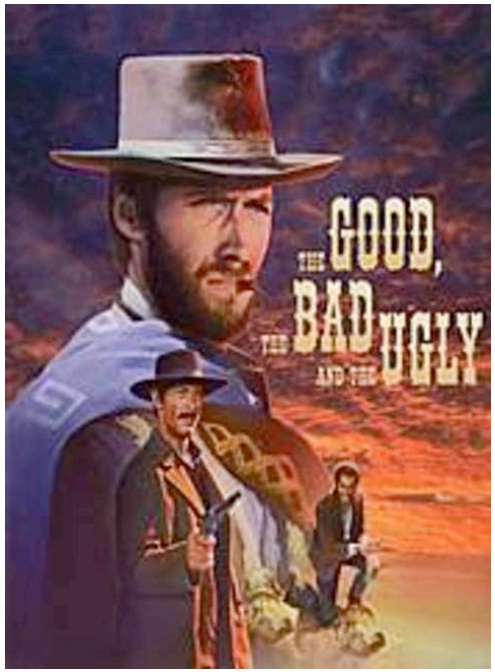
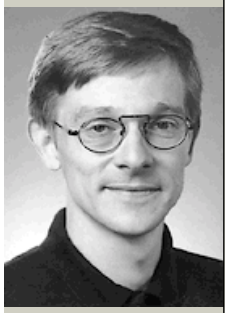



# What are Ideal Parts for Part-Based Object Models?

# the Good, the Bad, and the Ugly

**Bernt Schiele**

**Max Planck Institute for Informatics**

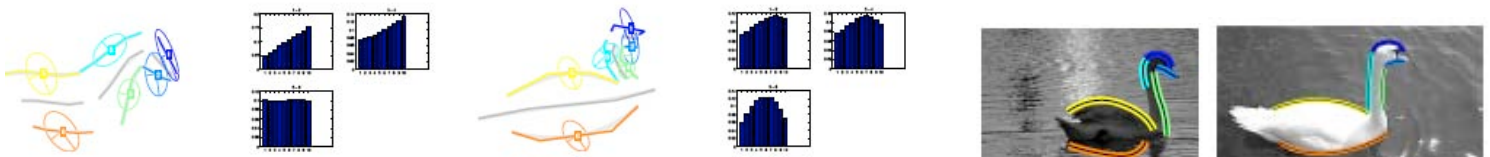


**thanks to: Micha Andriluka, Bastian Leibe, Sandra Ebert, Mario Fritz, Diane Larlus, Marcus Rohrbach, Paul Schnitzspan, Stefan Roth, Michael Stark, Michael Goesele**

# A Key Challenge for Object Recognition

- **Scalability: Large Scale Object Class Recognition**
  - goal: from 100's of classes to 1'000 and 10'000's of classes
  - methods:
    - unsupervised & semi-supervised methods for large-scale mining of multimodal databases (e.g. web, images, videos)
    - learning of object class hierarchies
    - use language to support learning and to derive semantic information
    - knowledge transfer across object classes
    - ...

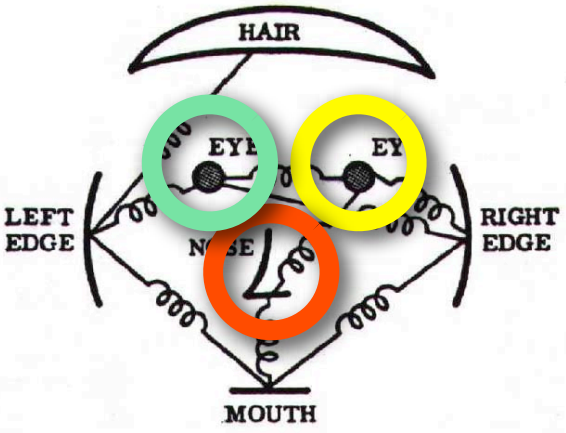
Knowledge Transfer from Horse to Swan [Stark,Goesele,Schiele@iccv09]
**(based on part-configurations):**

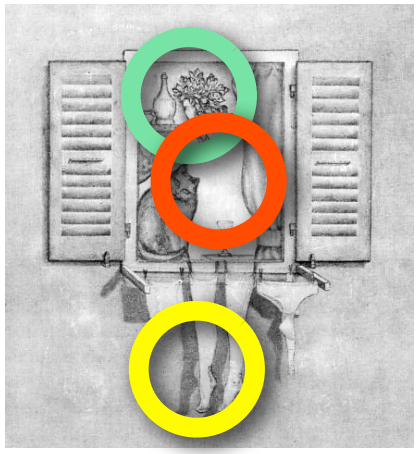


(a) Partial horse model (b) Combined swan model

# Class of Object Models: Part-Based Models / Pictorial Structures

- Pictorial Structures [Fischler & Elschlager 1973]
  - Model has two components
    - parts (2D image fragments)
    - structure (configuration of parts)

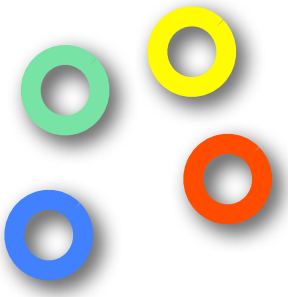


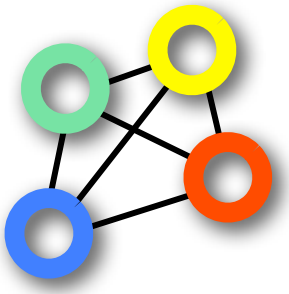
# “State-of-the-Art” in Object Class Representations

- Bag of Words Models (BoW)
  - object model = histogram of local features
  - e.g. local feature around interest points
- Global Object Models
  - object model = global feature object feature
  - e.g. HOG (Histogram of Oriented Gradients)
- Part-Based Object Models
  - object model = models of parts & spatial topology model
  - e.g. constellation model or ISM (Implicit Shape Model)
- **But: What is the Ideal Notion of Parts here?**
- **And: Should those Parts be Semantic?**

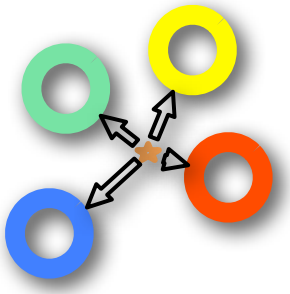
BoW: no spatial relationships



e.g. HOG: fixed spatial relationships



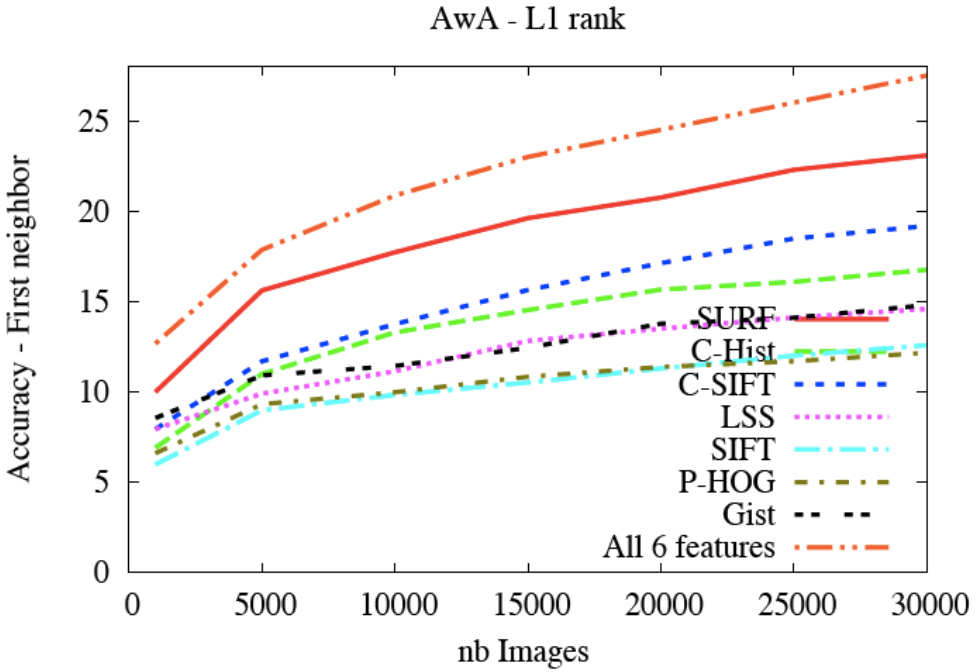
e.g. ISM: flexible spatial relationships



[Larlus,Ebert,Schiele@eccv10]

# Work on Semi-Supervised-Learning

- AwA (Animals with Attributes) [Lambert@cvpr09]
  - 50 animals classes




- Observations of Nearest Neighbor Quality
  - more & combining features is better
  - more images are better !
  - but: today's object class representations are not good enough !

# What are Good Object Representations?

- Parts-Based Object Models?
- Attribute-Based Models?

- Multiple motivations for such models exist:
  - **intuitiveness**: semantic meaning of parts/attributes is attractive (e.g. enables use of language sources)
  - **scalability**: transferability of parts/attributes across classes
  - **learnability**: sharing of parts/attributes across instances/classes
  - ...
- But:
  - **What is the ideal / correct notion of parts & attributes?**
  - **How semantic should those be?**

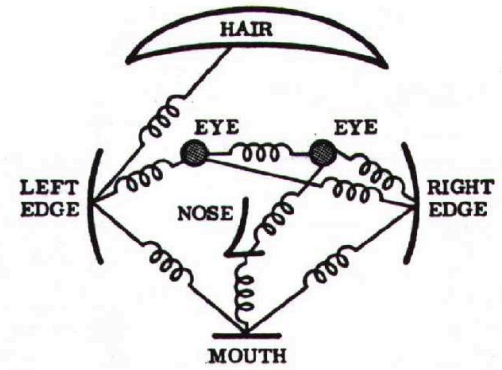
# Overview

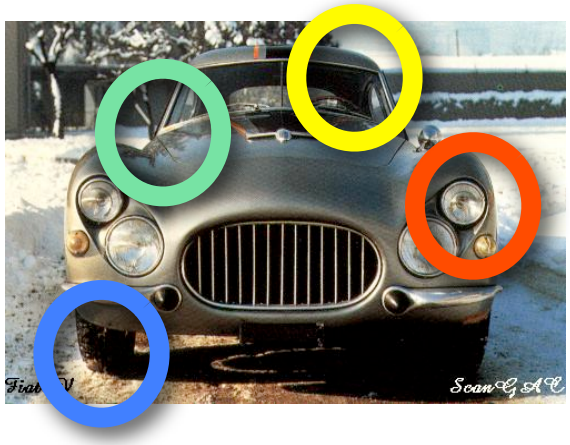
- What are Ideal Parts for Part-Based Object models
- Part-Based Models for Object and People Detection
  - Implicit Shape Model [bmvc03,ijcv08]
  - Pictorial Structures Model for Articulated Pose Estimation [cvpr09]
  - Hierarchical Latent CRF Model for Objects [cvpr10]
  - Learning Shape Models from 3D CAD Data [bmvc10]
- Discussion
  - Semantic vs. Non-Semantic Object Parts

# Overview

- What are Ideal Parts for Part-Based Object models
- Part-Based Models for Object and People Detection
  - Implicit Shape Model [bmvc03,ijcv08]
  - Pictorial Structures Model for Articulated Pose Estimation [cvpr09]
  - Hierarchical Latent CRF Model for Objects [cvpr10]
  - Learning Shape Models from 3D CAD Data [bmvc10]
- Discussion
  - Semantic vs. Non-Semantic Object Parts

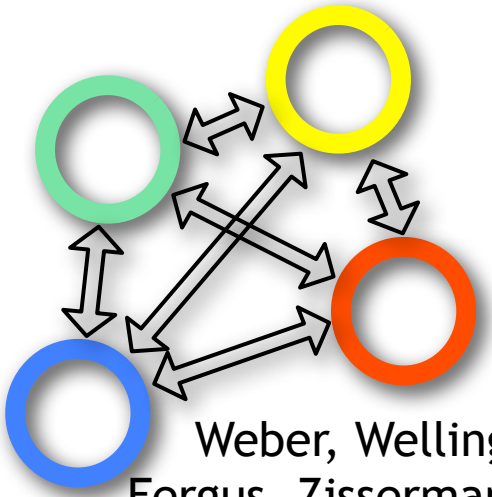
# Model of Parts & Structure: Constellation Model vs. Implicit Shape Model



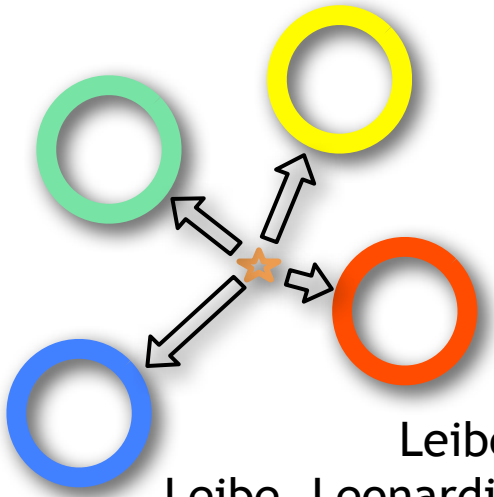



Constellation Model:
Fully connected shape model

Implicit Shape Model:
Star-Model w.r.t. Reference Point

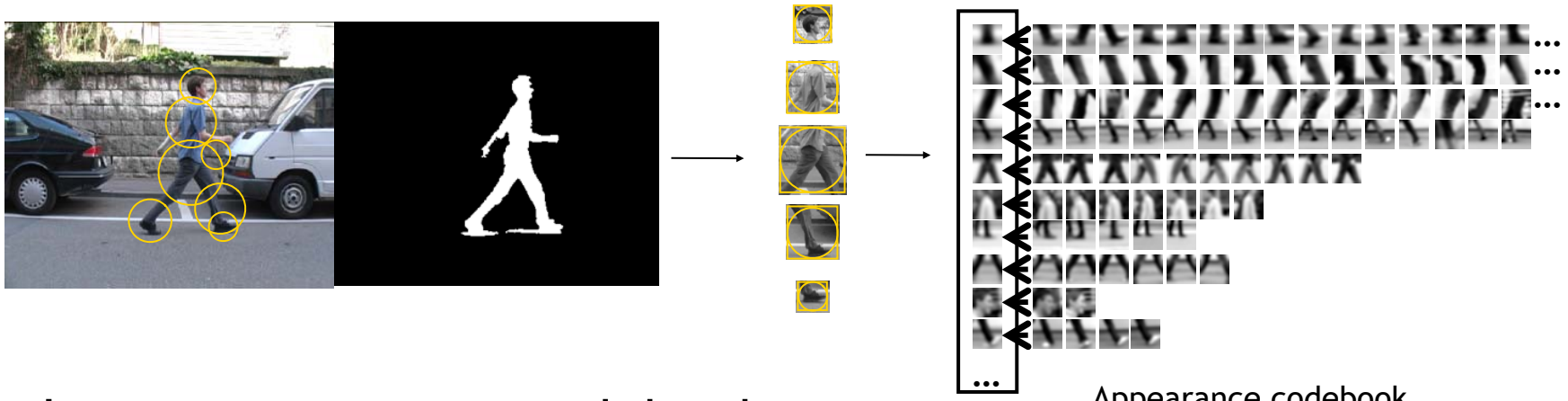


Weber, Welling, Perona ’00
Fergus, Zisserman, Perona ’03



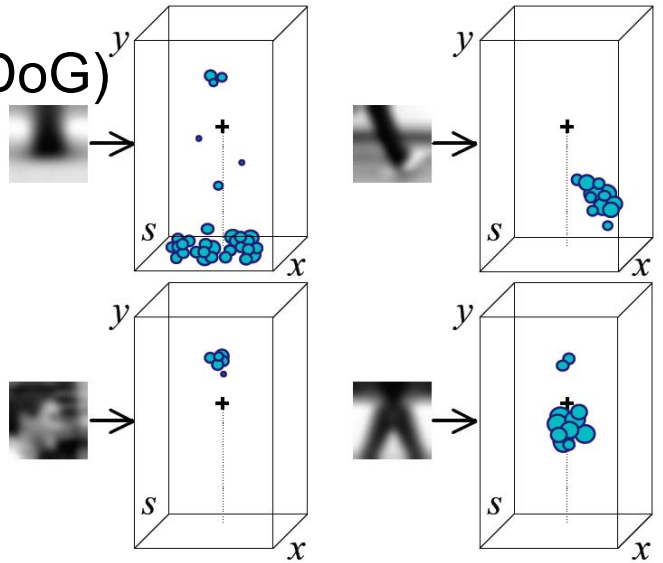
Leibe, Schiele ’03
Leibe, Leonardis, Schiele ’04

# Implicit Shape Model - Representation



Appearance codebook

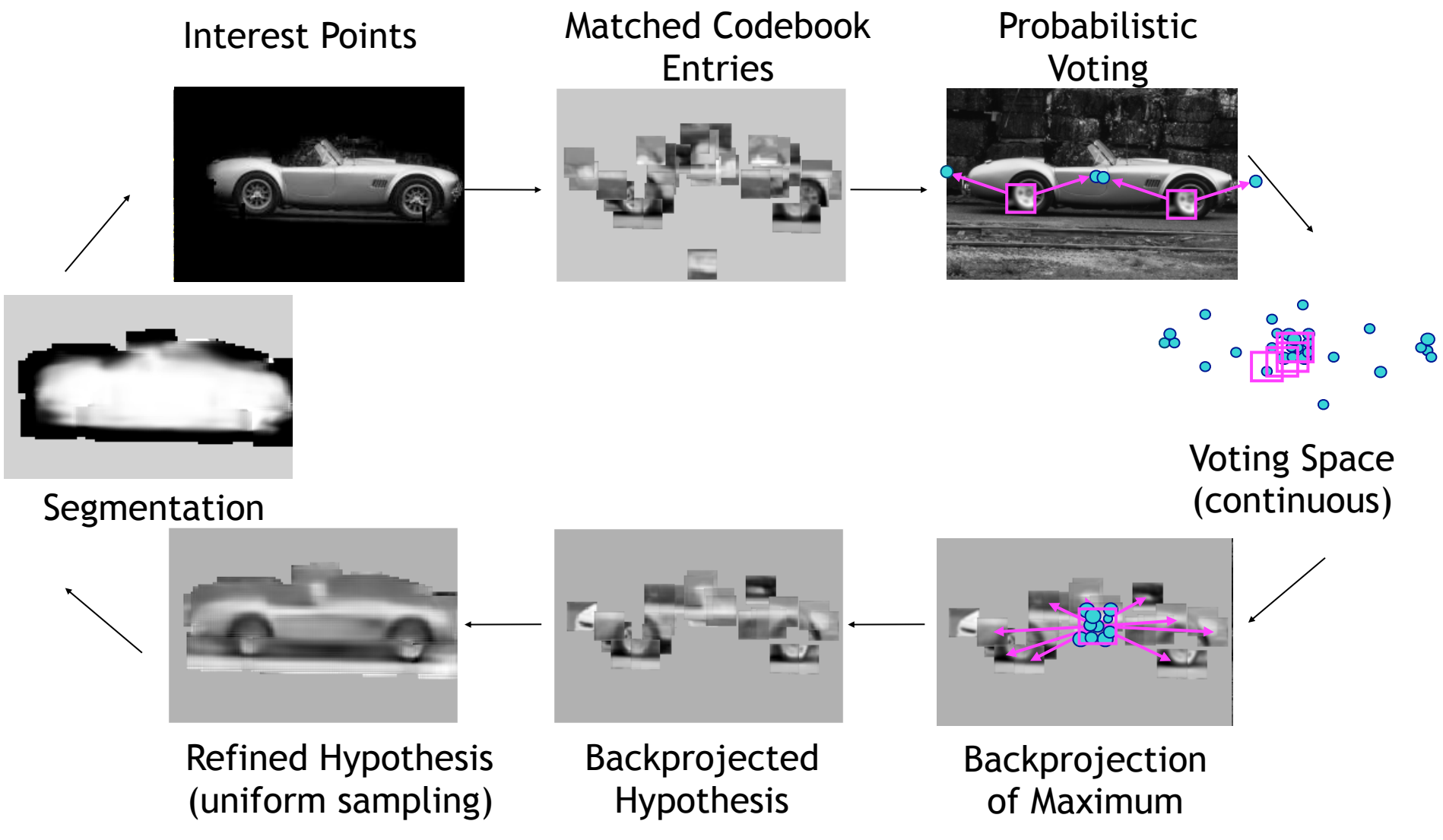
- Learn appearance codebook
  - Extract features at interest points (e.g. DoG)
  - Agglomerative clustering ⇒ codebook
- Learn codebook distributions (position & scale)
  - Match codebook to training images
  - Record matching positions on object




Spatial occurrence distributions

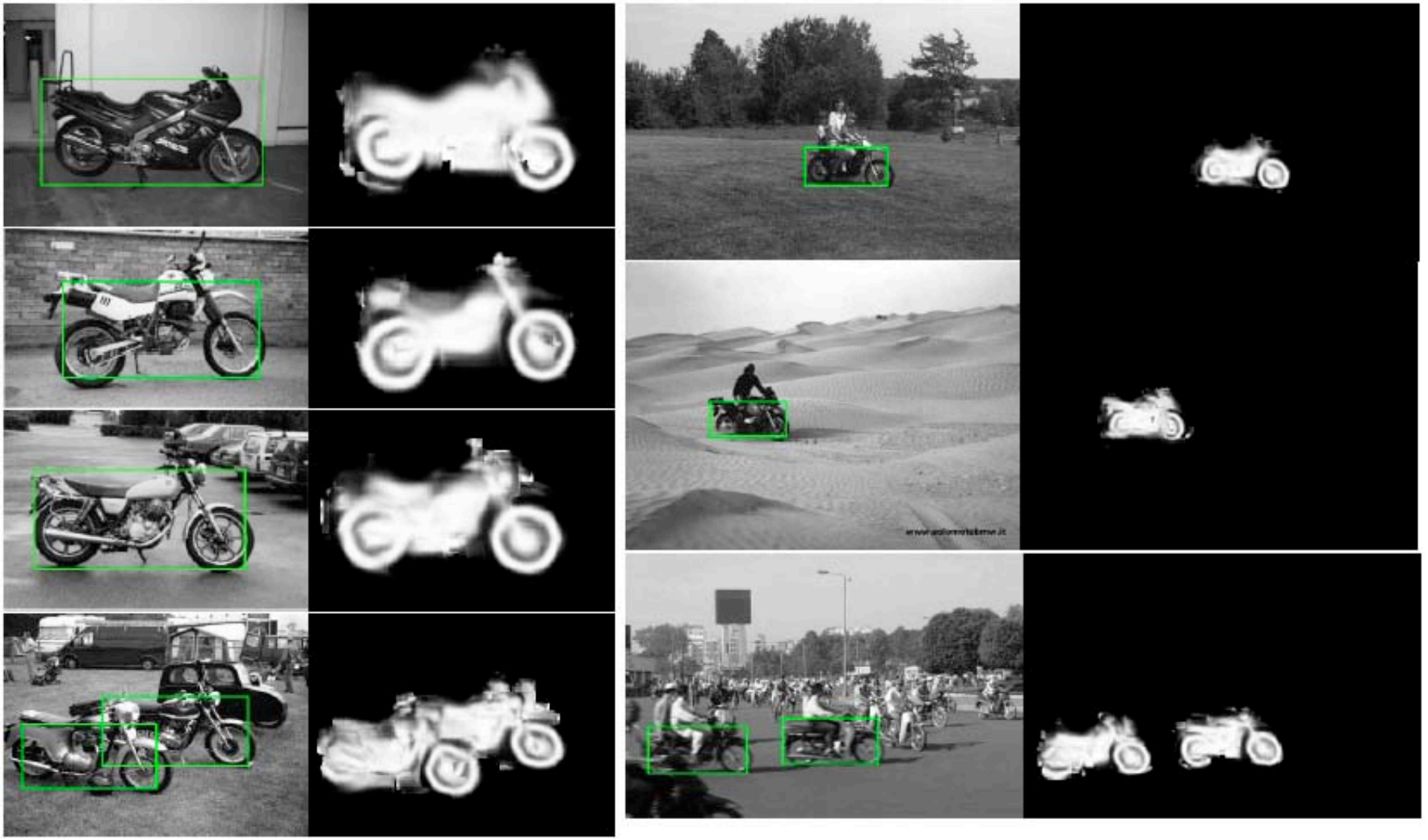
[Leibe,Schiele@bmvc03]

# Categorization: “Closing the Loop”



[Leibe,Schiele@bmvc03]

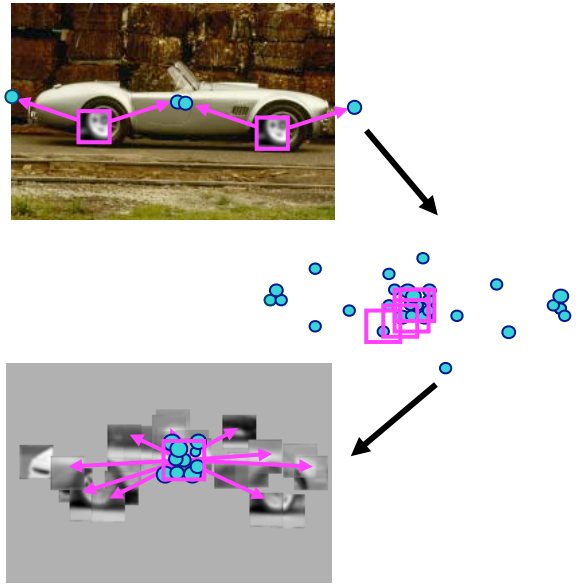
# Other Categories: Motorbikes



# Implicit Shape Model: What are Good Parts?

[Leibe,Schiele@bmvc03]
[Leibe,Leonardis,Schiele@ijcv08]

- Parts of the Implicit Shape Model
  - "parts" = feature clusters
  - lots of "parts" (in the order of 1'000 - 10'000 codebook entries) the more the better !
  - "parts" are mostly non-semantic



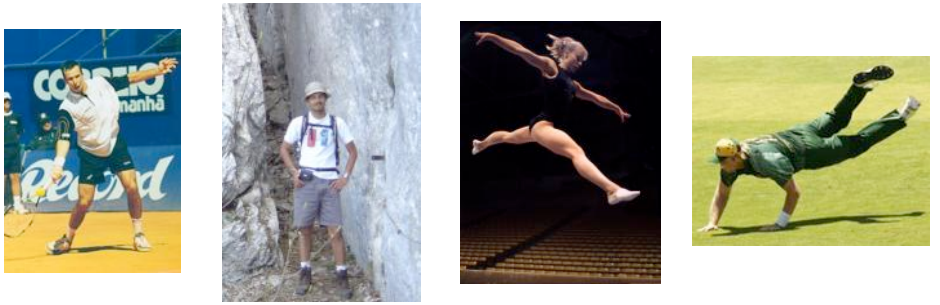
- "parts" = (mostly non semantic) feature clusters also true for
  - bag of words models
  - constellation model
  - ...

# Overview

- What are Ideal Parts for Part-Based Object models
- Part-Based Models for Object and People Detection
  - Implicit Shape Model [bmvc03,ijcv08]
  - Pictorial Structures Model for Articulated Pose Estimation [cvpr09]
  - Hierarchical Latent CRF Model for Objects [cvpr10]
  - Learning Shape Models from 3D CAD Data [bmvc10]
- Discussion
  - Semantic vs. Non-Semantic Object Parts

# Person Detection & Pose Estimation: Different Scenarios

1. Human Pose Estimation
   "People" dataset
   [Ramanan, NIPS'06]



2. Upper-body Pose Estimation
   "Buffy" dataset
   [Ferrari et al., CVPR'08]

3. Pedestrian Detection
   "TUD Pedestrians" dataset
   [Andriluka et al., CVPR'08]

- Typical approach: human body is represented as a flexible configuration of (**semantic**) body parts

[Andriluka,Roth,Schiele@cvpr09]
[Felzenszwalb,Huttenlocher@ijcv05]

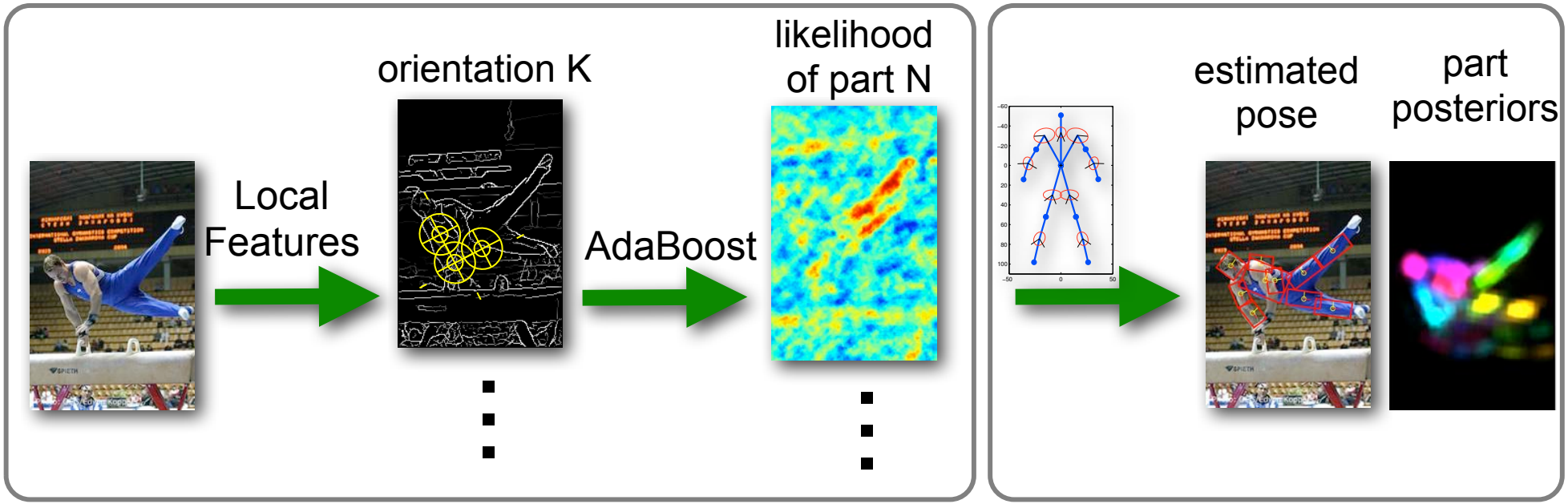
# Pictorial Structures Revisited:

Posterior over body poses

$$p(L|D) \propto p(D|L)p(L)$$

likelihood of part observations (appearance model)

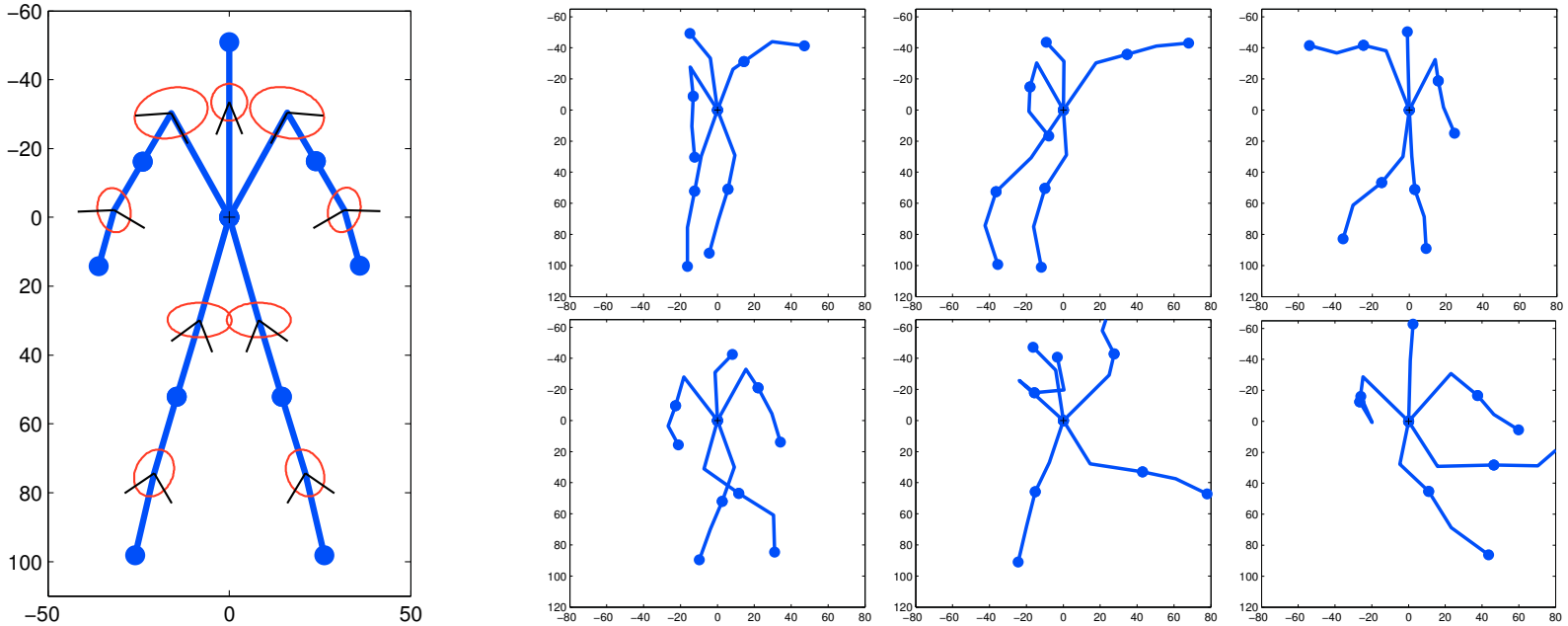
prior on body poses



[Andriluka,Roth,Schiele@cvpr09]

# Configuration of Body Parts: p(L)

- Kinematic Tree Prior
  - Samples from our generic prior on body poses:

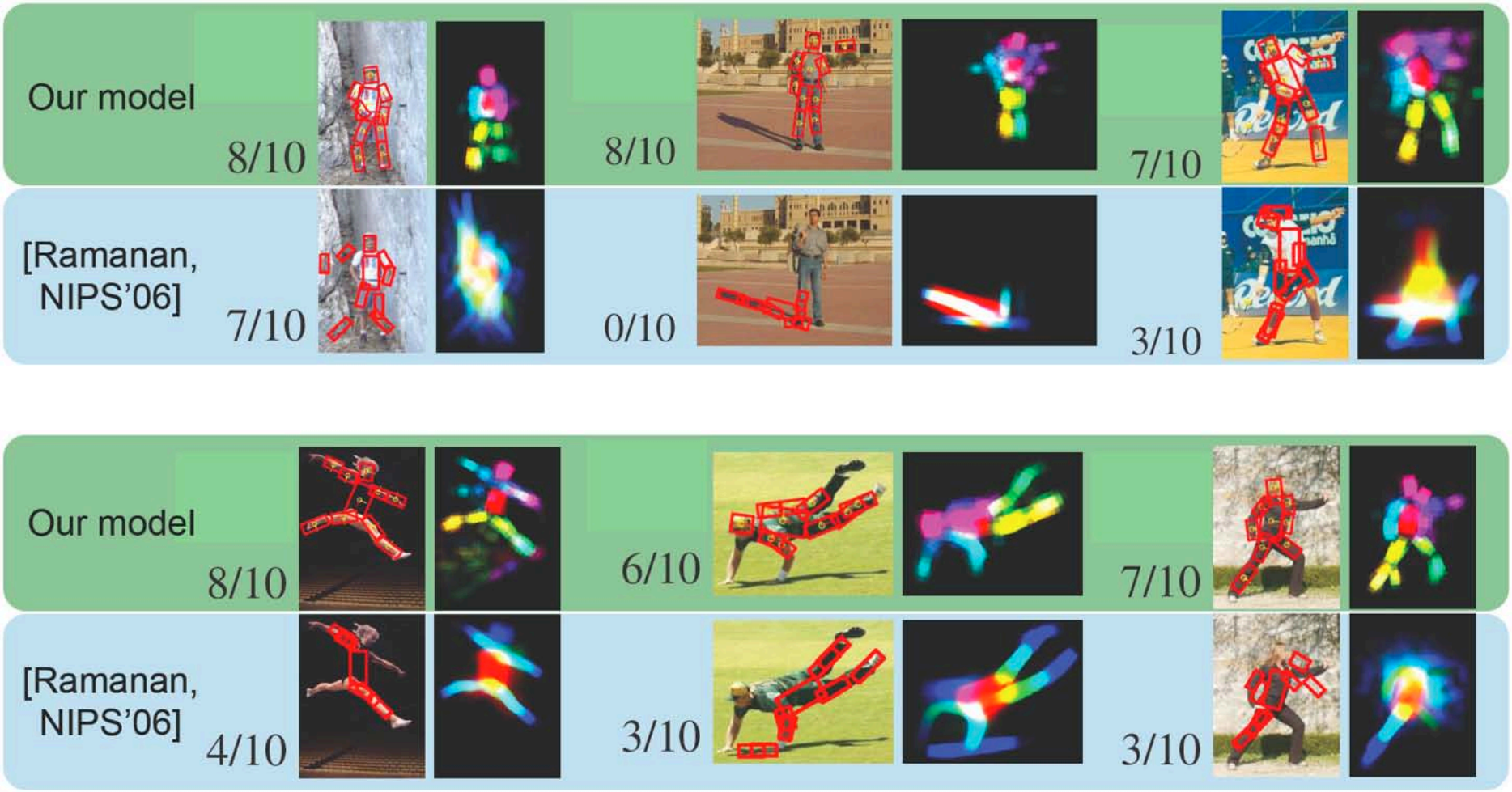


$$p(L|D) \propto p(D|L)p(L)$$

prior on body poses

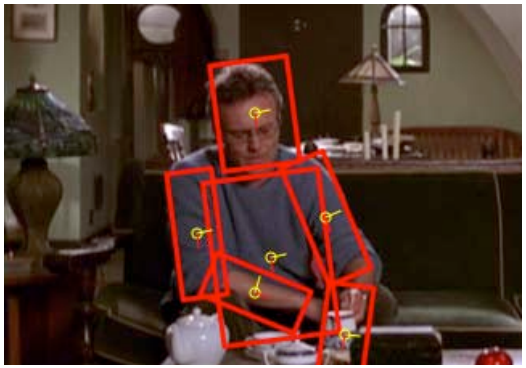
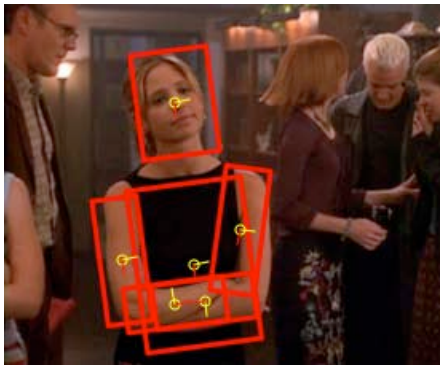
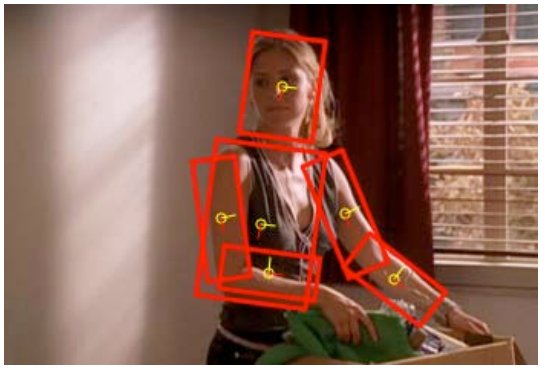
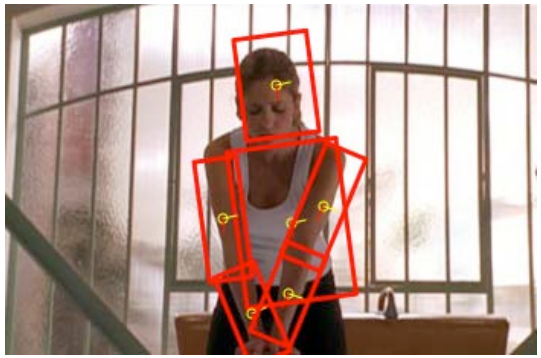
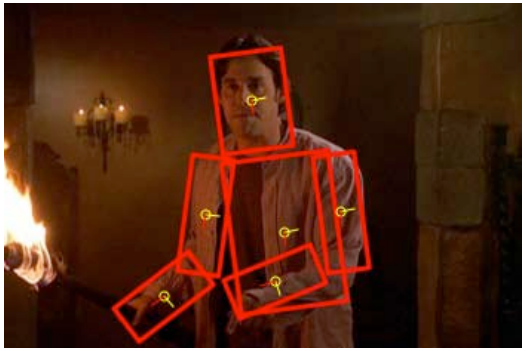
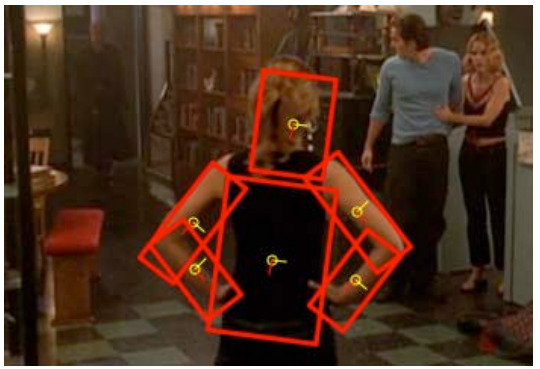
# Part-Based Model: 2D Human Pose Estimation

[Andriluka,Roth,Schiele@cvpr09]



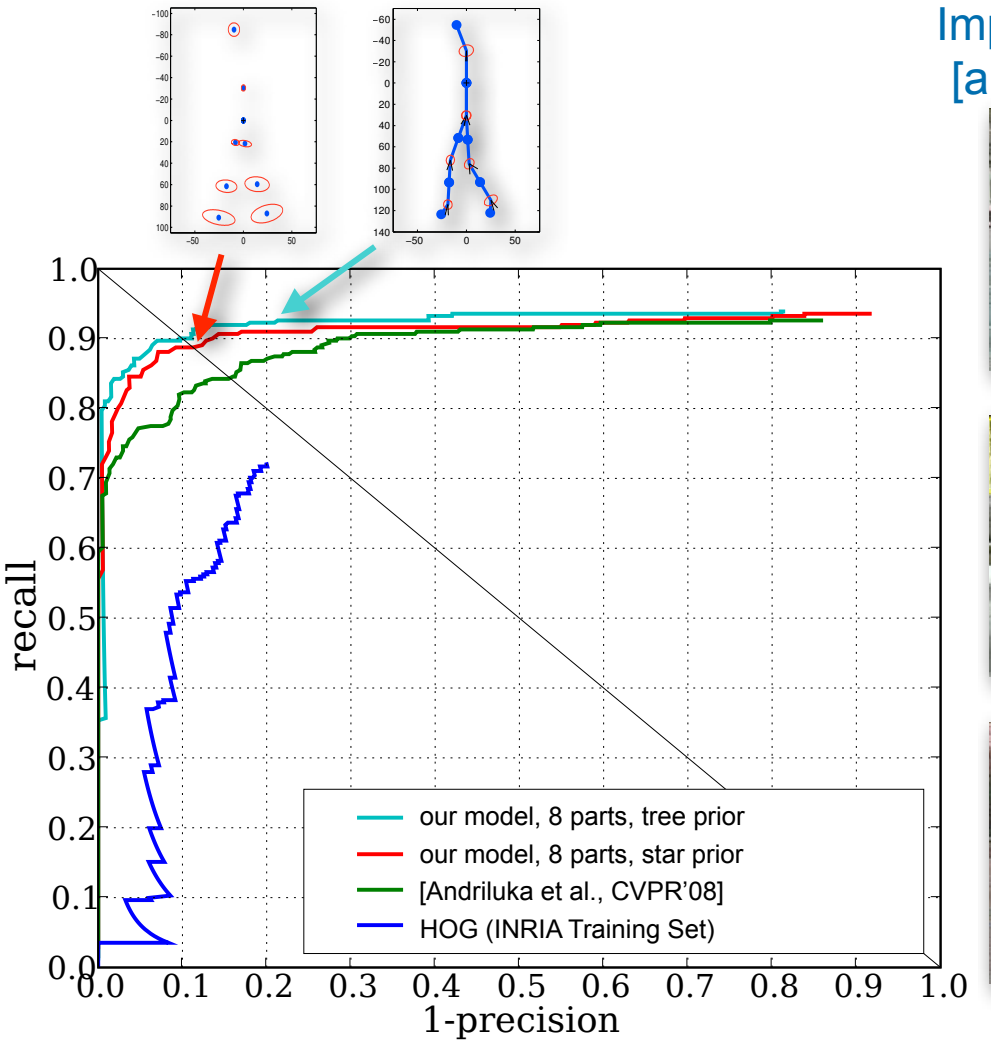
# Part-Based Model: Upper-Body 2D Pose Estimation

[Andriluka,Roth,Schiele@cvpr09]

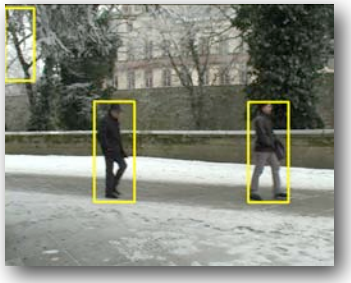







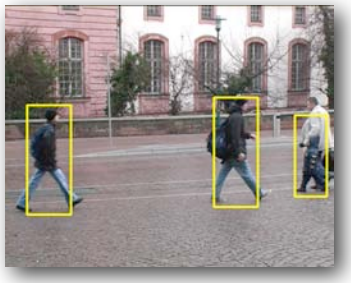
# Special Case: Upright People Detection

- Comparison of ISM and Pictorial Structures Model

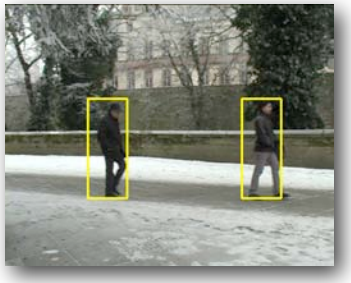


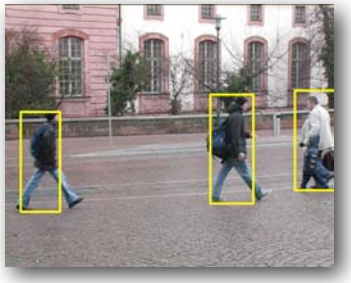

Implicit Shape Model: [andriluka@cvpr08]:





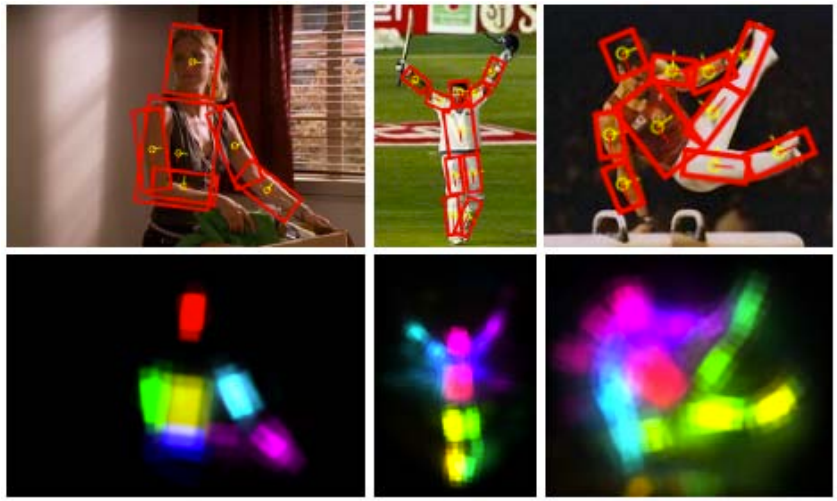
Pictorial Structures model [andriluka@cvpr09]





# Pictorial Structures for Human Pose Estimation: What are Good Parts?

- Parts of the Pictorial Structures Model
  - “parts” = semantic body parts
  - pose estimation = estimation of body part configuration
  - semantic body parts allow to use motion capture data, etc. to improve kinematic tree prior



  - non-semantic parts (e.g. in the ISM-model) are more difficult to generalize across human body poses

**sidenote: Jitendra will proof me wrong later today ;-)**

# Overview

- What are Ideal Parts for Part-Based Object models
- Part-Based Models for Object and People Detection
  - Implicit Shape Model [bmvc03,ijcv08]
  - Pictorial Structures Model for Articulated Pose Estimation [cvpr09]
  - Hierarchical Latent CRF Model for Objects [cvpr10]
  - Learning Shape Models from 3D CAD Data [bmvc10]
- Discussion
  - Semantic vs. Non-Semantic Object Parts

# “Standard Models” seen as Graphical Models

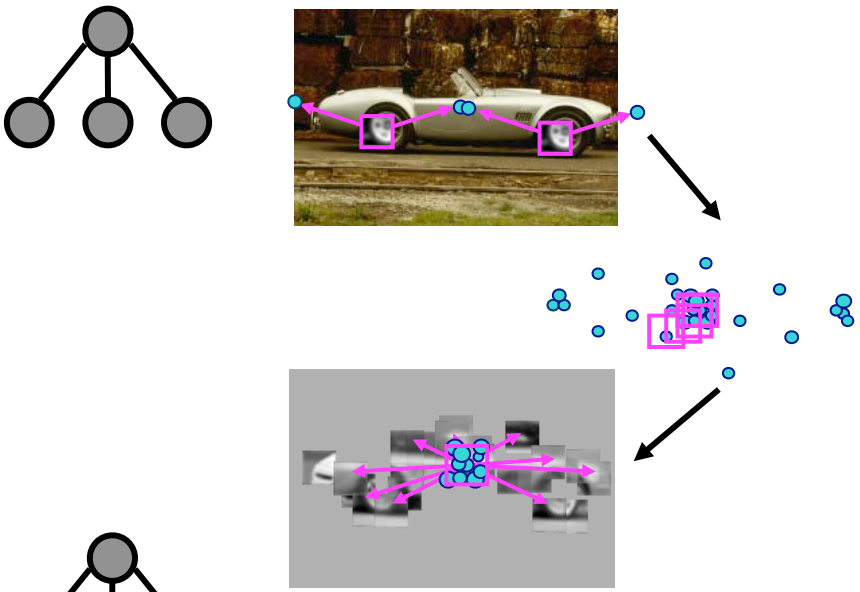
- Monolithic (global)

HOG [Dalal & Triggs, 2005]

also: Spatial Pyramid Kernel
[Lazebnik et al., 2006]

- Part-based (local)

Implicit Shape Model
[Leibe&Schiele, 2003]

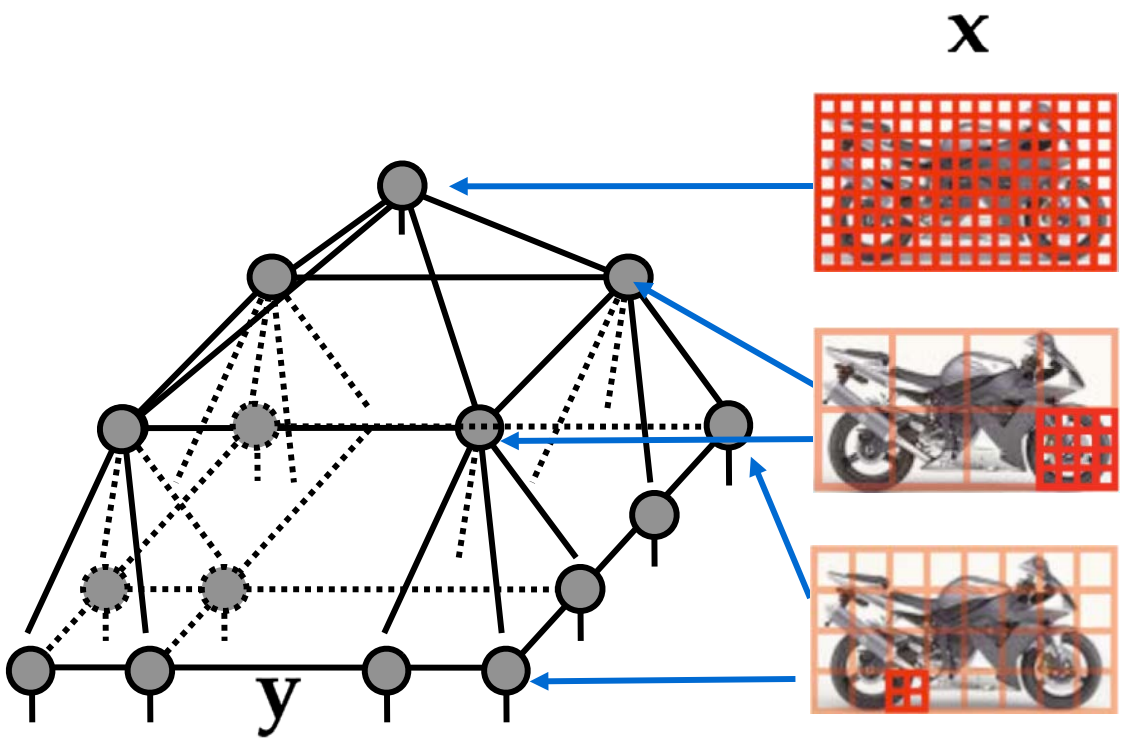


Pictorial Structures
[Felzenszwalb et al., 2005]

[Schnitzspan,Fritz,Schiele@eccv08]

# Hierarchical Graphical Model

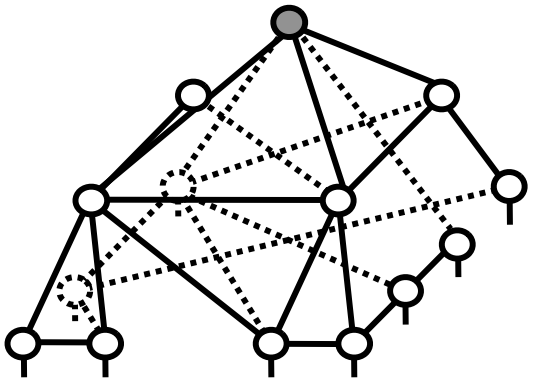
- Generalization of “Standard Models”
  - New hierarchical structure
  - Short and longer range dependencies
  - Joint training of local and global representations
    - Cyclic graph structure
    - Bottom-up top-down propagation



[Schnitzspan,Roth,Schiele@cvpr10]

# Hierarchical CRF Latent Model

- Goal: discover meaningful parts
- Labels of nodes not known
  - Now: part labels
  - Allow for ambiguities of part assignments
- Combine with structure learning
- Keep hierarchical representation



# Experiments on PASCAL VOC 2007: Hierarchical CRF Latent Model

[Schnitzspan,Roth,Schiele@cvpr10]

| VOC 2007 | aero | bicyc | bird | boat | bottle | bus | car | cat | chair | cow |
|---|---|---|---|---|---|---|---|---|---|---|
| Our model | 31.9 | **57.0** | 9.1 | 15.2 | 26.0 | 42.7 | 49.3 | 14.5 | 15.2 | 18.5 |
| HOG only | 29.1 | 56.4 | 4.6 | 13.0 | 25.2 | 40.7 | 47.3 | 13.5 | 10.1 | 18.8 |
| No parts | 31.7 | 56.3 | 1.7 | 15.1 | **27.6** | 41.3 | 48.0 | 15.2 | 9.5 | 18.3 |
| DPM | 28.1 | 55.4 | 1.4 | 14.5 | 25.4 | 38.9 | 46.6 | 14.3 | 9.4 | 16.0 |

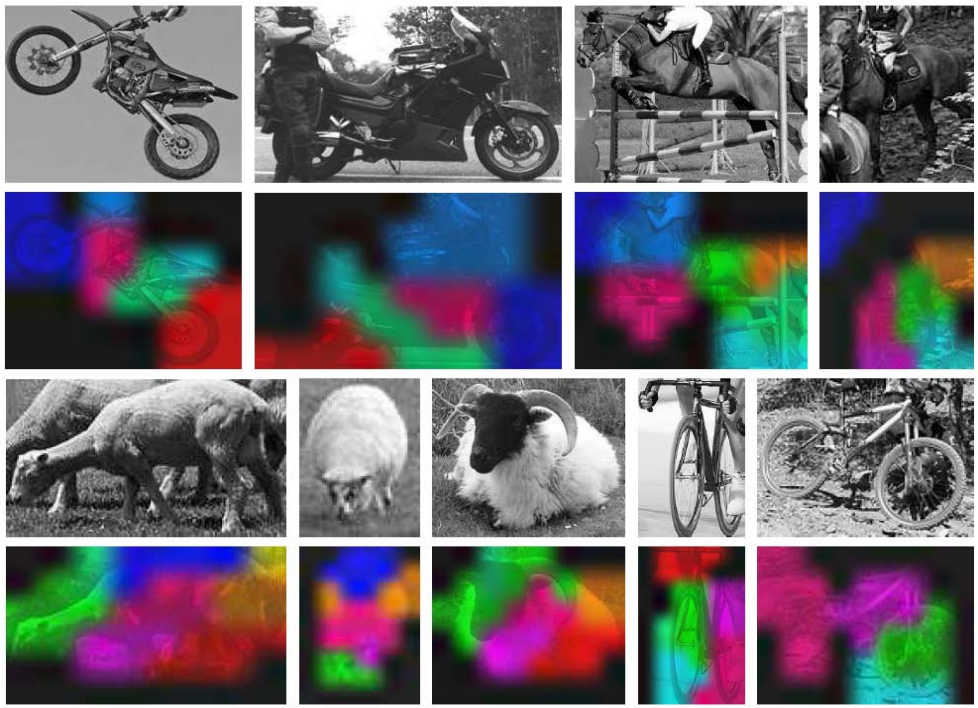
| | table | dog | horse | mbike | person | plant | sheep | sofa | train | tv | average |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Our model | 24.2 | 11.8 | 49.1 | 41.9 | 35.7 | 14.5 | 18.9 | 23.3 | 34.3 | 41.3 | 28.7 |
| HOG only | 23.1 | 10.9 | 48.0 | 38.4 | 34.7 | 14.3 | 17.1 | 21.0 | 32.7 | 38.8 | 26.9 |
| No parts | **26.1** | 11.3 | 48.5 | 38.9 | **35.8** | **14.8** | 17.7 | 18.8 | 34.1 | 39.8 | 27.5 |
| DPM | 22.8 | 10.6 | 44.1 | 37.0 | 35.2 | 13.6 | 16.1 | 18.5 | 31.8 | 36.9 | 25.9 |

- Consistently outperforms DPM [Felzenszwalb et al., 2008]
- MKL [Vedaldi et al., 2009] better due to more features

# Hierarchical Latent CRF-Model: What are Good Parts?

[Schnitzspan,Roth,Schiele@cvpr10]
similar observations hold for DPM
[Felzenszwalb et al@cvpr08]

- Parts of the Latent CRF Model:
  - "parts" have three main properties
    - "parts" = hierarchical appearance abstractions
    - "parts" = "local appearance" shared across objects
    - "parts" = discriminat "local appearance" of the objects
  - parts enable effective learning
    - by finding correspondences between discriminative local object appearance
  - semantics of parts is a secondary effect and not necessary



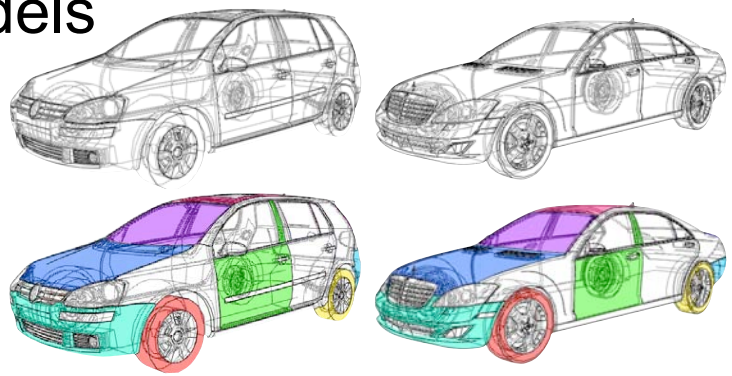
# Overview

- What are Ideal Parts for Part-Based Object models
- Part-Based Models for Object and People Detection
  - Implicit Shape Model [bmvc03,ijcv08]
  - Pictorial Structures Model for Articulated Pose Estimation [cvpr09]
  - Hierarchical Latent CRF Model for Objects [cvpr10]
  - Learning Shape Models from 3D CAD Data [bmvc10]
- Discussion
  - Semantic vs. Non-Semantic Object Parts

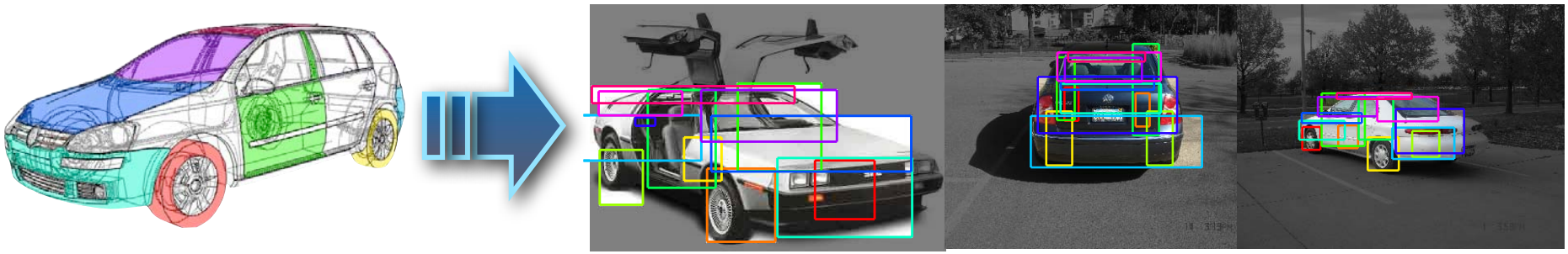
# Back to the Future: Learning Shape Models from 3D CAD Data

[Stark,Goesele,Schiele@bmvc10]

- 3D Computer Aided Design (CAD) Models
  - Computer graphics, game design
  - Polygonal meshes + texture descriptions
  - semantic part annotations (may) exist



- Can we learn Object Class Models directly from 3D CAD data?
  - Issue: Transition between 3D CAD models and 2D real-world images



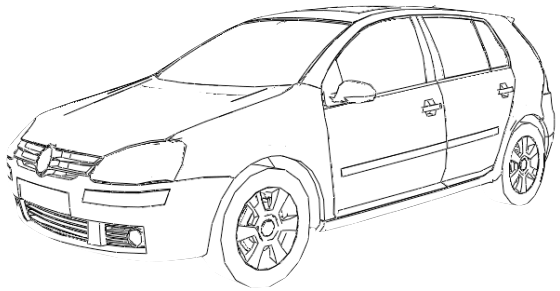
# Shape-based Appearance Abstraction
# ➳ Non-Photorealistic Rendering

- Learn shape models from rendered images
  - we do NOT render photo-realistically / texture
  - But focus on 3D CAD model edges (mimic real-world image edges)

**Part boundaries**

**Mesh creases**

**Silhouette**



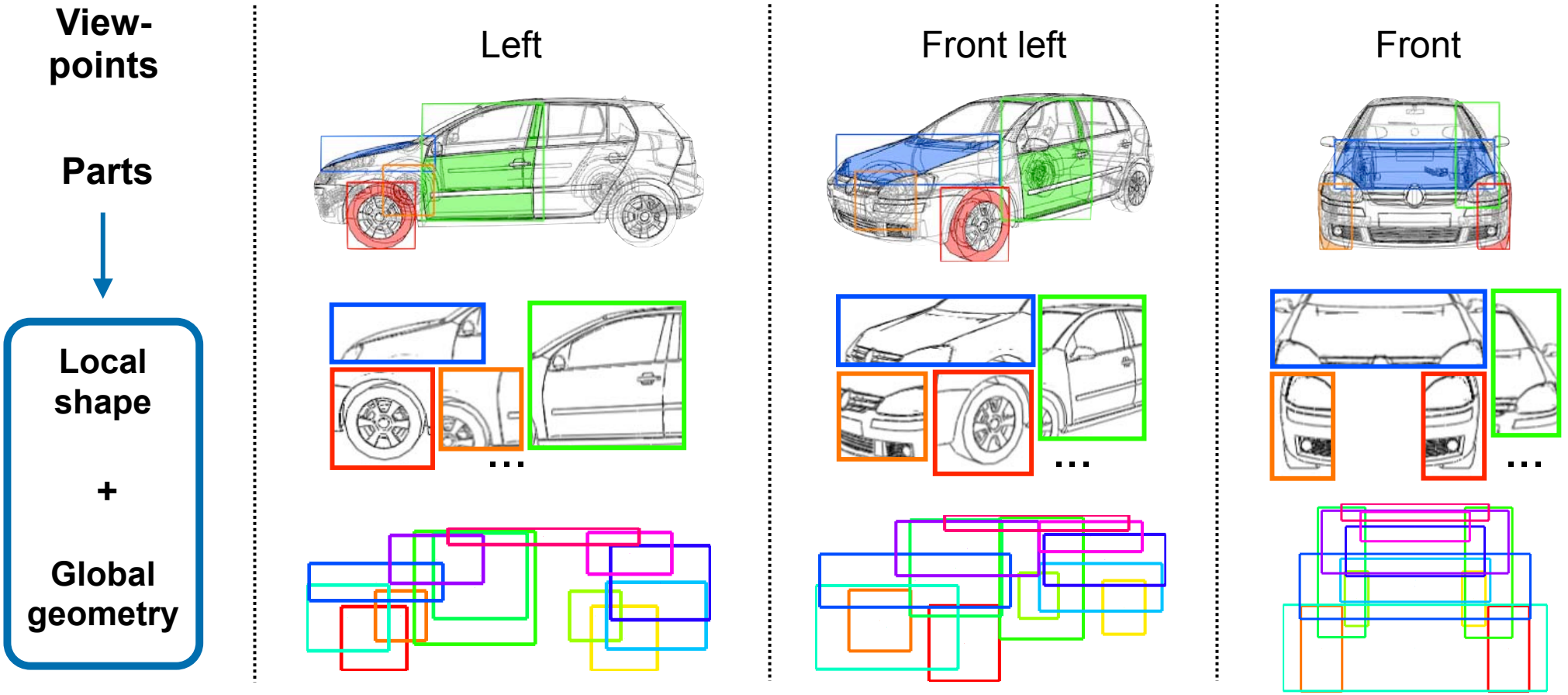
**Final edges**
(hidden edges removed)

[Stark,Goesele,Schiele@bmvc10]

[Stark,Goesele,Schiele@bmvc10]

# Shape - Local Shape + Global Geometry

- Part-based object class representation
  - Semantic parts from 3D CAD models: *left front wheel*, *left front door*, etc.



# Experimental Evaluation - Test Data Set

- 3D Object Classes *Cars* [Savarese and Fei-Fei ICCV'07]

8 azimuth angles


back


back-left


left


front-left

...

2 elevation angles


low


high

5 cars





3 distances


near

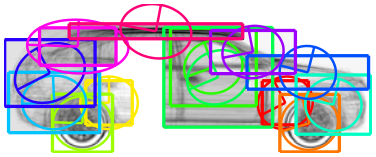

medium


far

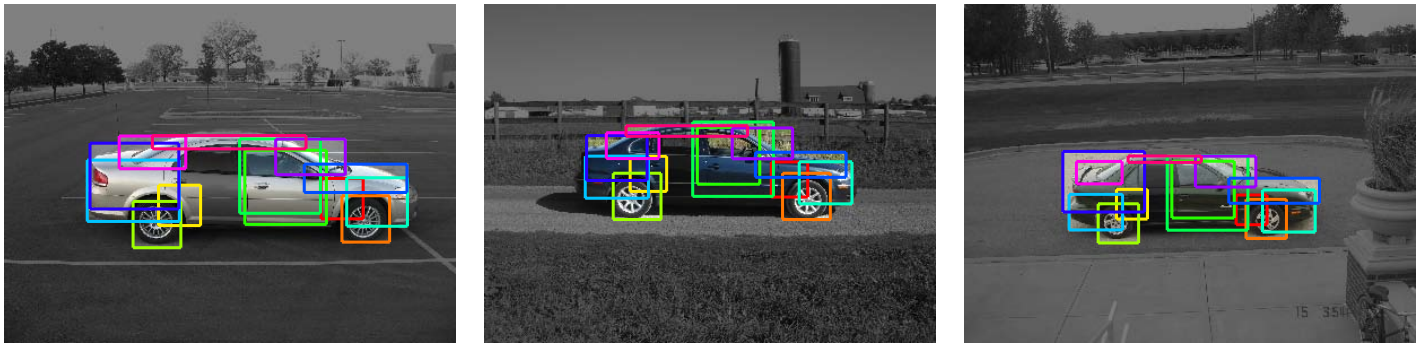
**240 images**

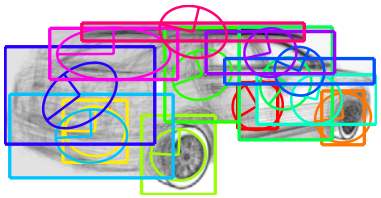
# Qualitative Results

- Three strongest true positive detections per viewpoint model

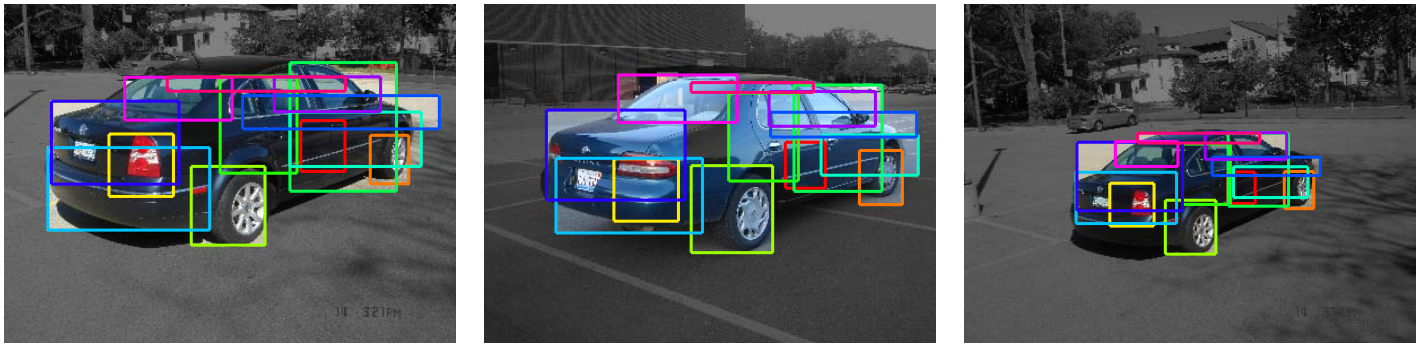


Right





Back-right

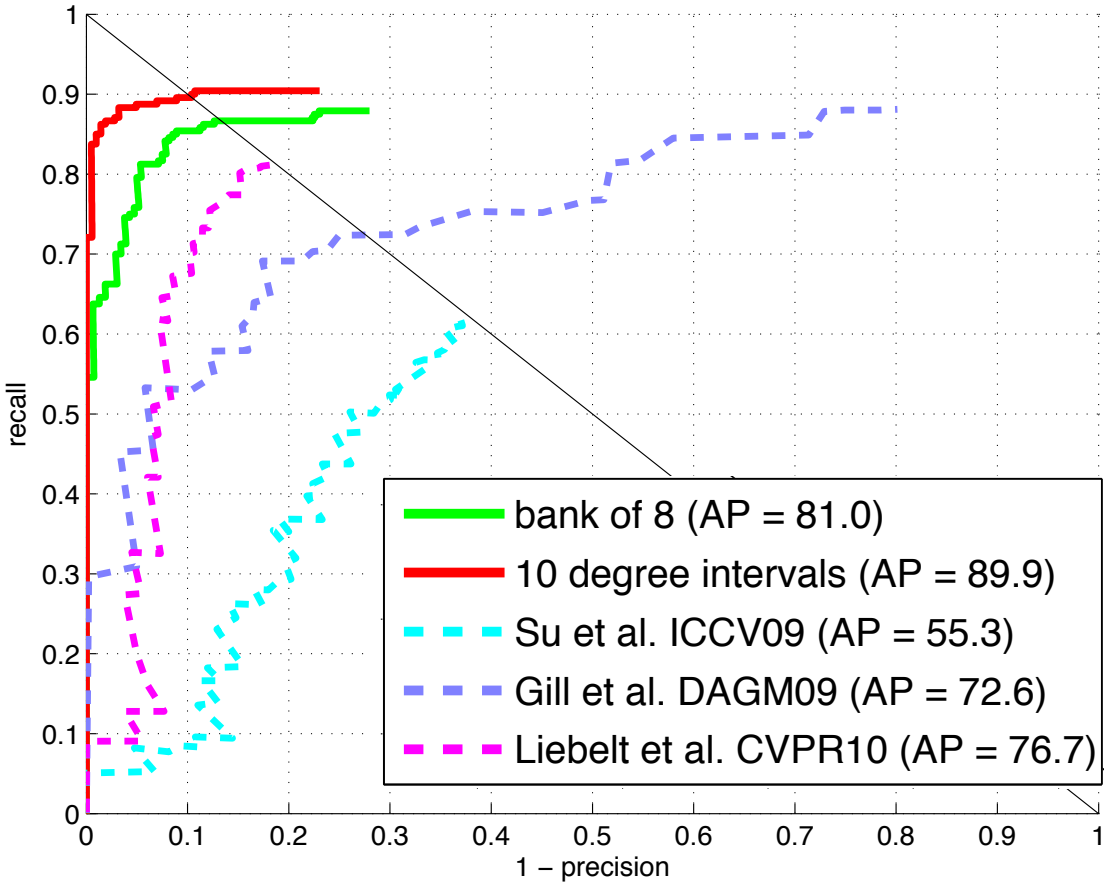


- Observations
  - Accurate part localization

# Quantitative Results - Multi-View Detection

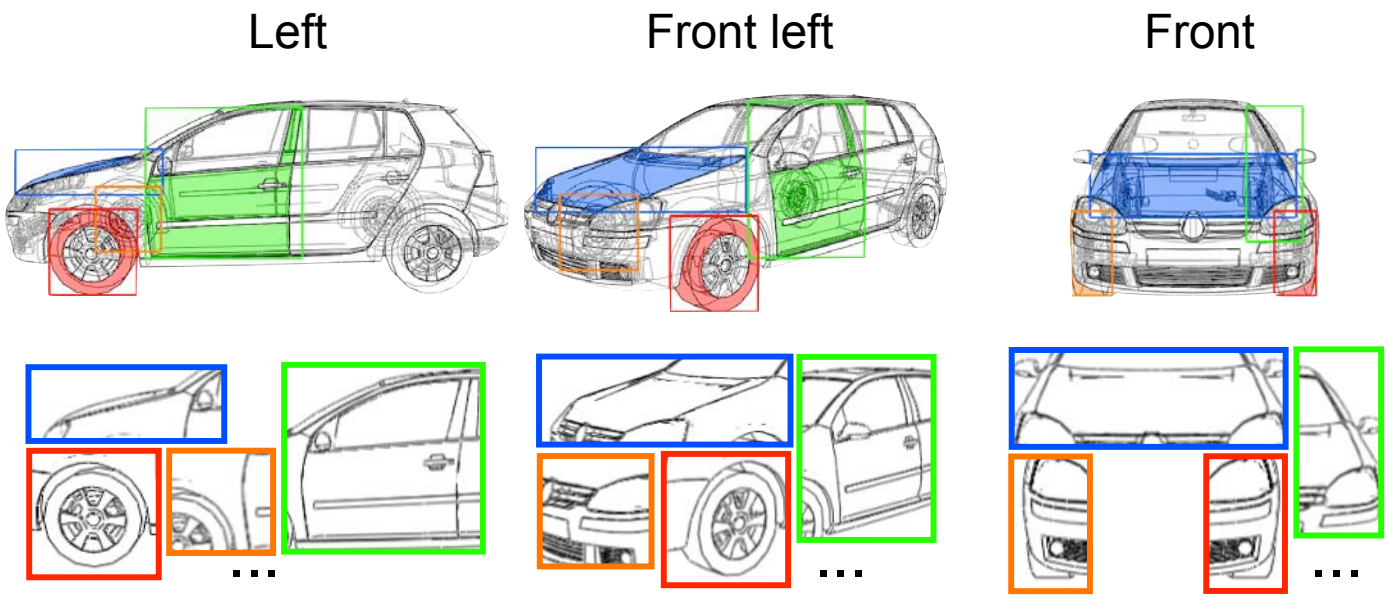
- Evaluation protocol
  - Precision / recall
  - PASCAL overlap criterion

- Observations
  - 4.3% AP improvement (■) [Liebelt et al. CVPR'10] (■)
  - We can further improve performance by 8.9% AP (■)



# Shape Model learned from 3D CAD-Data
# What are Good Parts?

- Parts of the Shape Model:
  - “parts” = just means to enable correspondence across 3D-models
  - semantics of parts:
    - in our case: yes - because of the employed 3D models
    - but: semantics neither necessary nor important

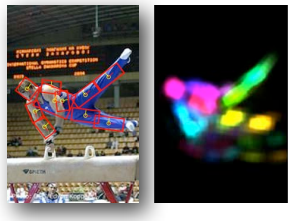


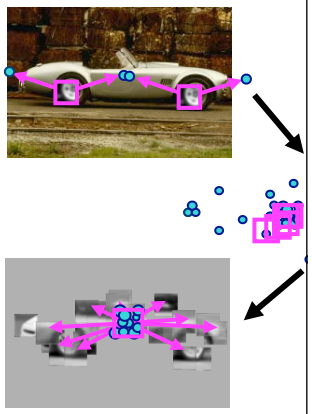
# Overview

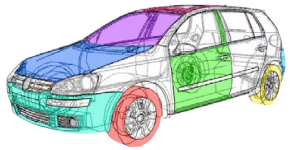
- What are Ideal Parts for Part-Based Object models
- Part-Based Models for Object and People Detection
  - Implicit Shape Model [bmvc03,ijcv08]
  - Pictorial Structures Model for Articulated Pose Estimation [cvpr09]
  - Hierarchical Latent CRF Model for Objects [cvpr10]
  - Learning Shape Models from 3D CAD Data [bmvc10]
- Discussion
  - Semantic vs. Non-Semantic Object Parts
  - What are the Ideal Parts for Parts-Based Object Models

# What are Ideal Parts for Part-Based Object Models?

- Parts can/may
  - be semantic body parts - e.g. for articulated human body pose estimation
  - be feature clusters (typically many clusters) (e,g, ISM, constellation model, BoW)
  - support learnability of discriminant appearance
  - enable correspondence across 3D models
  - ...



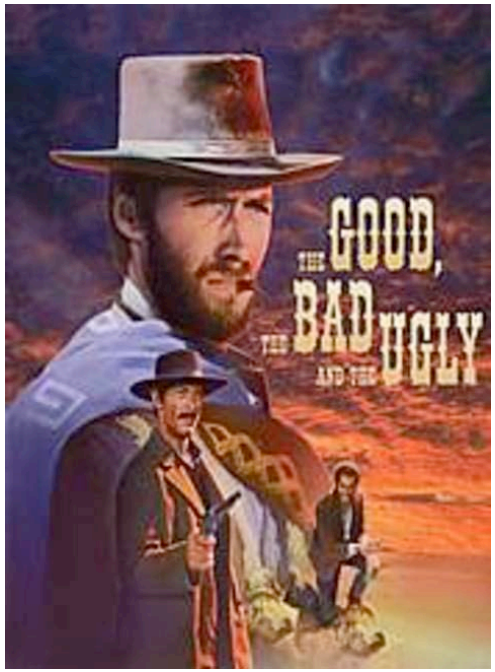




- in all those cases: the most important property is that "parts" facilitate correspondence across object instances

# What are Ideal Parts for Part-Based Object Models?

- Multiple motivations for part/attribute based models exist:
  - intuitiveness: semantic meaning of parts/attributes is attractive (e.g. enables use of language sources)
  - learnability: sharing of parts/attributes across instances/classes
  - scalability: transferability of parts/attributes across classes
  - ...
- in general, parts (and attributes) support learnability and scalability when they facilitate correspondence
  - across object instances
  - across object classes
  - across modalities (e.g. from language to visual appearance)
  - and semantics is only a secondary concern (for "intuitiveness")

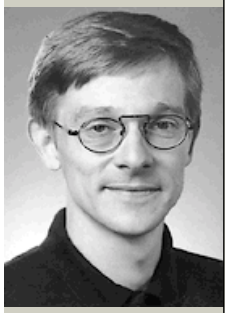



# What are Ideal Parts for Part-Based Object Models?

# the Good, the Bad, and the Ugly

**Bernt Schiele**

**Max Planck Institute for Informatics**



**thanks to: Micha Andriluka, Bastian Leibe, Sandra Ebert, Mario Fritz, Diane Larlus, Marcus Rohrbach, Paul Schnitzspan, Stefan Roth, Michael Stark, Michael Goesele**