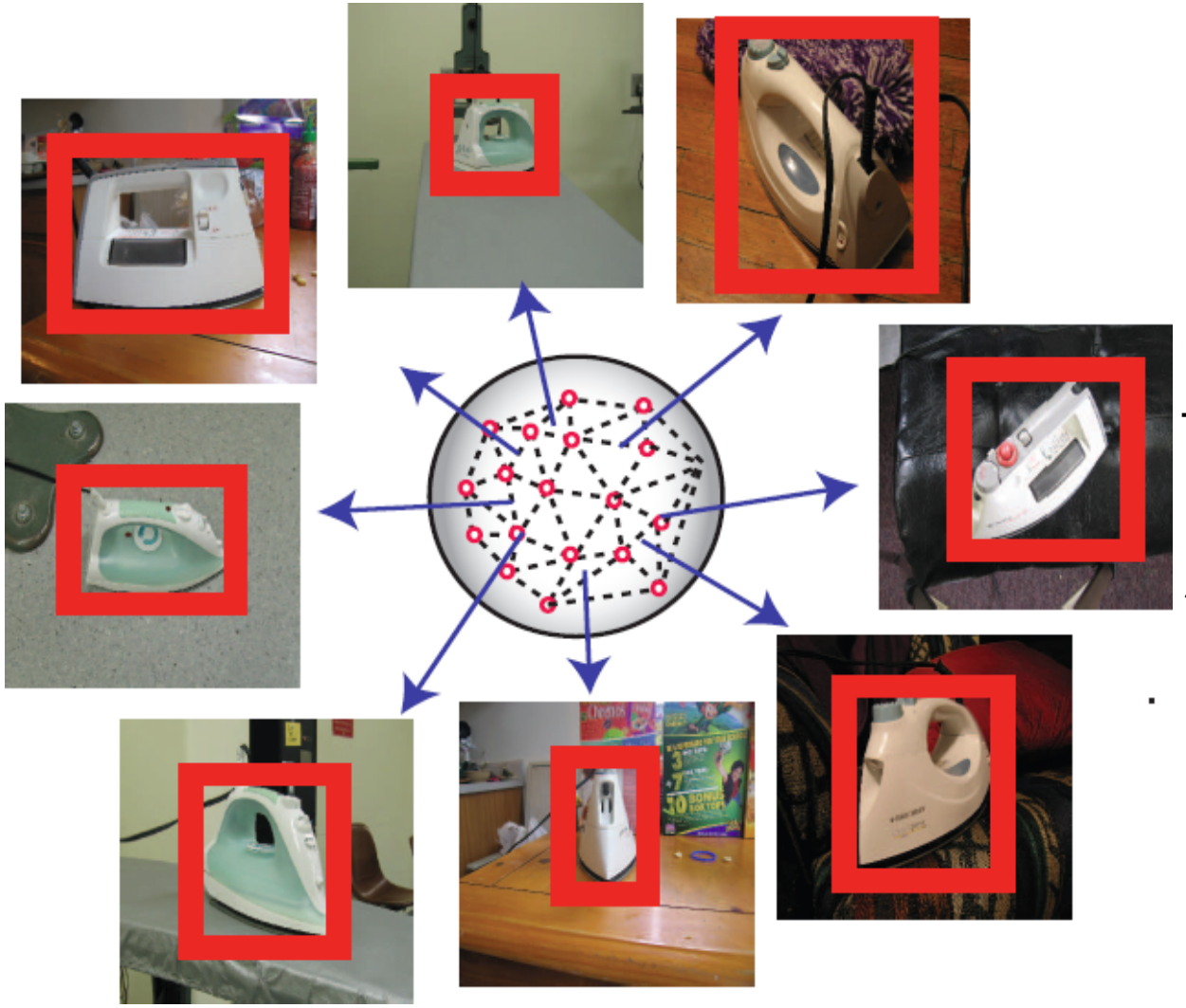
- Object detection from any viewing angles
- Accurate estimation of the object pose
- Synthesis of object appearance from unseen view points

Object detection and pose estimation



3D object dataset, 2007

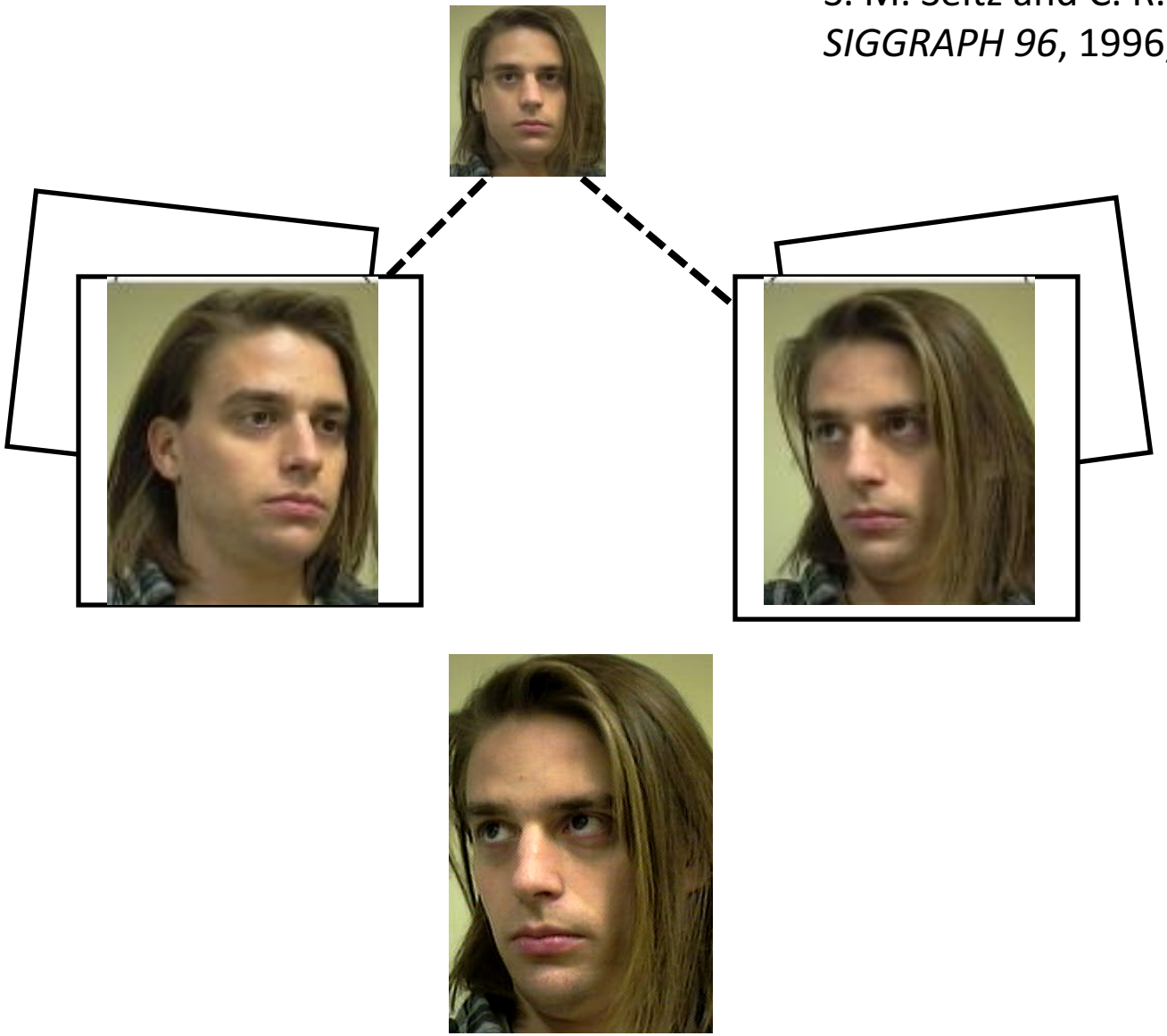
Object detection and pose estimation



3D object dataset, 2007

Synthesizing novel views

S. M. Seitz and C. R. Dyer, *Proc. SIGGRAPH 96*, 1996, 21-30

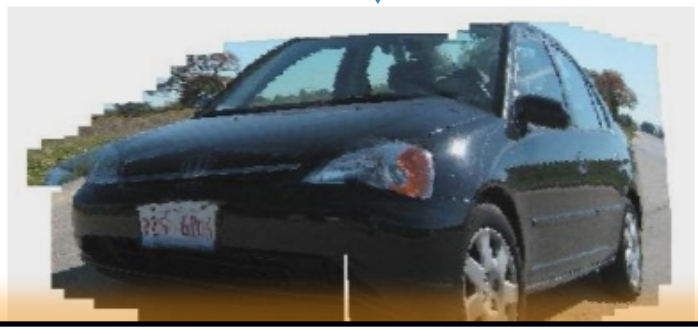


Predicting object appearance from novel views

Affine transformation

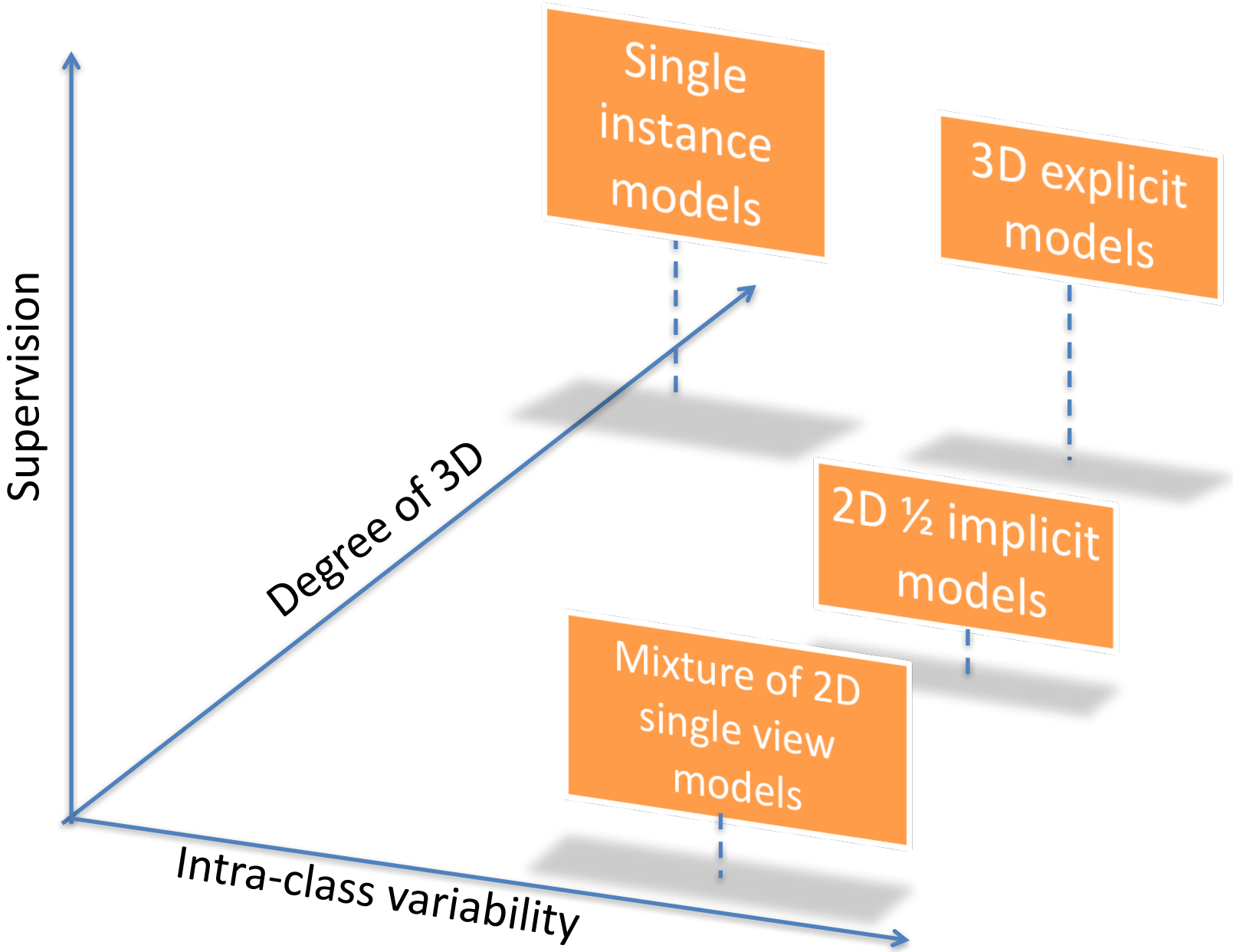
Learnt model

Learnt model



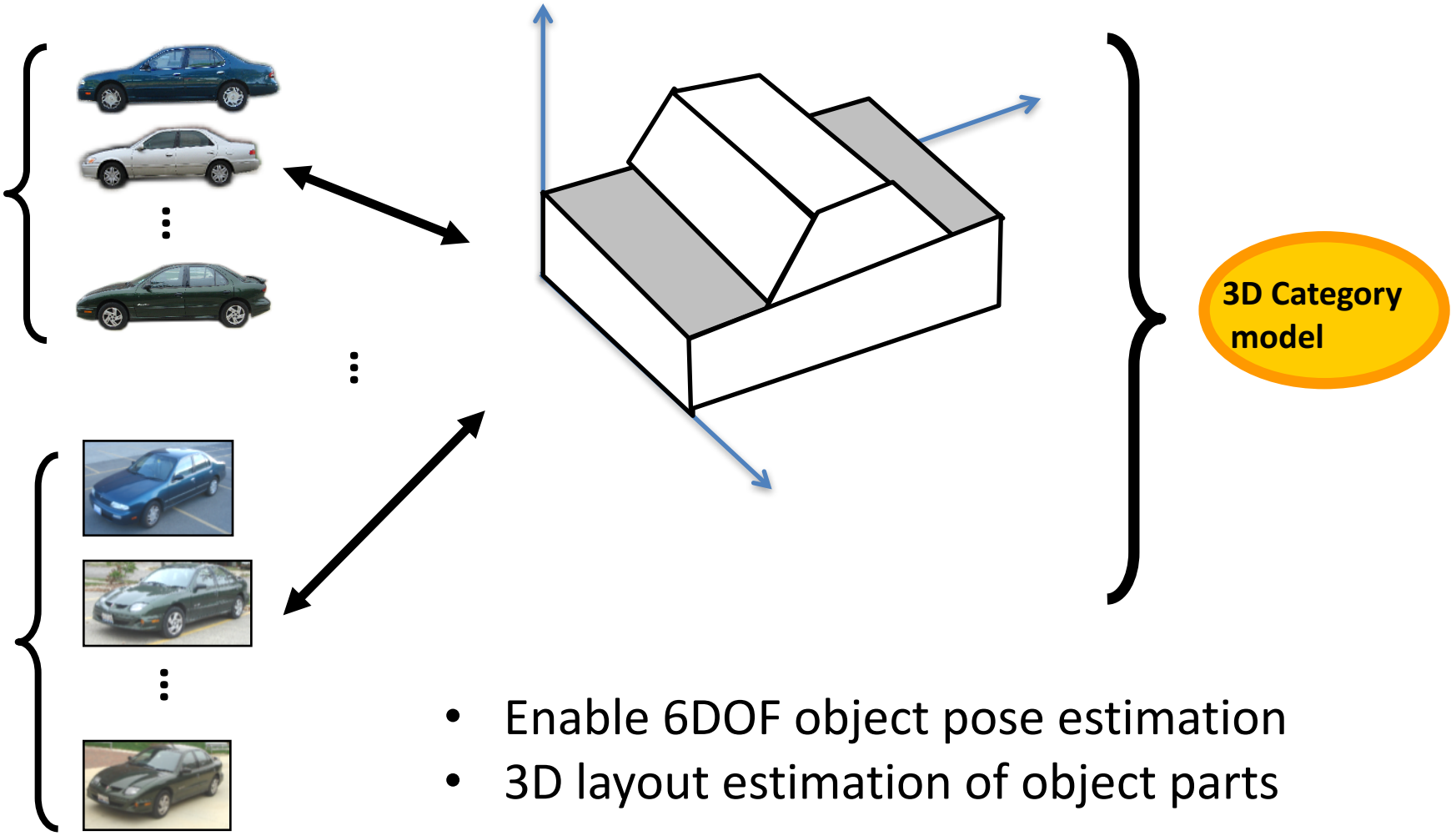
PROS: Flexible and easy to learn • Enable unsupervised discovery of parts
CONS: Limited accuracy • Unable to model part configurations in 3D

Models for 3d Object detection



3D explicit models

- Sun, Xu, Bradski, Savarese, ECCV 2010
- Sun, Kumar, Bradski, Savarese, 3DIM-PVT 2011
- Kumar, Sun, Savarese, CVPR 12
- Xiang & Savarese, CVPR 12
- Hoiem, et al. , '07
- Chiu et al . '07
- Liebelt et al. '08, 10
- Xiao et al . '08
- Yi et al. 09
- Arie-Nachimson & Barsi '09
- Sandhu et al . '09
- Hu & Zhu '10



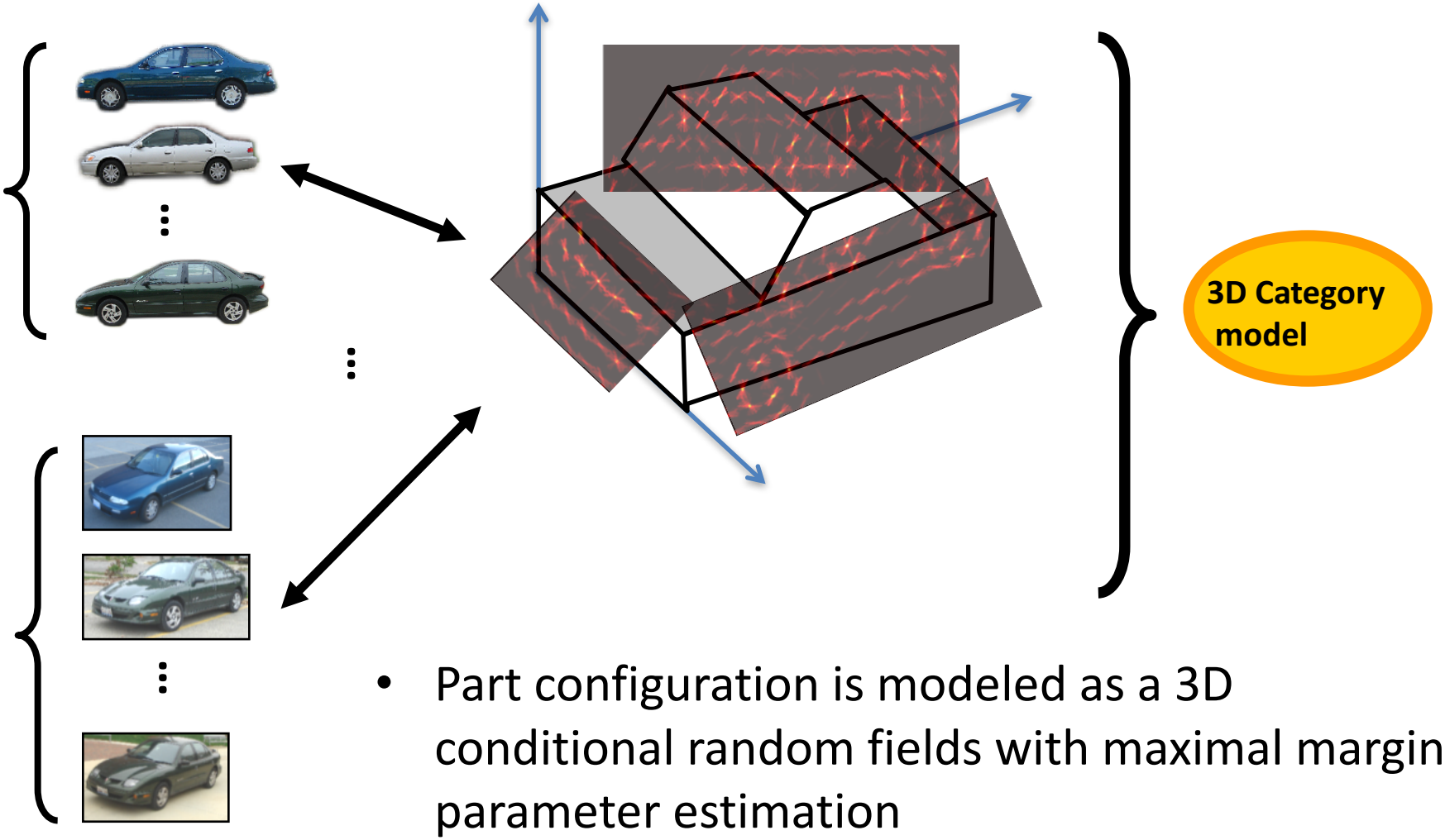
- Enable 6DOF object pose estimation
- 3D layout estimation of object parts

3D explicit models

Yan, et al. '07

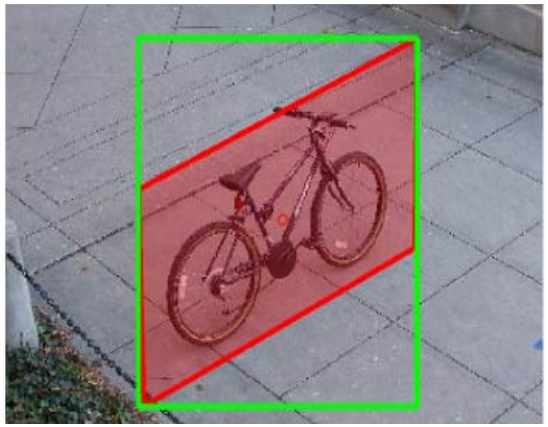
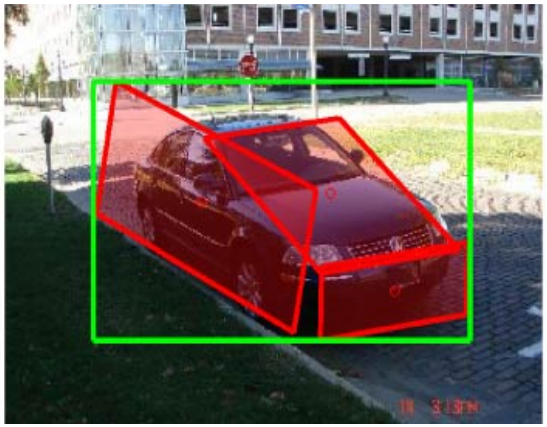
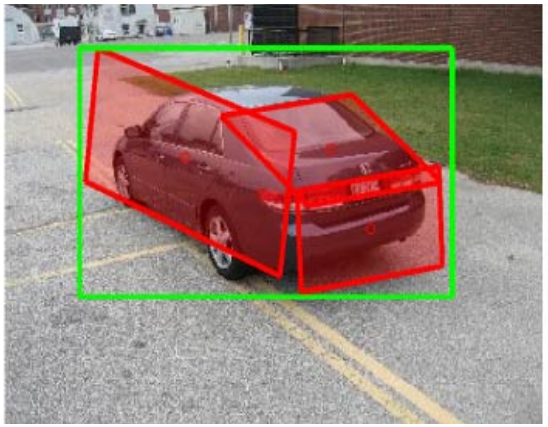
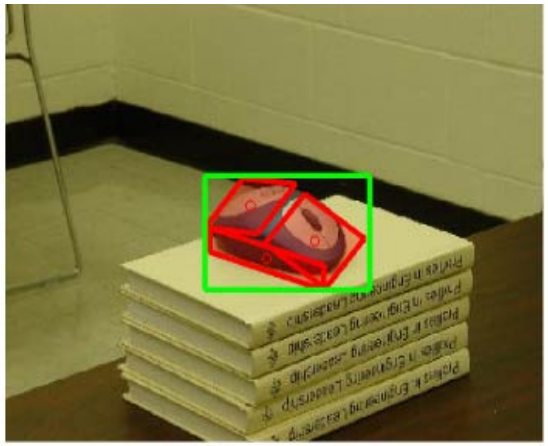
3D explicit models

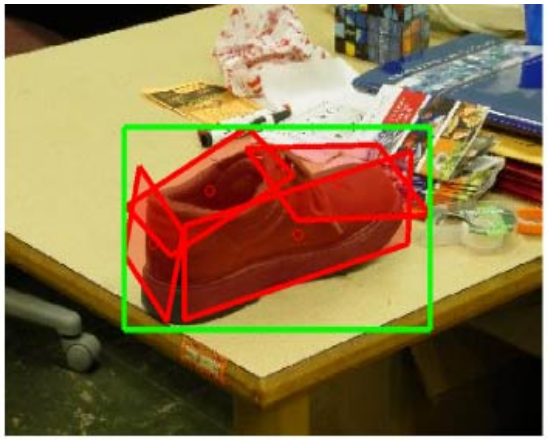
Xiang & Savarese, 2012
Pepik et al. 2013



- Part configuration is modeled as a 3D conditional random fields with maximal margin parameter estimation

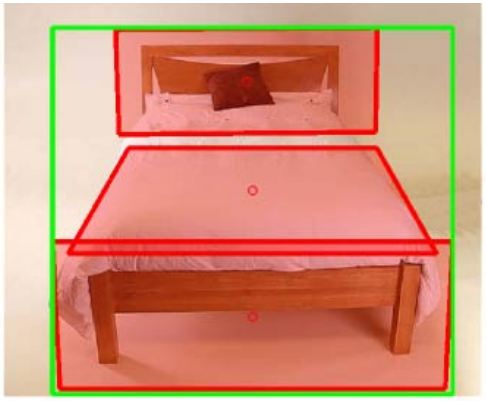
3D object detectors

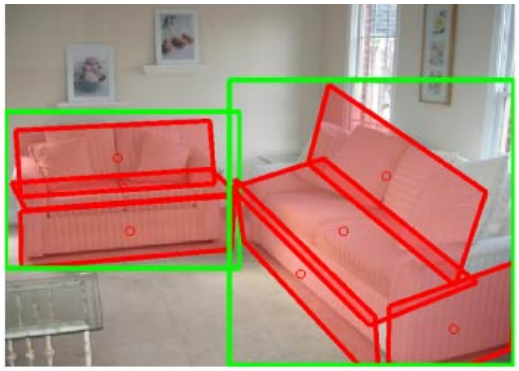





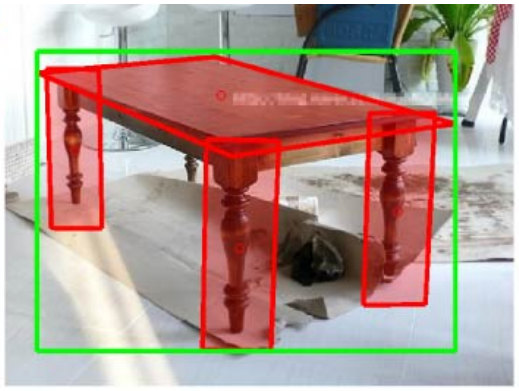


3D object dataset [Savarese & Fei-Fei, ICCV 07]

3D object detectors

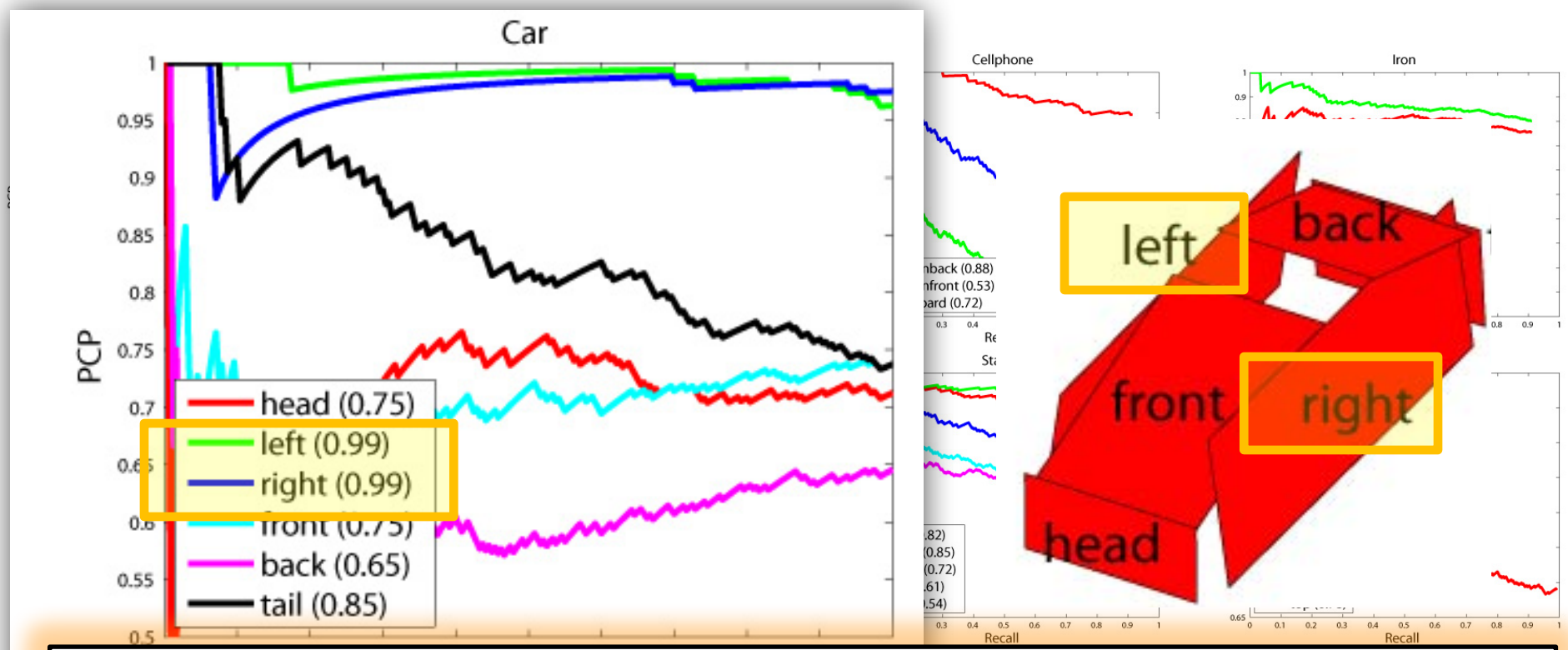




ImageNet dataset [Deng et al. 2010]

3D object detectors

- Part localization on the 3DObject dataset



PROS: Large discrimination power; Able to capture part configurations in 3D
CONS: Require more supervision; slow…

Next lecture

- 3D scene understanding

Agenda on recognition

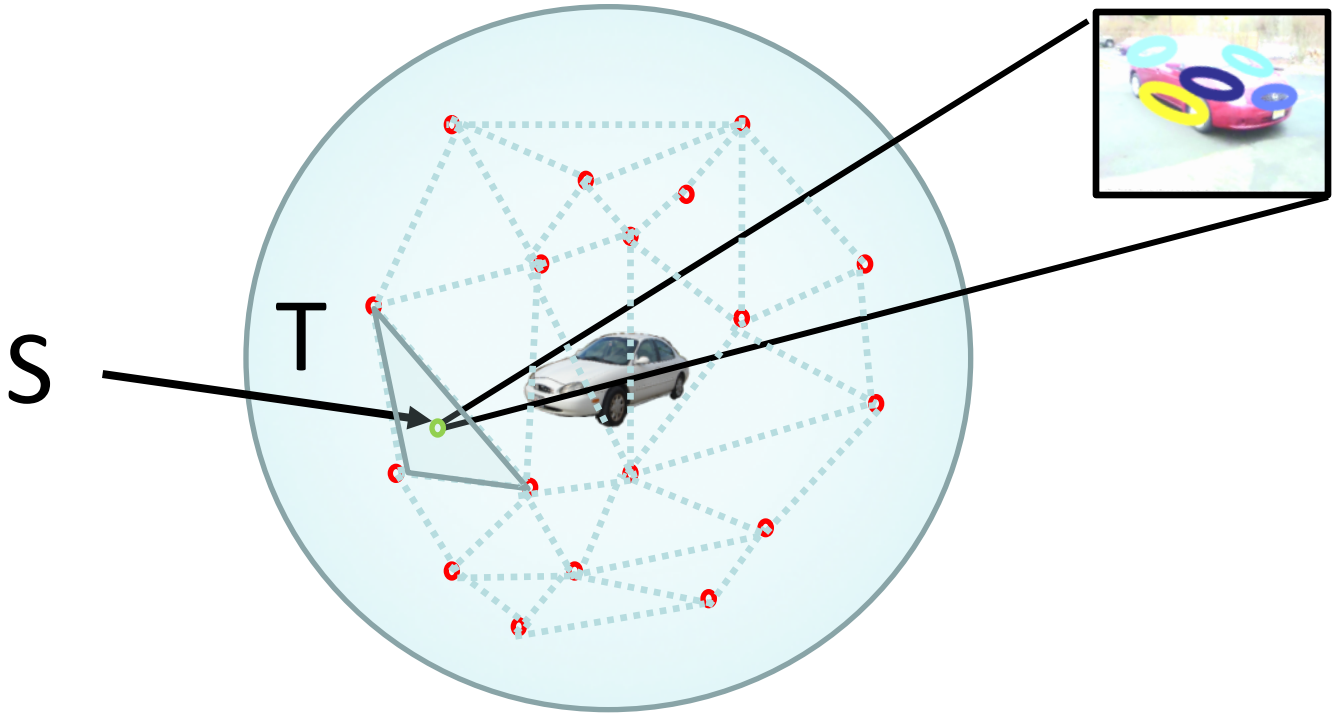
Classification (images; areas)
- bag of words
- Pyramid matching

Detection (use slides with 3 axis: category; supervision; 3D info. Use intro job talk)
- Template based (holistic; part based)
- Multi-view (single instance; categories; 3D pose)

Scene understanding
- Segmentation (bottom up; semantic)
- 3D scene understanding

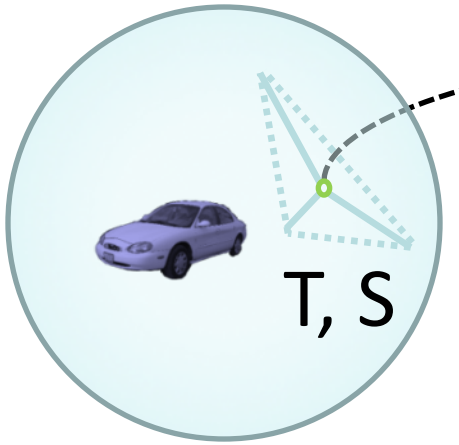
Activity understanding

Parameterization on view-sphere



- Model the object as collection of parts for any T and S on the viewing sphere

Multi-view generative part-based model



α = Part Prop. Prior

$$\pi \sim Dir(\alpha)$$

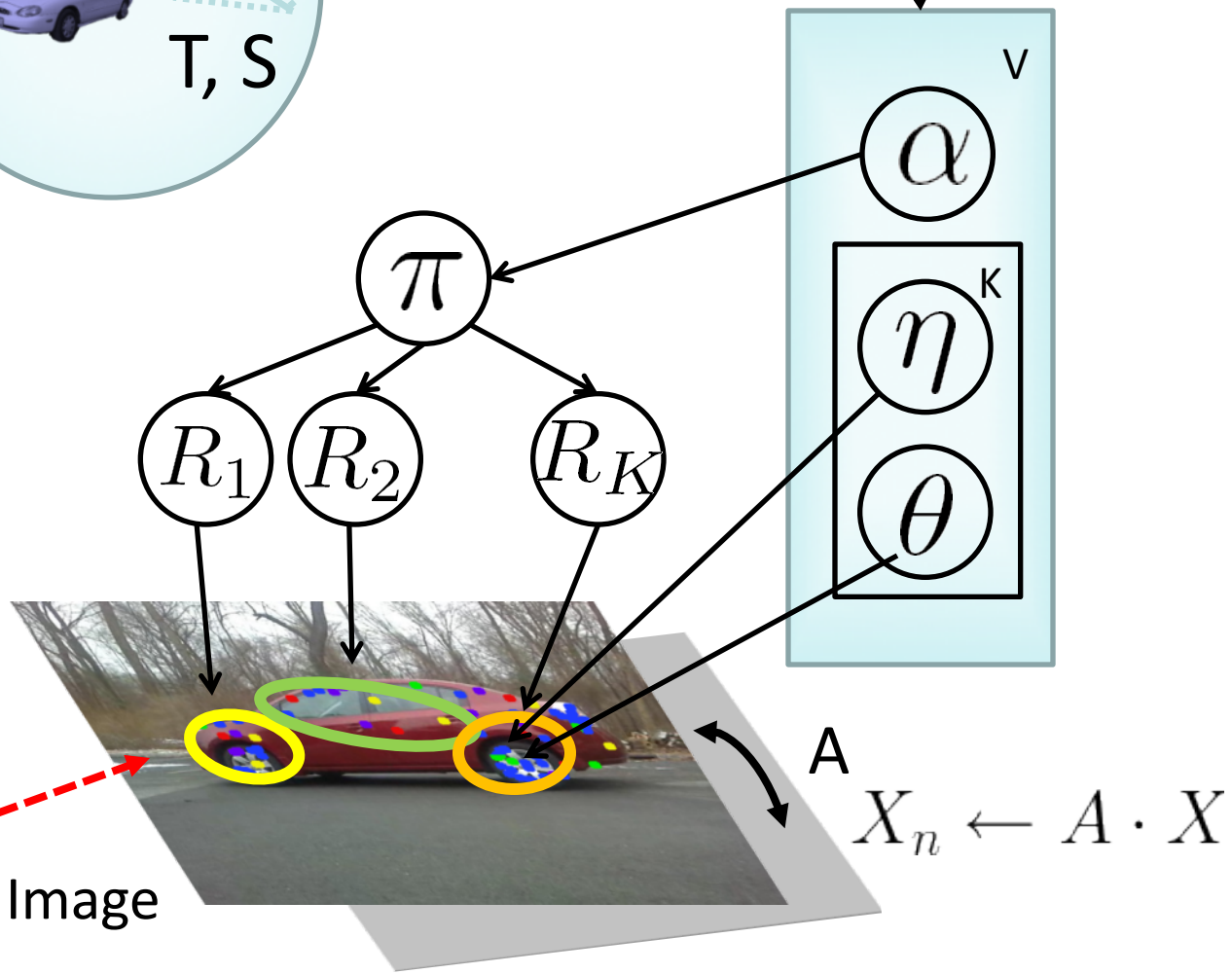
$$R \sim Mult(\pi)$$

$$Y_n \sim Mult(\eta)$$

η = Part Appearance

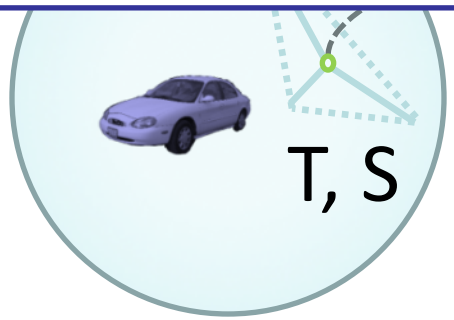
$$X_n \sim N(theta)$$

θ = Part Location/shape



Yn=Codeword
Xn=Location

- Learning: estimate the latent variables and relevant parameters, given the observations
- Variational EM can be used Blei, ICML 2004.



α = Part Prop. Prior

$$\pi \sim Dir(\alpha)$$

$$R \sim Mult(\pi)$$

$$Y_n \sim Mult(\eta)$$

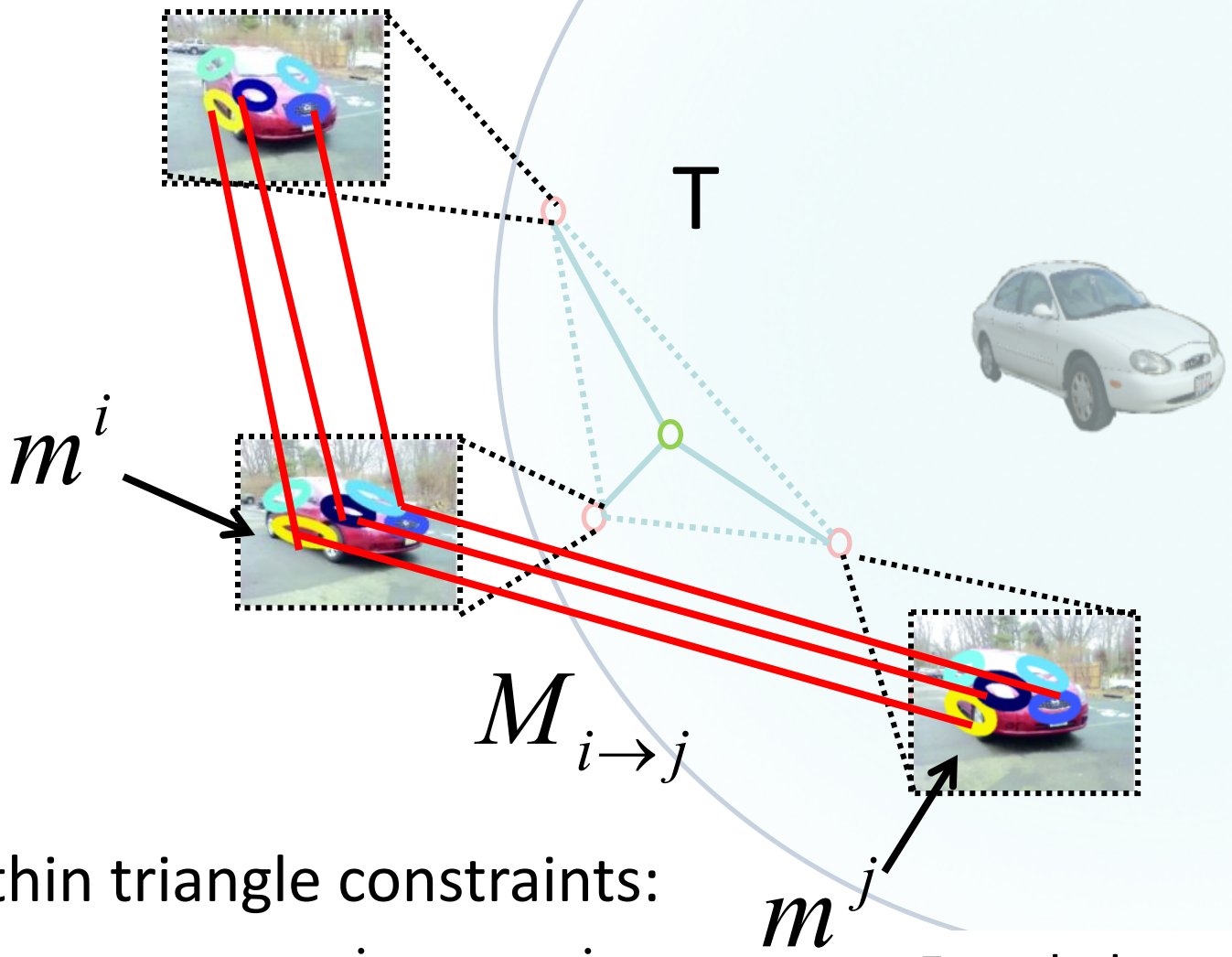
η = Part Appearance

$$X_n \sim N(theta)$$

θ = Part Location/shape

Yn=Codeword
Xn=Location

Incorporating geometrical constraints

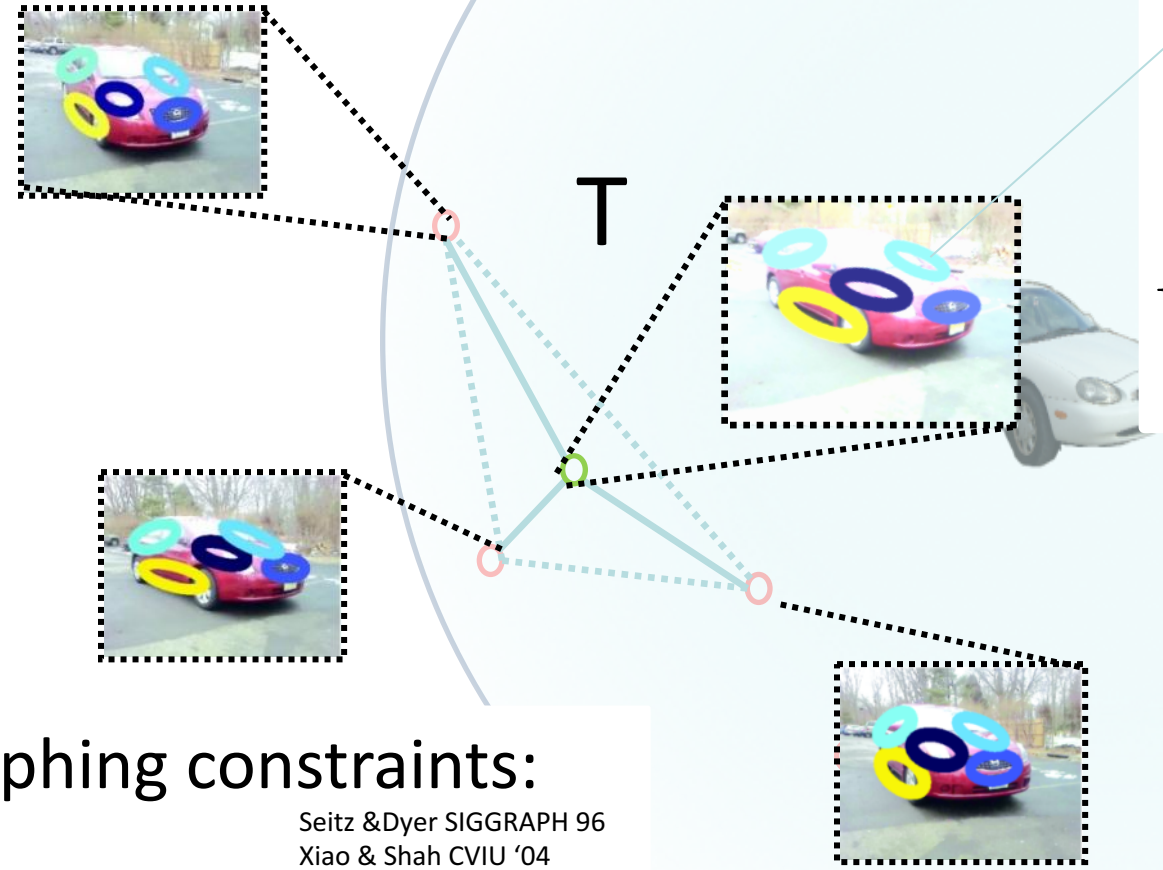


Within triangle constraints:

$$M_{i\to j}\cdot m^i \approx m^j$$

Encoded as a penalty term in variational *EM*

Incorporating geometrical constraints



$m(S)$ = Center

$W(S)$ = Shape

$$\begin{cases} \Sigma = WW^T \\ \theta = (m, \Sigma) \end{cases}$$

View morphing constraints:

Seitz &Dyer SIGGRAPH 96
Xiao & Shah CVIU '04

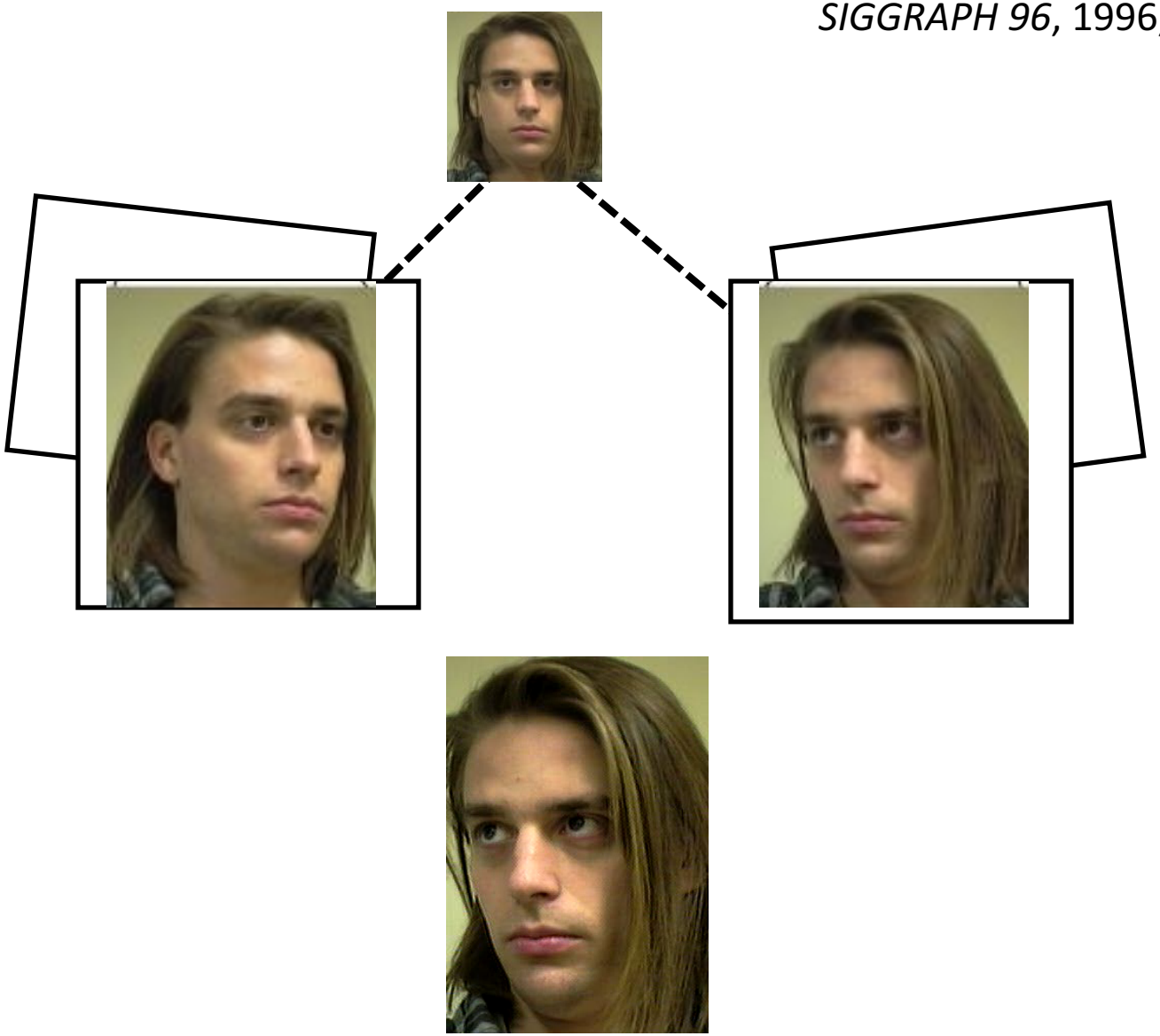
$$m(S) = \sum_{g=1}^{3} m_T^g \cdot s_g$$

$$W(S) = \sum_{g=1}^{3} W_T^g \cdot s_g$$

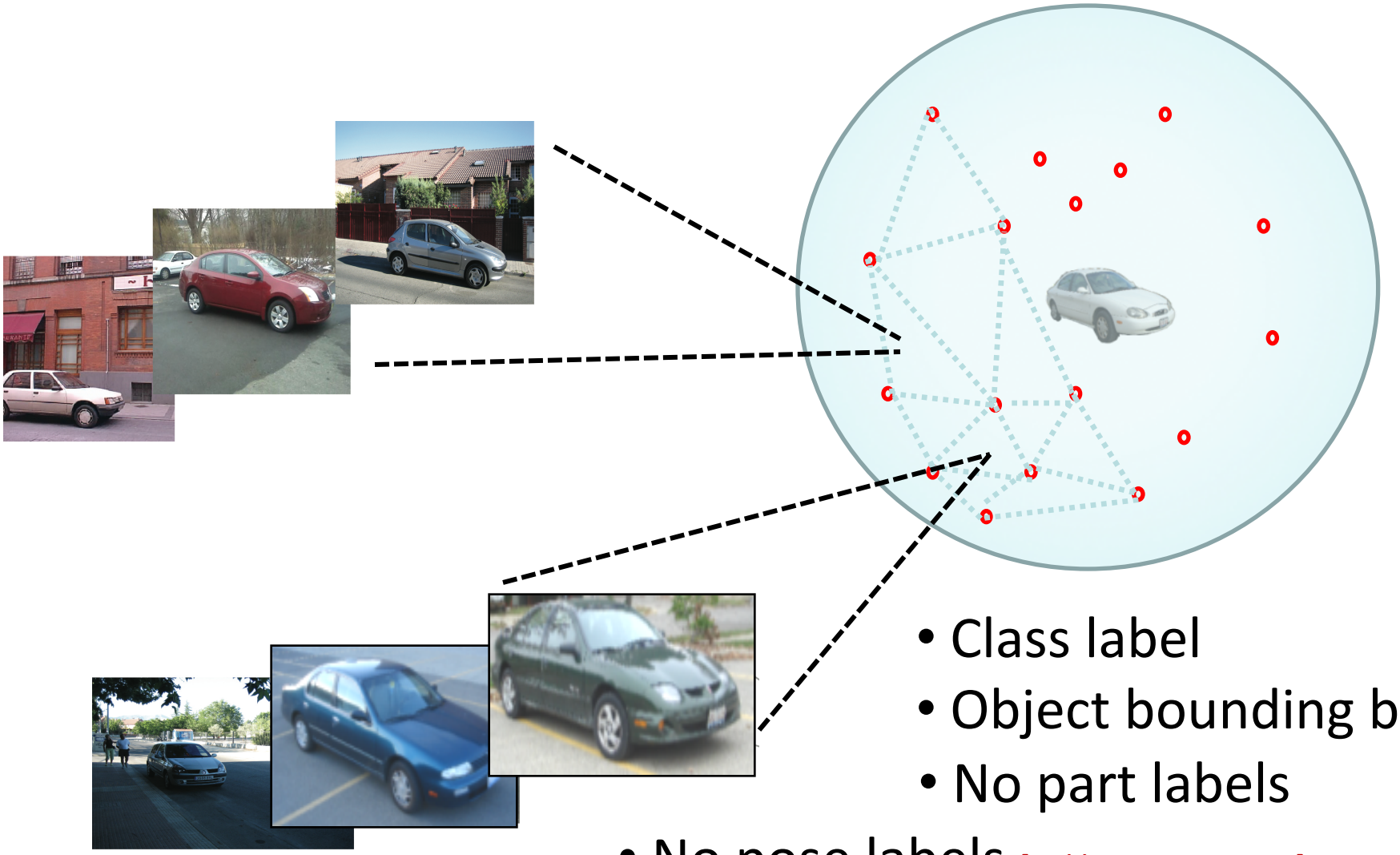
Encoded as a penalty term in variational *EM*

Incorporating geometrical constraints

S. M. Seitz and C. R. Dyer, *Proc. SIGGRAPH 96*, 1996, 21-30



Semi-supervised

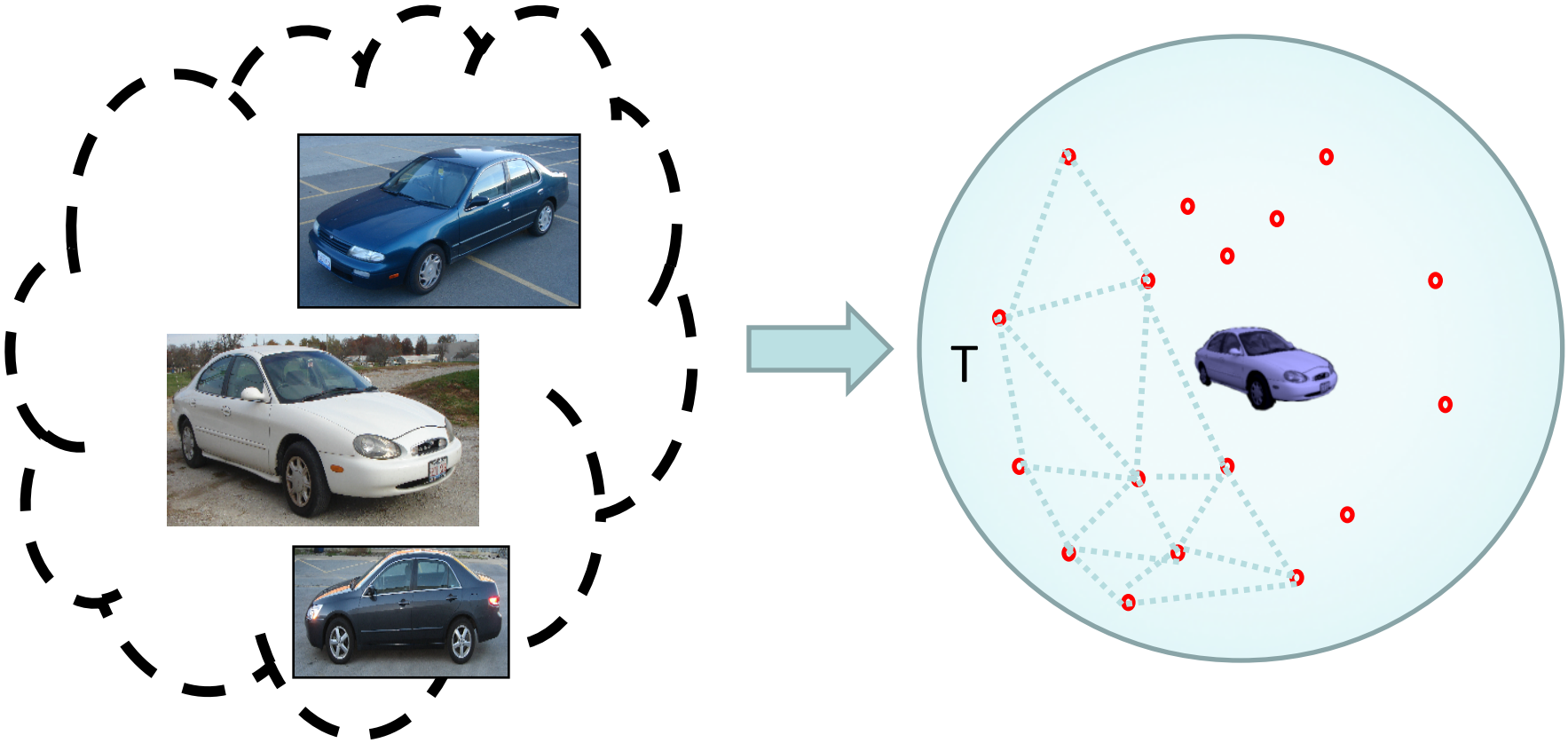


- Class label
- Object bounding box
- No part labels

- No pose labels [unlike Sun CVPR 09]

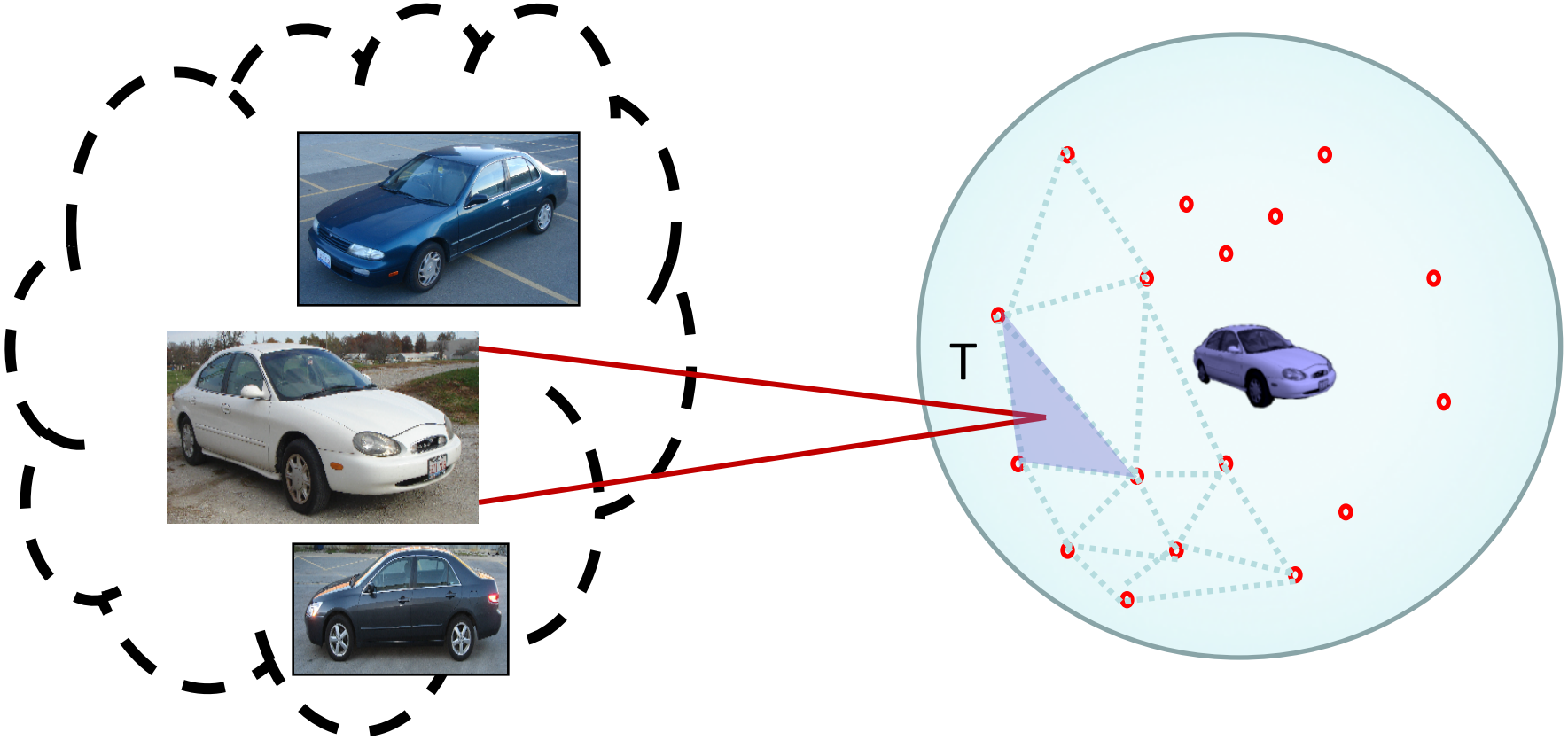
- No need to observe same object instance from multiple views
[unlike Savarese & Fei-Fei, 07, 08]

Incremental learning



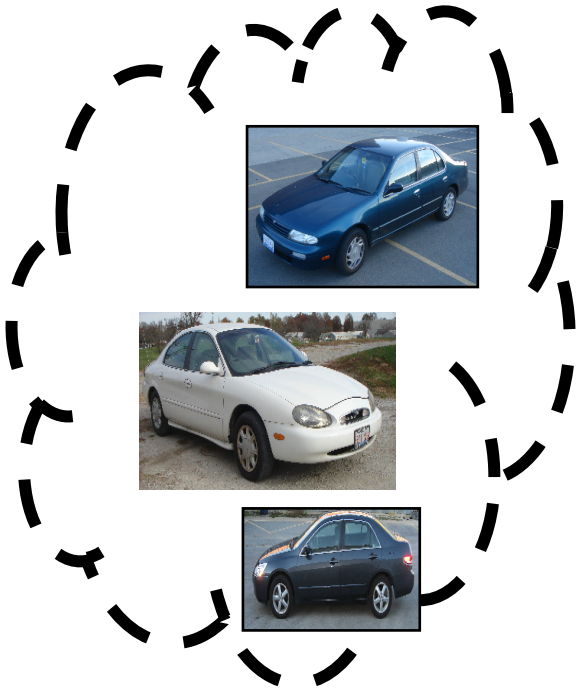
- Enable unorganized and on-line collection training images
- Increase efficiency in learning (no need large storage space)

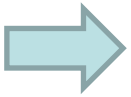
Incremental learning



- Assign new training image to a triangle of the view sphere
- Evidence of training image is used to update model parameters
- Re-estimate sufficient statistics in a iterative fashion

Evolution of learnt parts





3D object detectors

- Best results up-to-date in pose estimation and 3D part estimation

Cars from 3D Object dataset [Savarese 07]

Method	ours	[1]	[2]	[3]	[4]	[5]	[6]
Viewpoint (cars)	**93.4 %**	85.4	85.3	81	70	67	48.5

Cars from EPFL dataset [Ozuysal 09]

Method	ours	Ours - baseline	DPM [7]	[8]
Viewpoint (cars)	**64.8%**	58.1	56.6	41.6

Chairs, tables and beds from IMAGE NET [Deng et al. CVPR09]

Method	ours	Ours - baseline	DPM [7]
Viewpoint	**63.4%**	34.0	49.5

[1] N. Payet and S. Todorovic. From contours to 3d object detection and pose estimation. In ICCV, 2011.
[2] D. Glasner, M. Galun, S. Alpert, R. Basri, and G. Shakhnarovich. Viewpoint-aware object detection and pose estimation. In ICCV, 2011.
[3] M. Stark, M. Goesele, and B. Schiele. Back to the future: Learning shape models from 3d cad data. In BMVC, 2010.
[4] J. Liebelt and C. Schmid. Multi-view object class detection with a 3D geometric model. In CVPR, 2010.
[5] H. Su, M. Sun, L. Fei-Fei, and S. Savarese. Learning a dense multiview representation for detection, viewpoint classification. In ICCV, 2009.
[6] M. Arie-Nachimson and R. Basri. Constructing implicit 3d shape models for pose estimation. In ICCV, 2009.
[7] P. Felzenszwalb, R. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part-based models. TPAMI, 2010.
[8] M. Ozuysal, V. Lepetit, and P. Fua. Pose estimation for category specific multiview object localization. In CVPR, 2009.

3D object detectors

- Part localization on the 3DObject dataset