Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories

To appear in CVPR 2006

Svetlana Lazebnik (slazebni@uiuc.edu)
Beckman Institute, University of Illinois at Urbana-Champaign

Cordelia Schmid (cordelia.schmid@inrialpes.fr)
INRIA Rhône-Alpes, France

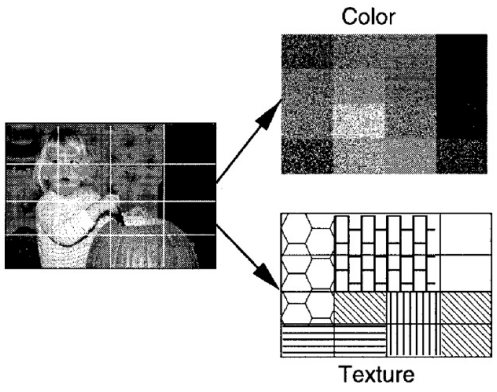
Jean Ponce (ponce@di.ens.fr)
Ecole Normale Supérieure, France

http://www-cvr.ai.uiuc.edu/ponce_grp

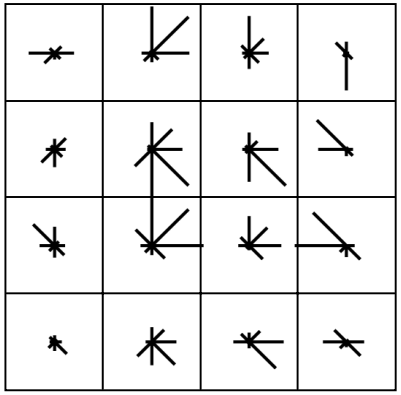
Overview

- A “pre-attentive” approach: recognize the scene as a whole without examining its constituent objects Biederman (1988), Thorpe et al. (1996), Fei-Fei et al. (2002), Renninger & Malik (2004)
- Inspiration: *locally orderless images* Koenderink & Van Doorn (1999)

- Previous work: “subdivide-and-disorder” strategy



Szummer & Picard (1997)

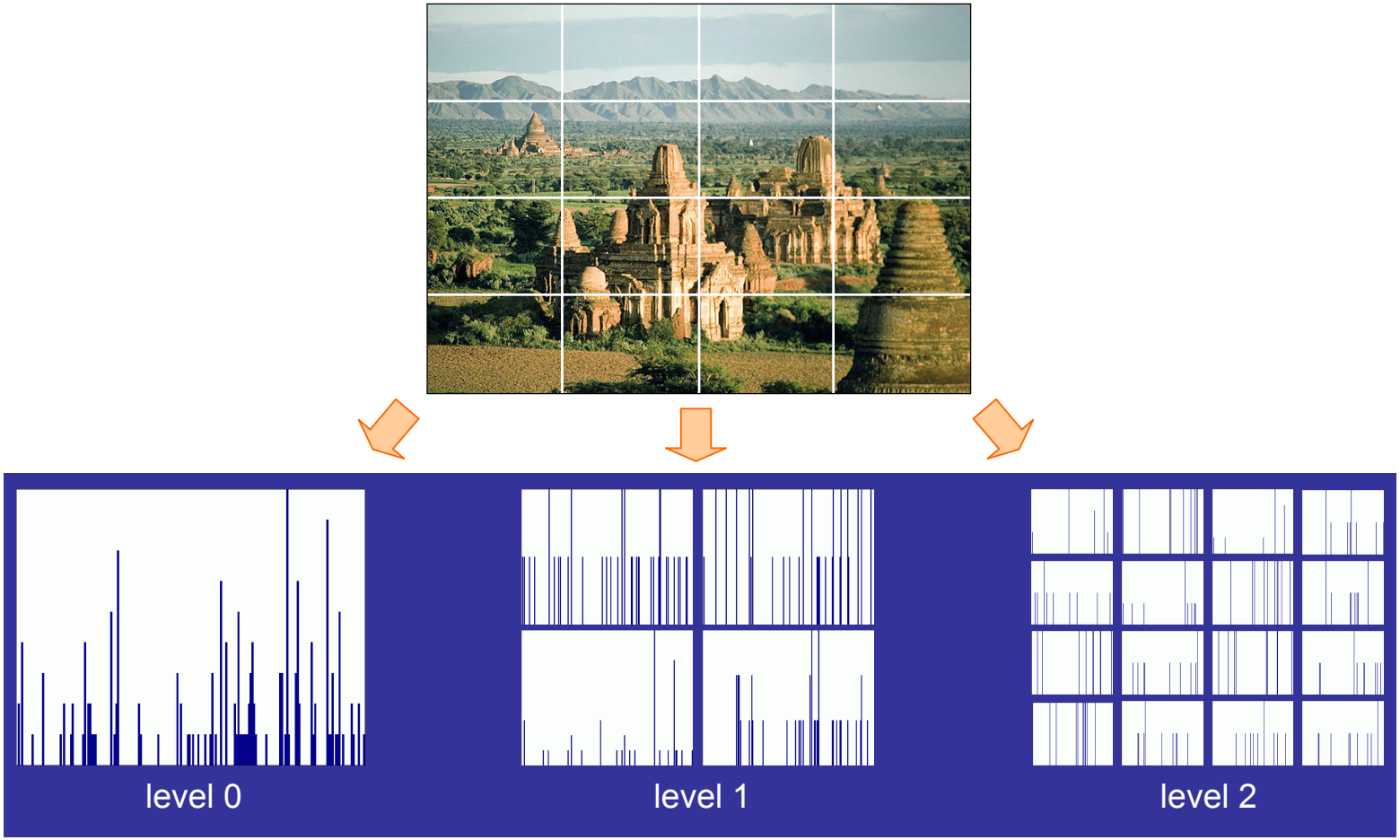


SIFT: Lowe (1999, 2004)

Gist: Torralba et al. (2003)

Spatial pyramid representation

- Extension of a bag of features
- Locally orderless representation at several levels of resolution
- Based on *pyramid match kernels* Grauman & Darrell (2005)
  - Grauman & Darrell: build pyramid in feature space, discard spatial information
  - Our approach: build pyramid in image space, quantize feature space



Pyramid matching

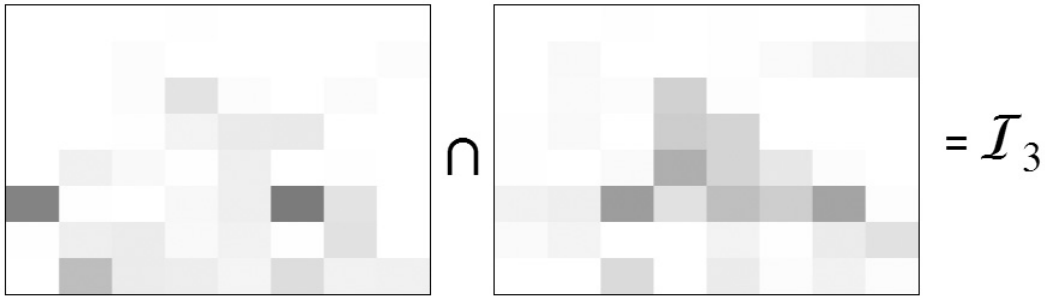
Indyk & Thaper (2003), Grauman & Darrell (2005)

Find maximum-weight matching (weight is inversely proportional to distance)

Original images

Feature histograms:

Level 3



Level 2

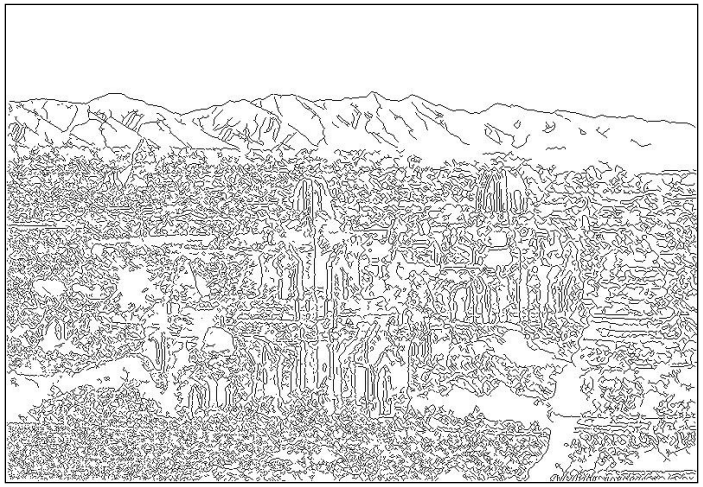
Level 1

Level 0

Total weight (value of *pyramid match kernel*): $\mathcal{I}_3 + \frac{1}{2}(\mathcal{I}_2 - \mathcal{I}_3) + \frac{1}{4}(\mathcal{I}_1 - \mathcal{I}_2) + \frac{1}{8}(\mathcal{I}_0 - \mathcal{I}_1)$

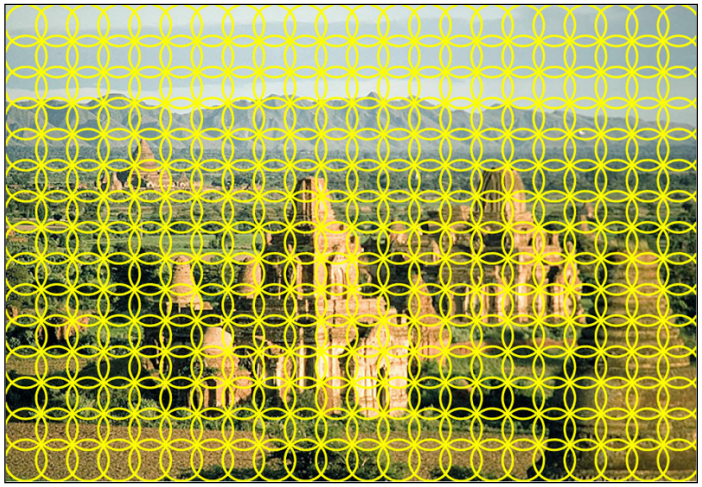
Feature extraction

Weak features



Edge points at 2 scales and 8 orientations (vocabulary size 16)

Strong features



SIFT descriptors of 16x16 patches sampled on a regular grid, quantized to form visual vocabulary (size 200, 400)

Scene category dataset

Fei-Fei & Perona (2005), Oliva & Torralba (2001)

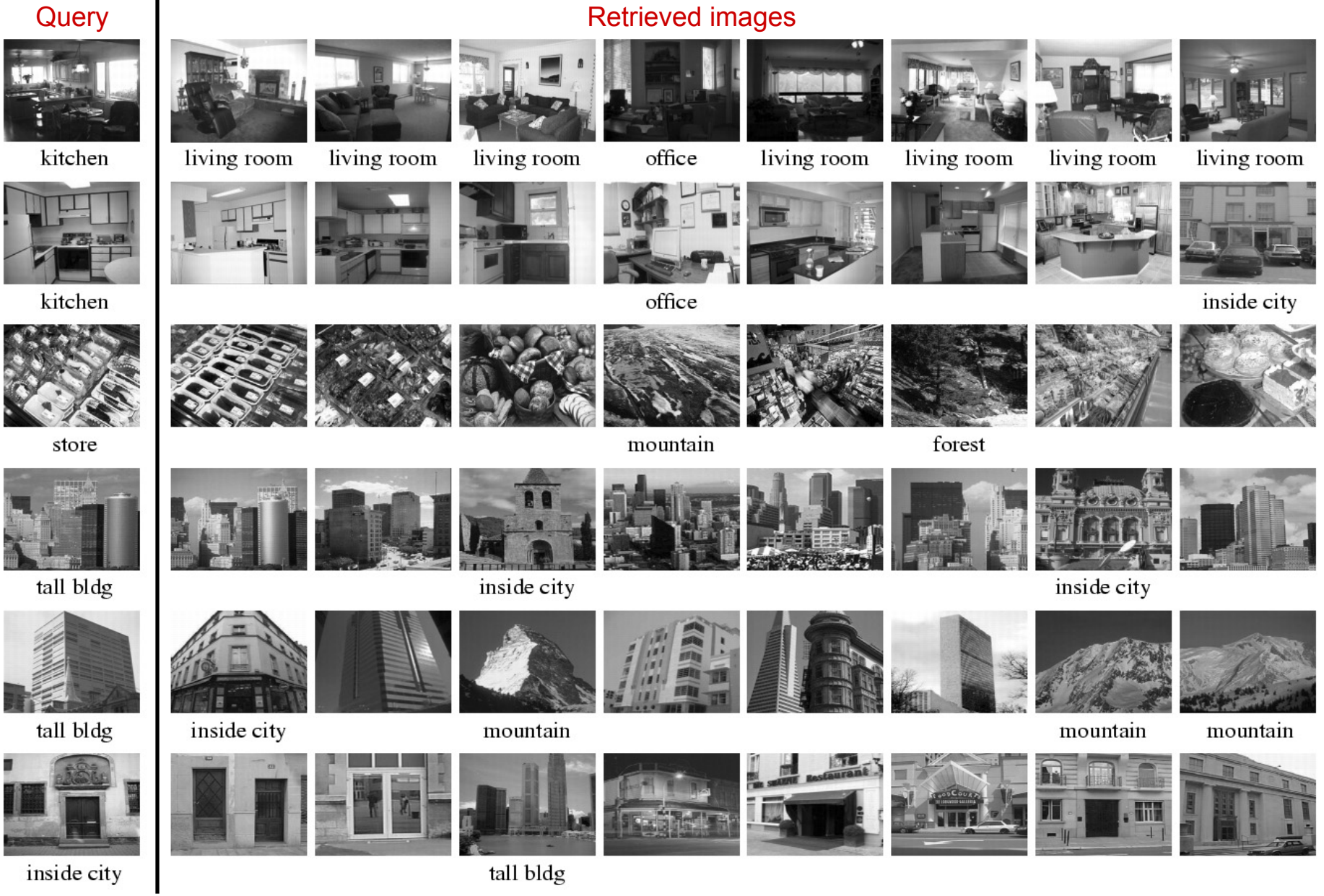
http://www-cvr.ai.uiuc.edu/ponce_grp/data

Multi-class classification results (100 training images per class)

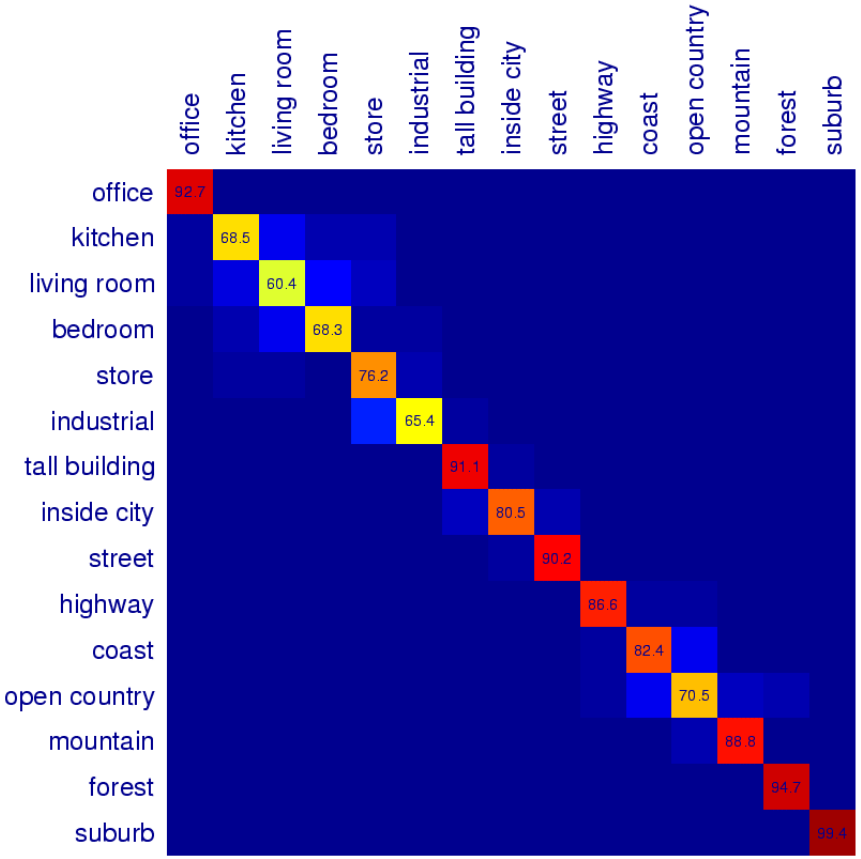
	Weak features (vocabulary size: 16)		Strong features (vocabulary size: 200)	
Level	Single-level	Pyramid	Single-level	Pyramid
0 (1×1)	45.3 ±0.5		72.2 ±0.6	
1 (2×2)	53.6 ±0.3	56.2 ±0.6	77.9 ±0.6	79.0 ±0.5
2 (4×4)	61.7 ±0.6	64.7 ±0.7	79.4 ±0.3	**81.1** ±0.3
3 (8×8)	63.3 ±0.8	**66.8** ±0.6	77.2 ±0.4	80.7 ±0.3

Fei-Fei & Perona: 65.2%

Scene category retrieval



Scene category confusions



Difficult indoor images

kitchen

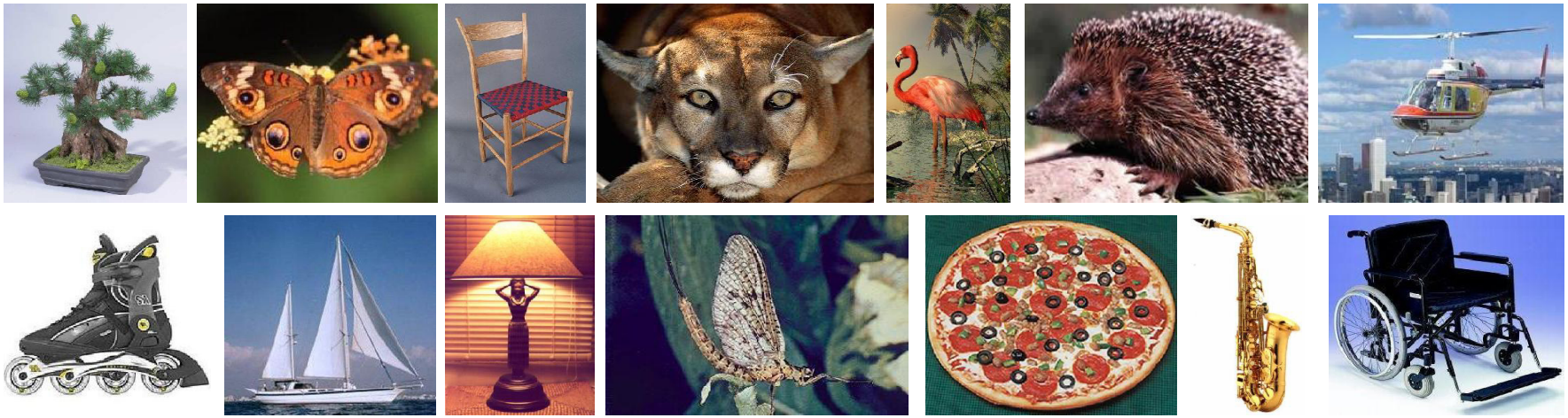
living room

bedroom

Caltech101 dataset

Fei-Fei et al. (2004)

http://www.vision.caltech.edu/Image_Datasets/Caltech101/Caltech101.html

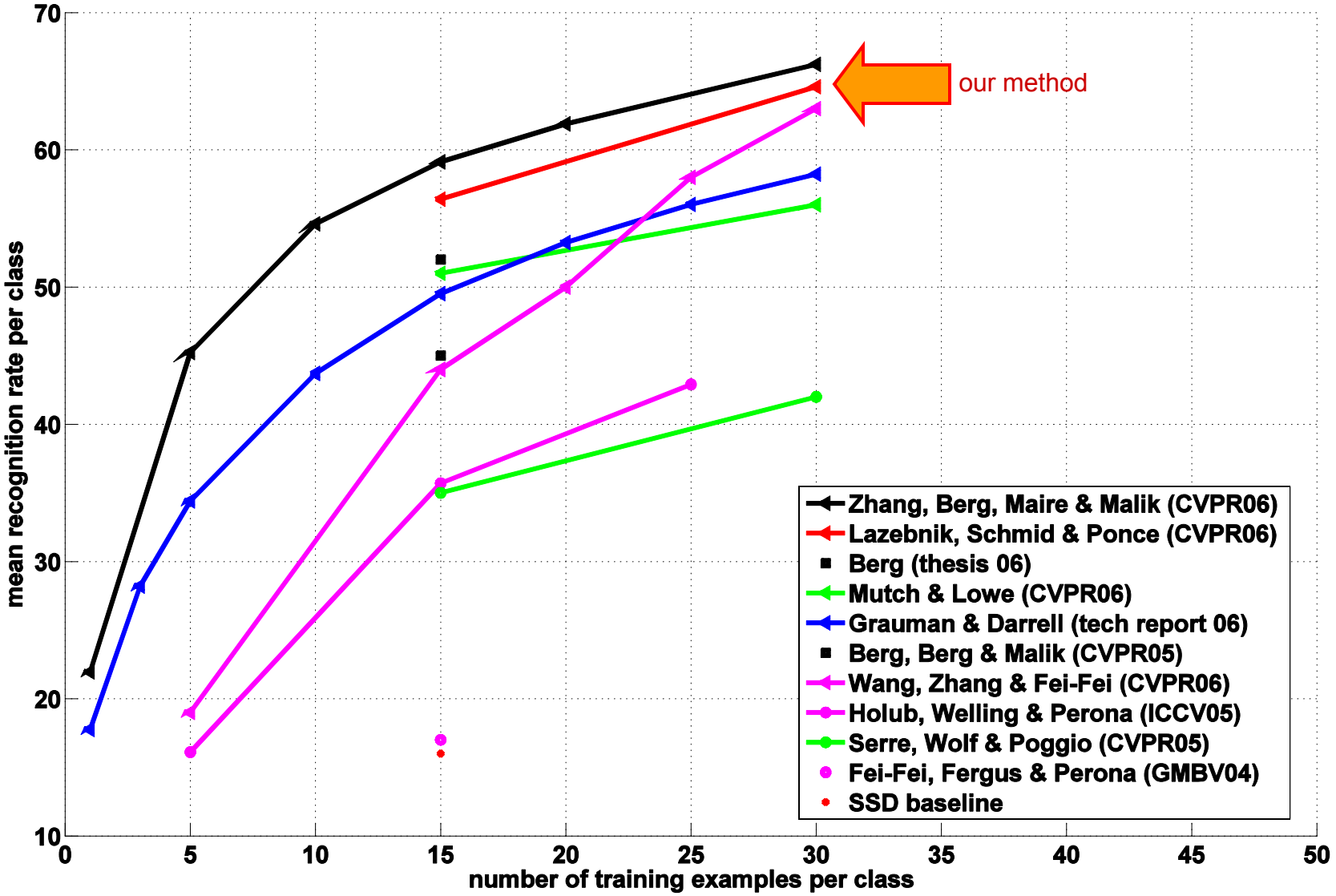


Multi-class classification results (30 training images per class)

	Weak features (16)		Strong features (200)	
Level	Single-level	Pyramid	Single-level	Pyramid
0	15.5 ±0.9		41.2 ±1.2	
1	31.4 ±1.2	32.8 ±1.3	55.9 ±0.9	57.0 ±0.8
2	47.2 ±1.1	49.3 ±1.4	63.6 ±0.9	**64.6** ±0.8
3	52.2 ±0.8	**54.0** ±1.1	60.3 ±0.9	64.6 ±0.7

Caltech101 comparison

Zhang, Berg, Maire & Malik, 2006

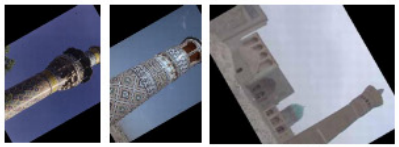


Caltech101 challenges

Top five confusions

class 1 / class 2	class 1 mis-classified as class 2	class 2 mis-classified as class 1
ketch / schooner	21.6	14.8
lotus / water lily	15.3	20.0
crocodile / crocodile head	10.5	10.0
crayfish / lobster	11.3	9.1
flamingo / ibis	9.5	10.4

Easiest and hardest classes


minaret (97.6%)


windsor chair (94.6%)


joshua tree (87.9%)


okapi (87.8%)


cougar body (27.6%)


beaver (27.5%)


crocodile (25.0%)


ant (25.0%)

- Sources of difficulty: lack of texture, camouflage, “thin” objects, highly deformable shape

Graz dataset

Opelt et al. (2004)

`http://www.emt.tugraz.at/~pinz/data/`

bike

person

background

Detection results (100 pos./100 neg. training images)

Class	Spatial pyramid ($M = 200$) $L = 0$	Spatial pyramid ($M = 200$) $L = 2$	Opelt et al. (2004)	Zhang et al. (2005)
Bikes	82.4 ±2.0	86.3 ±2.5	86.5	**92.0**
People	79.5 ±2.3	82.3 ±3.1	80.8	**88.0**

bag-of-features methods

- Global spatial regularities (natural scene statistics) help even in databases with high geometric variability!