

Computer Vision - Lecture 14

Part-based Models for Object Categorization

14.12.2016

Bastian Leibe
 RWTH Aachen
<http://www.vision.rwth-aachen.de>
 leibe@vision.rwth-aachen.de

Course Outline

- Image Processing Basics
- Segmentation & Grouping
- Object Recognition
 - Object Categorization I
 - Sliding Window based Object Detection
 - Local Features & Matching
 - Local Features - Detection and Description
 - Recognition with Local Features
 - Indexing & Visual Vocabularies
 - Object Categorization II
 - Bag-of-Words Approaches & Part-based Approaches
 - Deep Learning Methods
- 3D Reconstruction

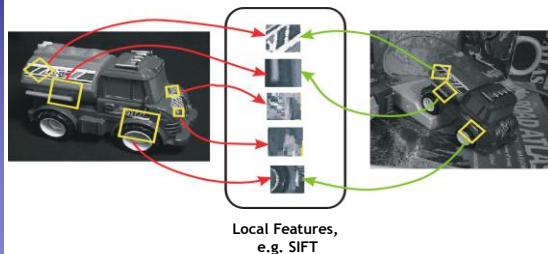
Topics of This Lecture

- Recap: Specific Object Recognition with Local Features
 - Matching & Indexing
 - Geometric Verification
- Part-Based Models for Object Categorization
 - Structure representations
 - Different connectivity structures
- Bag-of-Words Model
 - Use for image classification
- Implicit Shape Model
 - Generalized Hough Transform for object category detection
- Deformable Part-based Model
 - Discriminative part-based detection

B. Leibe

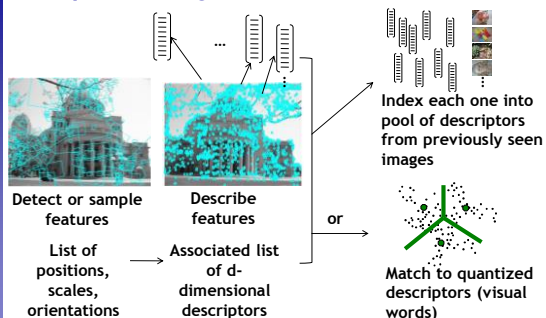
Recap: Recognition with Local Features

- Image content is transformed into local features that are invariant to translation, rotation, and scale
- Goal: Verify if they belong to a consistent configuration



B. Leibe

Recap: Indexing features



⇒ Shortlist of possibly matching images + feature correspondences

B. Leibe

Extension: *tf-idf* Weighting

- Term frequency - inverse document frequency
 - Describe frame by frequency of each word within it, downweight words that appear often in the database
 - (Standard weighting for text retrieval)

$$t_i = \frac{n_{id}}{n_d} \log \frac{N}{n_i}$$

Number of occurrences of word i in document d → n_{id}

Number of words in document d → n_d

Total number of documents in database → N

Number of occurrences of word i in whole database → n_i

B. Leibe

RWTH AACHEN UNIVERSITY

Recap: Fast Indexing with Vocabulary Trees

- Recognition

Geometric verification

[Nister & Stewenius, CVPR'06]

8

Computer Vision WS 15/16

Slide credit: David Nister

B. Leibe

RWTH AACHEN UNIVERSITY

Recap: Geometric Verification by Alignment

- Assumption
 - Known object, rigid transformation compared to model image
 - ⇒ If we can find evidence for such a transformation, we have recognized the object.
- You learned methods for
 - Fitting an *affine transformation* from ≥ 3 correspondences
 - Fitting a *homography* from ≥ 4 correspondences

Affine: solve a system Homography: solve a system

$$At = b \qquad Ah = 0$$

- Correspondences may be noisy and may contain outliers
 - ⇒ Need to use robust methods that can filter out outliers
 - ⇒ Use **RANSAC** or the **Generalized Hough Transform**

9

Computer Vision WS 15/16

B. Leibe

RWTH AACHEN UNIVERSITY

Topics of This Lecture

- Recap: Specific Object Recognition with Local Features
- Part-Based Models for Object Categorization
 - Structure representations
 - Different connectivity structures
- Bag-of-Words Model
 - Use for image classification
- Implicit Shape Model
 - Generalized Hough Transform for object category detection
- Deformable Part-based Model
 - Discriminative part-based detection

10

Computer Vision WS 15/16

B. Leibe

RWTH AACHEN UNIVERSITY

Recognition of Object Categories

- We no longer have exact correspondences...
- On a local level, we can still detect similar parts.
- Represent objects by their parts
 - ⇒ Bag-of-features
- How can we improve on this?
 - Encode structure

11

Computer Vision WS 15/16

Slide credit: Rob Fergus

RWTH AACHEN UNIVERSITY

Part-Based Models

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

12

Computer Vision WS 15/16

B. Leibe

RWTH AACHEN UNIVERSITY

Different Connectivity Structures

$\mathcal{O}(N)$

a) Bag of visual words
Csurka et al. '04
Vasconcelos et al. '00

$\mathcal{O}(N^2)$

b) Constellation
Fergus et al. '03
Fel-Fei et al. '03

$\mathcal{O}(N^2)$

c) Star shape
Leibe et al. '04, '08
Crandall et al. '05
Fergus et al. '05

$\mathcal{O}(N^2)$

d) Tree
Felzenszwalb & Huttenlocher '05

$\mathcal{O}(N^3)$

e) k-fan (k = 2)
Crandall et al. '05

Center {
Part {
Subpart {

f) Hierarchy
Bouchard & Triggs '05

$k=1$
 $k=2$

g) Sparse flexible model
Carneiro & Lowe '06

13

Computer Vision WS 15/16

Slide adapted from Rob Fergus

B. Leibe

Image from [Carneiro & Lowe, ECCV'06]

RWTH AACHEN UNIVERSITY

Topics of This Lecture

- Recap: Specific Object Recognition with Local Features
- Part-Based Models for Object Categorization
 - Structure representations
 - Different connectivity structures
- Bag-of-Words Model
 - Use for image classification
- Implicit Shape Model
 - Generalized Hough Transform for object category detection
- Deformable Part-based Model
 - Discriminative part-based detection

14

RWTH AACHEN UNIVERSITY

Recap: Analogy to Documents

Of all the sensory impressions proceeding to the brain, the visual experiences are the dominant ones. Our perception of the world around us is based essentially on the messages that come from our eyes. For this reason, we might think that the retina is the point by which the brain receives its information from the screen. In the discovery of the visual system, we know that the retina undergoes a step-wise analysis of the messages that fall on it. In this system each cell has its specific function and is responsible for a specific detail in the pattern of the retinal image.

sensory, brain, visual, perception, retinal, cerebral cortex, eye, cell, optical nerve, image Hubel, Wiesel

China is forecasting a trade surplus of \$90bn (£51bn) to \$100bn this year, a threefold increase on 2004's \$32bn. The Commerce Ministry said the surplus would be created by a 30% jump in exports to \$100bn. The 18% rise in exports is a 18% rise in exports. China has long had a trade surplus, but it has long been unfair to the rest of the world. Under the current trade agreement, the surplus is only on the order of \$30bn. Zhou Xiaochuan, China's vice premier, said he needed to demand so much more from the country. China has long had a trade surplus against the dollar by 2.1% and permitted it to trade within a narrow band, but the US wants the yuan to be allowed to trade freely. However, Beijing has made it clear that it will take its time and tread carefully before allowing the yuan to rise further in value.

China, trade, surplus, commerce, exports, imports, US, yuan, bank, domestic, foreign, increase, trade, value

15

RWTH AACHEN UNIVERSITY

Recap: Visual Words

- Quantize the feature space into "visual words"
- Perform matching only to those visual words.

Exact feature matching → Match to same visual word

16

RWTH AACHEN UNIVERSITY

Recap: Bag-of-Word Representations (BoW)

Object → Bag of "words"

17

RWTH AACHEN UNIVERSITY

Recap: Categorization with Bags-of-Words

- Compute the word activation histogram for each image.
- Let each such BoW histogram be a feature vector.
- Use images from each class to train a classifier (e.g., an SVM).

18

RWTH AACHEN UNIVERSITY

Recap: Advantage of BoW Histograms

- Bag of words representations make it possible to describe the unordered point set with a single vector (of fixed dimension across image examples).
- Provides easy way to use distribution of feature types with various learning algorithms requiring vector input.

19

Computer Vision WS 15/16

Limitations of BoW Representations

- The bag of words removes spatial layout.
- This is both a strength and a weakness.
- Why a strength?
- Why a weakness?

Slide adapted from Bill Freeman B. Leibe 20

Computer Vision WS 15/16

Spatial Pyramid Representation

- Representation in-between orderless BoW and global appearance

Slide credit: Svetlana Lazebnik B. Leibe (Lazebnik, Schmid & Ponce, CVPR'06) 21

Computer Vision WS 15/16

Spatial Pyramid Representation

- Representation in-between orderless BoW and global appearance

Slide credit: Svetlana Lazebnik B. Leibe (Lazebnik, Schmid & Ponce, CVPR'06) 22

Computer Vision WS 15/16

Spatial Pyramid Representation

- Representation in-between orderless BoW and global appearance

Slide credit: Svetlana Lazebnik B. Leibe (Lazebnik, Schmid & Ponce, CVPR'06) 23

Computer Vision WS 15/16

Summary: Bag-of-Words

- Pros:**
 - Flexible to geometry / deformations / viewpoint
 - Compact summary of image content
 - Provides vector representation for sets
 - Empirically good recognition results in practice
- Cons:**
 - Basic model ignores geometry - must verify afterwards, or encode via features.
 - Background and foreground mixed when bag covers whole image
 - When using interest points or sampling: no guarantee to capture object-level parts \Rightarrow Dense sampling is often better.

Slide credit: Kristen Grauman B. Leibe 24

Computer Vision WS 15/16

Topics of This Lecture

- Recap: Specific Object Recognition with Local Features
- Part-Based Models for Object Categorization
 - Structure representations
 - Different connectivity structures
- Bag-of-Words Model
 - Use for image classification
- Implicit Shape Model
 - Generalized Hough Transform for object category detection
- Deformable Part-based Model
 - Discriminative part-based detection


B. Leibe 25

Computer Vision WS 15/16

RWTH AACHEN UNIVERSITY

Implicit Shape Model (ISM)

- Basic ideas
 - Learn an appearance codebook
 - Learn a star-topology structural model
 - Features are considered independent given obj. center
- Algorithm: probabilistic Gen. Hough Transform
 - Exact correspondences → Prob. match to object part
 - NN matching → Soft matching
 - Feature location on obj. → Part location distribution
 - Uniform votes → Probabilistic vote weighting
 - Quantized Hough array → Continuous Hough space



B. Leibe

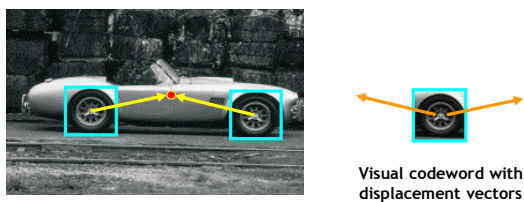
26

Computer Vision WS 15/16

RWTH AACHEN UNIVERSITY

Implicit Shape Model: Basic Idea

- Visual vocabulary is used to index votes for object position [a visual word = "part"].



Training image

Visual codeword with displacement vectors

B. Leibe, A. Leonardis, and B. Schiele, [Robust Object Detection with Interleaved Categorization and Segmentation](#), International Journal of Computer Vision, Vol. 77(1-3), 2008.


B. Leibe

Computer Vision WS 15/16

RWTH AACHEN UNIVERSITY

Implicit Shape Model: Basic Idea

- Objects are detected as consistent configurations of the observed parts (visual words).



Test image

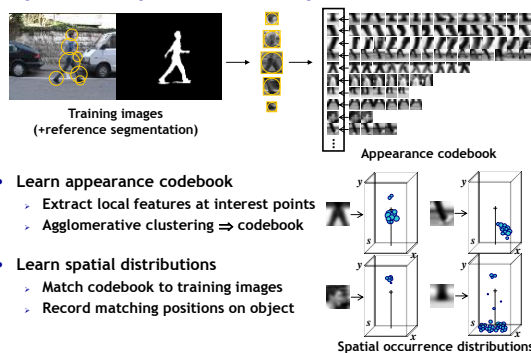
B. Leibe, A. Leonardis, and B. Schiele, [Robust Object Detection with Interleaved Categorization and Segmentation](#), International Journal of Computer Vision, Vol. 77(1-3), 2008.

B. Leibe

Computer Vision WS 15/16

RWTH AACHEN UNIVERSITY

Implicit Shape Model - Representation



- Learn appearance codebook
 - Extract local features at interest points
 - Agglomerative clustering ⇒ codebook
- Learn spatial distributions
 - Match codebook to training images
 - Record matching positions on object

Spatial occurrence distributions

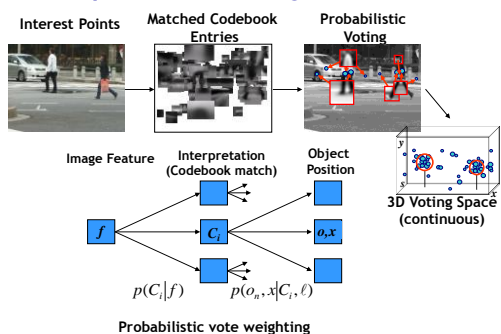
B. Leibe

29

Computer Vision WS 15/16

RWTH AACHEN UNIVERSITY

Implicit Shape Model - Recognition



Interest Points

Matched Codebook Entries

Probabilistic Voting

Image Feature

Interpretation (Codebook match)

Object Position

3D Voting Space (continuous)

Probabilistic vote weighting

$p(C_i|f)$

$p(a_i x|C_i, l)$

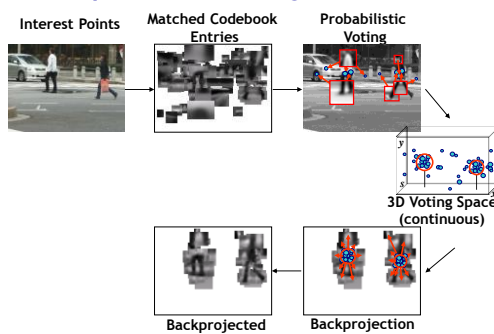
B. Leibe, Leonardis, Schiele, SLCV'04; IJCV'08

30

Computer Vision WS 15/16

RWTH AACHEN UNIVERSITY

Implicit Shape Model - Recognition



Interest Points

Matched Codebook Entries

Probabilistic Voting

3D Voting Space (continuous)

Backprojected Hypotheses

Backprojection of Maxima


B. Leibe, Leonardis, Schiele, SLCV'04; IJCV'08

31

Computer Vision WS 15/16

RWTH AACHEN UNIVERSITY

Example: Results on Cows



Original image


B. Leibe

32

Computer Vision WS 15/16

RWTH AACHEN UNIVERSITY

Example: Results on Cows



Interest points

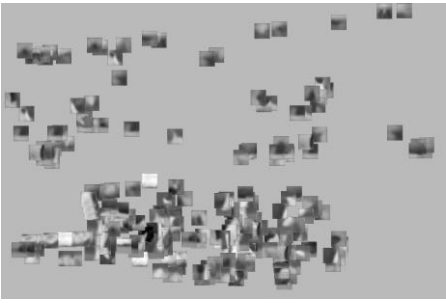
B. Leibe

33

Computer Vision WS 15/16

RWTH AACHEN UNIVERSITY

Example: Results on Cows



Matched patches

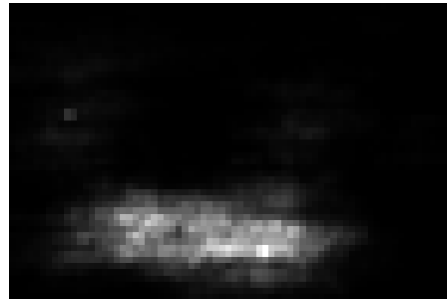
B. Leibe

34

Computer Vision WS 15/16

RWTH AACHEN UNIVERSITY

Example: Results on Cows



Prob. Votes


B. Leibe

35

Computer Vision WS 15/16

RWTH AACHEN UNIVERSITY

Example: Results on Cows



1st hypothesis


K. Grauman, B. Leibe

36

Computer Vision WS 15/16

RWTH AACHEN UNIVERSITY

Example: Results on Cows



2nd hypothesis

B. Leibe

37

Computer Vision WS 15/16

RWTH AACHEN UNIVERSITY

Example: Results on Cows

3rd hypothesis

B. Leibe

38

Computer Vision WS 15/16

RWTH AACHEN UNIVERSITY

Scale Invariant Voting

- Scale-invariant feature selection
 - Scale-invariant interest regions
 - Extract scale-invariant descriptors
 - Match to appearance codebook
- Generate scale votes
 - Scale as 3rd dimension in voting space

$$x_{vote} = x_{img} - x_{occ}(s_{img}/s_{occ})$$

$$y_{vote} = y_{img} - y_{occ}(s_{img}/s_{occ})$$

$$s_{vote} = (s_{img}/s_{occ})$$

- Search for maxima in 3D voting space

B. Leibe

39

Computer Vision WS 15/16

RWTH AACHEN UNIVERSITY

Detection Results

- Qualitative Performance
 - Recognizes different kinds of objects
 - Robust to clutter, occlusion, noise, low contrast

B. Leibe

42

Computer Vision WS 15/16

RWTH AACHEN UNIVERSITY

Detections Using Ground Plane Constraints

Battery of 5 ISM detectors for different car views

left camera
1175 frames

B. Leibe [Leibe, Cornelis, Cornelis, Van Gool, CVPR'07]

43

Computer Vision WS 15/16

RWTH AACHEN UNIVERSITY

Extension: Rotation-Invariant Detection

- Polar instead of Cartesian voting scheme

- Benefits:
 - Recognize objects under image-plane rotations
 - Possibility to share parts between articulations.
- Caveats:
 - Rotation invariance should only be used when it's really needed. (Also increases false positive detections)

B. Leibe [Mikolajczyk, Leibe, Schiele, CVPR'06]

44

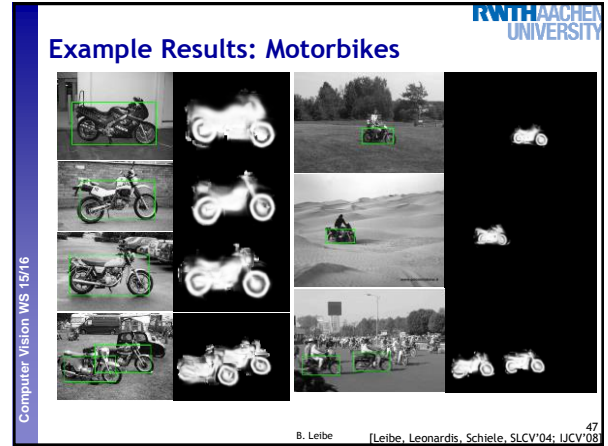
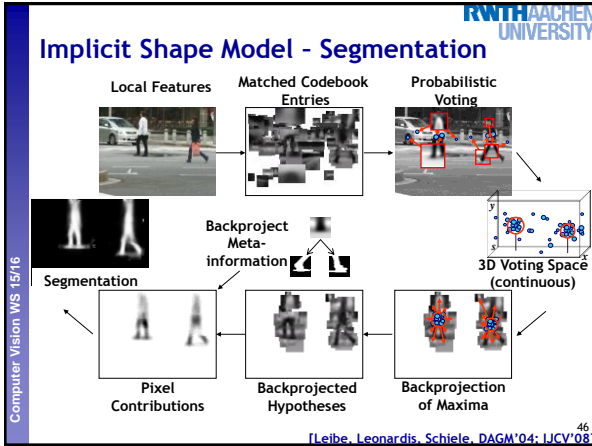
Computer Vision WS 15/16

RWTH AACHEN UNIVERSITY

Sometimes, Rotation Invariance Is Needed...

Figure from [Mikolajczyk et al., CVPR'06]

B. Leibe



RWTH AACHEN UNIVERSITY

You Can Try It At Home...

- Linux source code & binaries available
 - Including datasets & several pre-trained detectors
 - <http://www.vision.rwth-aachen.de/software>

B. Leibe

- RWTH AACHEN UNIVERSITY
- ## Topics of This Lecture
- Recap: Specific Object Recognition with Local Features
 - Part-Based Models for Object Categorization
 - Structure representations
 - Different connectivity structures
 - Bag-of-Words Model
 - Use for image classification
 - Implicit Shape Model
 - Generalized Hough Transform for object category detection
 - Deformable Part-based Model
 - Discriminative part-based detection
- B. Leibe

RWTH AACHEN UNIVERSITY

Starting Point: HOG Sliding-Window Detector

- Array of weights for features in window of HOG pyramid
- Score is dot product of filter and vector

B. Leibe

RWTH AACHEN UNIVERSITY

Deformable Part-based Models

- Mixture of deformable part models (pictorial structures)
- Each component has global template + deformable parts
- Fully trained from bounding boxes alone

Slide credit: Pedro Felzenszwalb B. Leibe

RWTH AACHEN UNIVERSITY

2-Component Bicycle Model

Root filters
coarse resolution

Part filters
finer resolution

Deformation
models

Computer Vision WS 15/16 52

Slide credit: Pedro Felzenszwalb B. Leibe

RWTH AACHEN UNIVERSITY

Object Hypothesis

Computer Vision WS 15/16 53

Slide credit: Pedro Felzenszwalb B. Leibe

Score of filter:
 dot product of filter
 with HOG features
 underneath it

 Score of object
 hypothesis is sum of
 filter scores minus
 deformation costs

- Multiscale model captures features at two resolutions

RWTH AACHEN UNIVERSITY

Score of a Hypothesis

$$\text{score}(p_0, \dots, p_n) = \sum_{i=0}^n F_i \cdot \phi(H, p_i) - \sum_{i=1}^n d_i \cdot (dx_i^2 + dy_i^2)$$

Computer Vision WS 15/16 54

Slide credit: Pedro Felzenszwalb B. Leibe

concatenation filters and
deformation parameters

concatenation of HOG
features and part
displacement features

$\text{score}(z) = \beta \cdot \Psi(H, z)$

RWTH AACHEN UNIVERSITY

Recognition Model

$$f_w(x) = w \cdot \Phi(x)$$

$$f_w(x) = \max_z w \cdot \Phi(x, z)$$

- z : vector of part offsets
- $\Phi(x, z)$: vector of HOG features (from root filter & appropriate part sub-windows) and part offsets

Computer Vision WS 15/16 55

Slide credit: Pedro Felzenszwalb B. Leibe

RWTH AACHEN UNIVERSITY

Results: Persons

- Results (after non-maximum suppression)
 - -1s to search all scales

Computer Vision WS 15/16 56

Slide credit: Pedro Felzenszwalb B. Leibe

RWTH AACHEN UNIVERSITY

Results: Bicycles


Computer Vision WS 15/16 57

Slide adapted from Trevor Darrell B. Leibe

RWTH AACHEN
UNIVERSITY

False Positives

- Bicycles



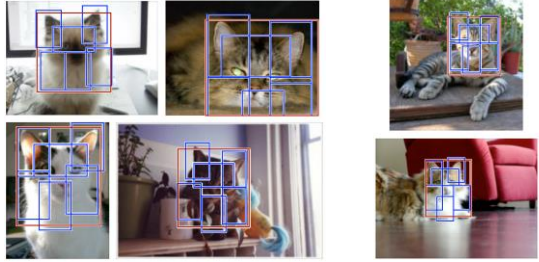
Computer Vision WS 15/16

B. Leibe

58

RWTH AACHEN
UNIVERSITY

Results: Cats



High-scoring true positives

High-scoring false positives
(not enough overlap)

Computer Vision WS 15/16

Slide credit: Pedro Felzenszwalb

59

RWTH AACHEN
UNIVERSITY

You Can Try It At Home...

- Deformable part-based models have been very successful at several recent evaluations.

⇒ State-of-the-art approach in object detection for several years

- Source code and models trained on PASCAL 2006, 2007, and 2008 data are available here:
<http://www.cs.uchicago.edu/~pff/latent>

Computer Vision WS 15/16

B. Leibe

60

RWTH AACHEN
UNIVERSITY

References and Further Reading

- Details about the ISM approach can be found in
 - *B. Leibe, A. Leonardis, and B. Schiele, [Robust Object Detection with Interleaved Categorization and Segmentation](#), International Journal of Computer Vision, Vol. 77(1-3), 2008.*
- Details about the DPMs can be found in
 - *P. Felzenszwalb, R. Girshick, D. McAllester, D. Ramanan, [Object Detection with Discriminatively Trained Part Based Models](#), IEEE Trans. PAMI, Vol. 32(9), 2010.*
- Try the ISM Linux binaries
 - <http://www.vision.ee.ethz.ch/bleibe/code>
- Try the Deformable Part-based Models
 - <http://www.cs.uchicago.edu/~pff/latent>

Computer Vision WS 15/16