Scalability of Object Class Recognition

Bernt Schiele
Max Planck Institute for Informatics, Saarbrücken
Saarland University, Saarbrücken

thanks to: Micha Andriluka, Sandra Ebert, Mario Fritz, Peter Gehler, Michael Goesele, Leonid Pishchulin, Marcus Rohrbach, Stefan Roth, Konrad Schindler, Bojan Pepik, Michael Stark, Christian Wojek

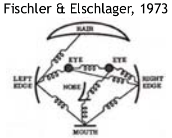
DAGM Workshop on

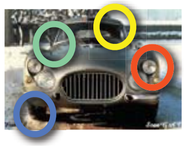
Unsolved Problems in ~~Pattern~~ Recognition

Motivation for Part-Based Models [slide from Pietro Perona]

Model of Parts & Structure: Part-Based Models...

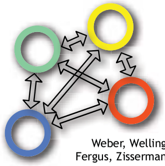
Fischler & Elschlager, 1973



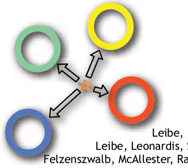


Constellation Model:
Fully connected shape model

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



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

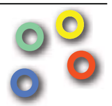


Leibe, Schiele '03
Leibe, Leonardis, Schiele '04
Felzenszwalb, McAllester, Ramanan '08

Object Models State-of-the-Art in Computer Vision

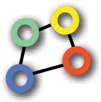
- Bag of Words Models (BoW)
  - object model = histogram of local features
  - e.g. local feature around interest points

BoW: no spatial relationships



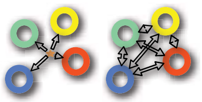
- Global Object Models
  - object model = global feature object feature
  - e.g. HOG (Histograms of Oriented Gradients)

e.g. HOG: fixed spatial relationships



- Part-Based Object Models
  - object model = models of parts & spatial topology model
  - e.g. DPM & ISM & constellation model

e.g. DPM, ISM or constellation model: flexible spatial relationships



Object Class Recognition

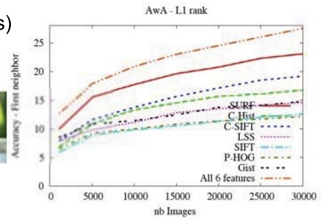
- State-of-the-Art in 2011
  - Computer Vision
    - recognition of few 100s, maybe 1,000s of object classes “possible”
  - Human Vision
    - recognition of several 10,000s of object classes
- Open Research Question for Computer Vision:
  - **how to scale to recognize 1,000s or 10,000s of object classes?**

Scalability of Object Class Recognition
Different Thoughts for Discussion…

- 1. Rethinking our Object Representations
  - current morels don't seem powerful enough
  - we heavily rely on machine learning to make our models "work"
- 2. Labels for Most Object Classes are Sparse
  - we need to reuse and transfer learned knowledge
- 3. Reformulation of Recognition Problem
  - are basic level categories the "right" problem to work on ?
  - what about new ways of addressing recognition ?
- 4. ??

1. Rethinking our Object Representations

- AwA (Animals with Attributes)
  - 50 animals classes



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

1. Rethinking our Object Representations

- ImageNet Challenge 2010
  - 1,000 object classes, about 1,000 images per class total of 1.2 Million images
- Some Results…

Model	Descriptor	Learning method	Total dim.	Err. top 5	Err. top 1
BoW [2]	Sift	LibLinear	1,000	80	91
BoW	Sift	MeanSGD	1,000	72	86
BoW + SPM	rgSift	MeanSGD	8,000	59	76
LLC + SPM	rgSift	MeanSGD	21,000	50	69
Fisher vector	rgSift	MeanSGD	32,768	43	61
LLC+SPM, Fisher	**rgSift**	**MeanSGD**	**53,768**	**38**	**57**
Fisher+SPM [25]	Sift, Color	SGD	262,144	34	
LLC,SVC+SPM [16]	Hog, Lbp	ASGD	1,179,648	28	47

  - **best published result: 47% error !**
  - **the higher the dimensionality - the better the performance !!**
  - **this suggests: our "object models" are not good enough !!!**

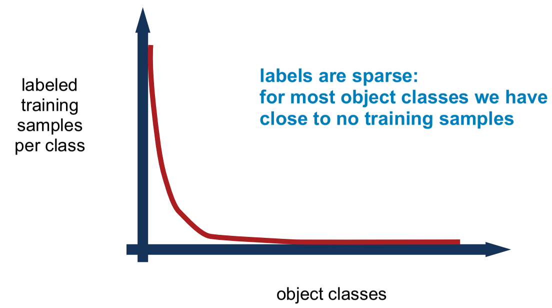
Discussion
1. Rethinking our Object Representations

- Pros of our current approaches:
  - nice scientific problem (learning object models from large databases with or without noisy labels)
  - have shown to work for "small scale" problem
- Cons & Pitfalls:
  - our object representations do not seem to be powerful enough (yet?)
  - powerful machine learning seems to be "key to success"
  - we never seem to have (or at least make use of) enough data
  - majority of work tries to recognize "basic level categories"
    - is that the right problem to work on?

Scalability of Object Class Recognition
Different Thoughts for Discussion…

- 1. Rethinking our Object Representations
  - current morels don't seem powerful enough
  - we heavily rely on machine learning make our models "work"
- 2. Labels for Most Object Classes are Sparse
  - we need to reuse and transfer learned knowledge
- 3. Reformulation of Recognition Problem
  - are basic level categories the "right" problem to work on ?
  - what about new ways of addressing recognition ?
- 4. ??

2. Labels for Most Object Classes are Sparse

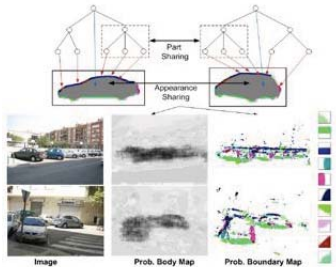


2. Labels for Most Object Classes are Sparse
Potential of Knowledge Transfer

- Assumptions (Observations)
  - different object classes share properties & features (texture, color, ...) or share parts (wheels, head, legs, ...) or ...
  - i.e. one should transfer knowledge
- Large Variety of Approaches:
  - distance metric learning
    (e.g. [Fink@nips04], [Bart&Ullmann@bmvc05], [Thrun@nips06])
  - joint learning of multiple object classes
    (e.g. [Torralba&al@cvpr04], [Amit&al@icml07], [Zhu&al@cvpr10])
  - use of prior information
    (e.g. [Levi&al-04], [Bart&Ullmann@cvpr05], [Fei-Fei@icml06], [Zweig&Weinshall@iccv07], [Stark,Goesele,Schiele@iccv09], [Rodner&Denzler@dagm10])

2. Knowledge Transfer
Some Approaches

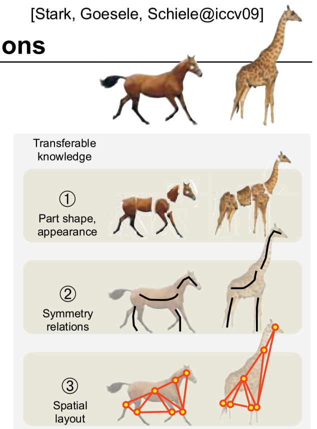
- Part and Appearance Sharing: Recursive Compositional Models for Multi-View Multi-Object Detection
  - [Zhu,Chen,Torralba,Freeman,Yuille@cvpr10]



[Stark, Goesele, Schiele@iccv09]

Motivation and Contributions

- Motivation
  - Scaling recognition to large numbers of object classes suffers from high amount of needed training data
  - Knowledge transfer from existing models may reduce number of required training examples
- Contributions
  - Novel shape-based object class model for explicit knowledge transfer (①, ②, ③)
  - Effective incorporation of symmetries
  - Competitive shape-based object recognition results

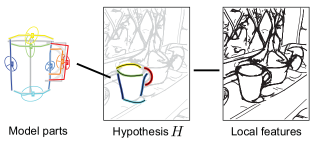


[Stark, Goesele, Schiele@iccv09]

Probabilistic Formulation

- Hypothesis space
  - Assignments between model parts and local image features
    [Fergus et al. CVPR'03]

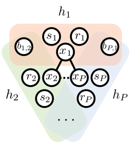
    $H = (h_1, \ldots, h_P)$



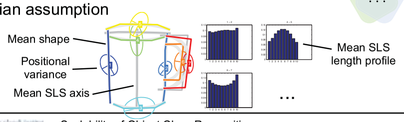
- Factorized joint density

  $p(X, R, S, B, H|\theta) =$

  $\underbrace{p(S|H,\theta)}_{Local\,Shape}\ \underbrace{p(B|H,\theta)}_{Symm.\,Rel.}\ \underbrace{p(R|H,\theta)}_{Rel.\,Scale}\ \underbrace{p(X|H,\theta)}_{Layout}\ \underbrace{p(H|\theta)}_{Prior}$



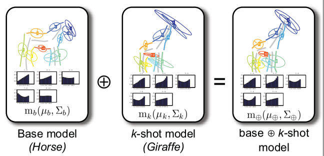
- Gaussian assumption

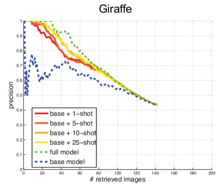


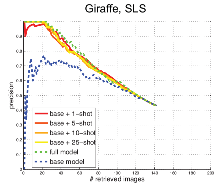
[Stark, Goesele, Schiele@iccv09]

Experiments - Full Model Transfer

- Observations
  - Base model alone *(Horse)* already good on *Elephant, Cheetah*
  - Combined base + *k*-shot models often comparable to full models
  - Models including symmetries (SLS) often superior



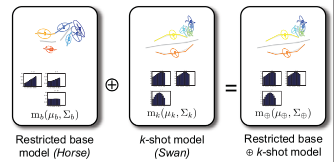


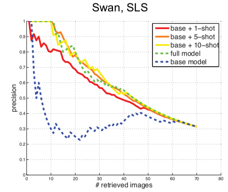


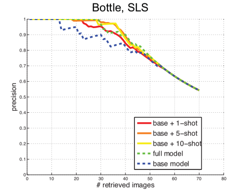
[Stark, Goesele, Schiele@iccv09]

Experiments - Partial Model Transfer

- Observations
  - Base model *(Horse)* fails on *Swan*
  - Base model *(Mug)* good for *Bottle*
  - Symmetries (SLS) often beneficial







2. Knowledge Transfer
Discussion

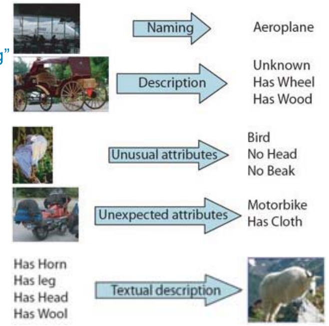
- Pros:
  - intuitive approach
  - typical object-centric representation and transfer
- Cons:
  - approaches still in their infancy
  - many open questions
    - which knowledge to transfer
    - how to transfer knowledge
    - how to identify knowledge to transfer
- Note: Knowledge Transfer is just one way to deal with the problem that labels are sparse for most object classes

Scalability of Object Class Recognition
Different Thoughts for Discussion...

- 1. Rethinking our Object Representations
  - current morels don't seem powerful enough
  - we heavily rely on machine learning make our models "work"
- 2. Labels for Most Object Classes are Sparse
  - we need to reuse and transfer learned knowledge
- 3. Reformulation of Recognition Problem
  - are basic level categories the "right" problem to work on ?
  - what about new ways of addressing recognition ?
- 4. ??

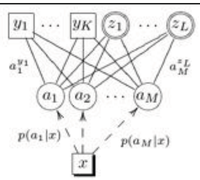
3. Reformulate Problem
Example 1: Attribute-Based...

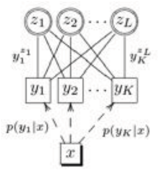
- Describing objects by their attributes [Farhardi&al@cvpr09]
  - "Shift goal of recognition from naming to describing"
  - unusual & unexpected attributes:
  - zero-shot recognition from textual description:



3. Reformulate Problem
Example 1: Attribute-Based...

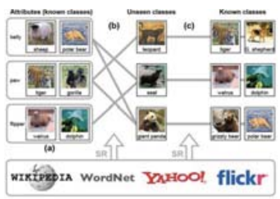
- [Lampert&al@cvpr09]
  - attribute classifiers trained on known classes
    - a_i: attributes
    - y_j: known classes
  - transfer to unknown classes
    - known associations between a_i and z_i: unknown classes
  - second approach: direct similarity (similar to Bart&Ullmann-05)
    - y_i: known classes
    - known associations between y_i and z_i: unknown classes





3. Reformulate Problem
Example 1: Attribute-Based...

- [Rohrbach,Stark,Szarvas,Gurevych,Schiele@cvpr10]:
  - question addressed: how to automatically decide which information to transfer across object classes (setting: zero-shot-learning)
  - using NLP (natural language processing)
    - 1. different language sources (Wikipedia, WordNet, Yahoo!, Flickr, Yahoo! Img Search, ...)
    - 2. different semantic relatedness measures (hitcount, path-length, vector-based, ...)



  - main results:
    - manual supervision of attribute-object-class associations can be fully replaced (in principle) by tapping into language sources
    - Yahoo! Img Search & Wikipedia much better suited than e.g. WordNet...

3. Reformulate Problem
Example 2: Direct Matching Based...

- Assumption:
  - assume large enough databases and efficient enough indexing algorithms
  - direct matching might solve a large portion of the recognition problem...
- Examples
  - [Russell&al@nips07] - Object recognition by scene alignment
  - [Torralba&al@pami08] - 80 million tiny images: a large dataset for non-parametric object and scene recognition
  - [Malisiewicz&Efros@cvpr08,nips09] - Recognition by associations, Visual Memex

[Hinterstoisser,Lepetit,Ilic,Fua,Navab@cvpr10]

Realtime 3D Object Detection

- Approach
  - Represent Object by a Stack of Templates (no machine learning ;-)
  - Real-Time Template Matching
    - Dominant Orientation Template matching (inspired by HOG)
    - Template Clustering & Branch-and-Bound
    - etc.

3. Reformulate Problem Discussion

- Pros:
  - promising ideas & new formulations offer new insights
  - more new ideas are needed !
- Cons:
  - approaches clearly in their infancy
  - open questions:
    - how scalable are those approaches really
    - do we need other / new representations for direct matching/attributes/...
    - ...
- Important Note:
  - these approaches fundamentally challenge the assumption that basic level categories is the "right" level to address object recognition

Scalability of Object Class Recognition Different Thoughts for Discussion...

- 1. Rethinking our Object Representations
  - current morels don't seem powerful enough
  - we heavily rely on machine learning make our models "work"
- 2. Labels for Most Object Classes are Sparse
  - we need to reuse and transfer learned knowledge
- 3. Reformulation of Recognition Problem
  - are basic level categories the "right" problem to work on ?
  - what about new ways of addressing recognition ?
- 4. ??
  - Larger Datasets: LabelMe, ImageNet, SUN Dataset, ...
  - Exploiting Context, Hierarchical Context, ...
  - ...???

max planck institut informatik

UNIVERSITÄT DES SAARLANDES

Scalability of Object Class Recognition

Bernt Schiele
Max Planck Institute for Informatics, Saarbrücken
Saarland University, Saarbrücken

thanks to: Micha Andriluka, Sandra Ebert, Mario Fritz, Peter Gehler, Michael Goesele, Leonid Pishchulin, Marcus Rohrbach, Stefan Roth, Konrad Schindler, Bojan Pepik, Michael Stark, Christian Wojek

Scalability of Object Class Recognition Different Thoughts for Discussion...

- 1. Rethinking our Object Representations
  - current morels don't seem powerful enough
  - we heavily rely on machine learning make our models "work"
- 2. Labels for Most Object Classes are Sparse
  - we need to reuse and transfer learned knowledge
- 3. Reformulation of Recognition Problem
  - are basic level categories the "right" problem to work on ?
  - what about new ways of addressing recognition ?
- 4. ??
  - Larger Datasets: LabelMe, ImageNet, SUN Dataset, ...
  - Exploiting Context, Hierarchical Context, ...
  - ...???