
Recognition

Lu Peng

School of Computer Science,
Beijing University of Posts and Telecommunications

Machine Vision Technology								
Semantic information				Metric 3D information				
Pixels	Segments	Images	Videos	Camera		Multi-view Geometry		
Convolutions Edges & Fitting Local features Texture	Segmentation Clustering	Recognition Detection	Motion Tracking	Camera Model	Camera Calibration	Epipolar Geometry	SfM	
10	4	4	2	2	2	2	2	

What we will learn today?

- Introduction to object recognition
 - Representation
 - Learning
 - Recognition
- Bag of Words models
 - Basic representation
 - Different learning and recognition algorithms

2020/4/26

Beijing University of Posts and Telecommunications

2

What are the different visual recognition tasks?



Source: Fei-Fei Li

2020/4/26

Beijing University of Posts and Telecommunications

3

Classification

Does this image contain a building? [yes/no]



Source: Fei-Fei Li

2020/4/26

Beijing University of Posts and Telecommunications

4

Classification

Is this an beach?



Source: Fei-Fei Li

2020/4/26

Beijing University of Posts and Telecommunications

5

Image Search & Organizing photo collections



Source: Fei-Fei Li

2020/4/26

Beijing University of Posts and Telecommunications

6

Detection

Does this image contain a car? [where?]



Source: Fei-Fei Li

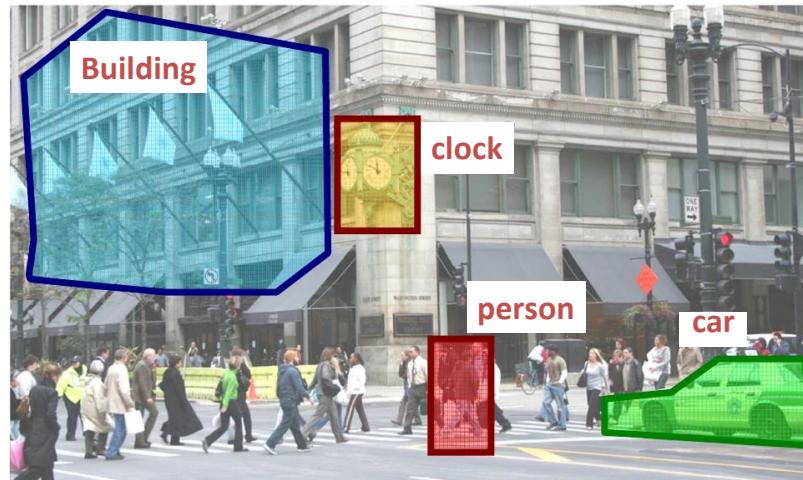
2020/4/26

Beijing University of Posts and Telecommunications

7

Detection

Which object does this image contain? [where?]



2020/4/26

Beijing University of Posts and Telecommunications

8

Detection

Accurate localization (segmentation)



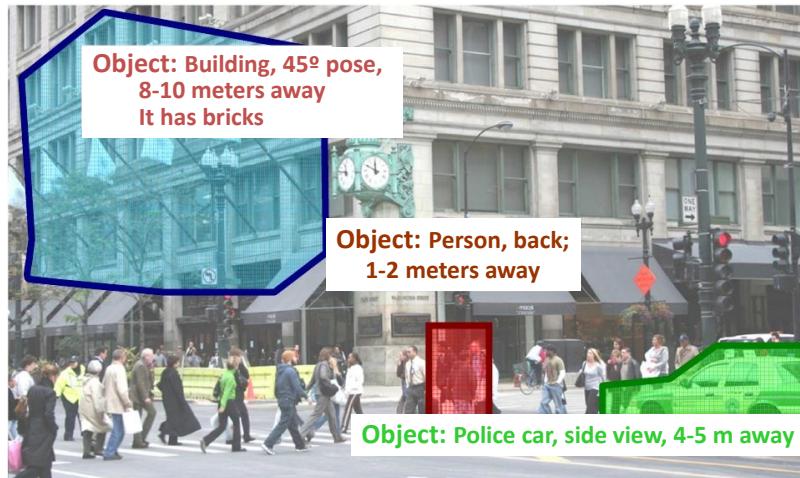
2020/4/26

Beijing University of Posts and Telecommunications

9

Detection

Estimating object semantic & geometric attributes



Source: Fei-Fei Li

2020/4/26

Beijing University of Posts and Telecommunications

10

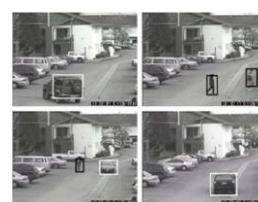
Applications of computer vision



Computational photography



Assistive technologies



Surveillance



Security



Assistive driving

Source: Fei-Fei Li

2020/4/26

Beijing University of Posts and Telecommunications

11

Categorization vs Single instance recognition

Does this image contain the Chicago Macy building's?



2020/4/26

Beijing University of Posts and Telecommunications

12

Categorization vs Single instance recognition

Where is the crunchy nut?



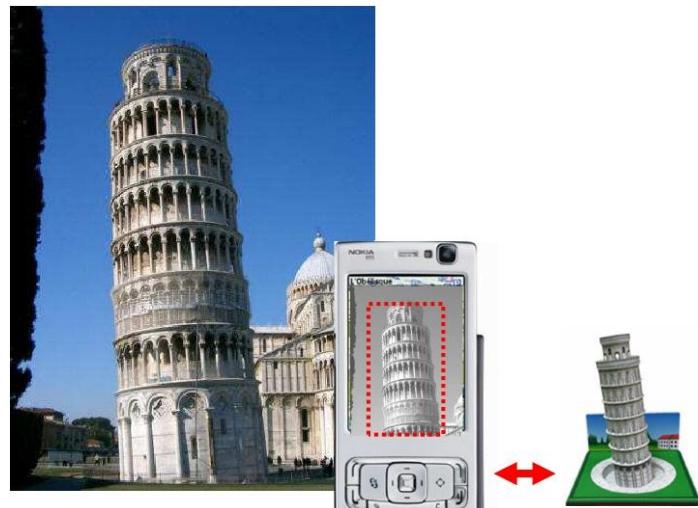
Source: Fei-Fei Li

2020/4/26

Beijing University of Posts and Telecommunications

13

Categorization vs Single instance recognition



Source: Fei-Fei Li

2020/4/26

Beijing University of Posts and Telecommunications

14

Activity or Event recognition

What are these people doing?



Source: Fei-Fei Li

2020/4/26

Beijing University of Posts and Telecommunications

15

Visual Recognition

- Design algorithms that are capable to
 - Classify images or videos
 - Detect and localize objects
 - Estimate semantic and geometrical attributes
 - Classify human activities and events

Why is this challenging?

Source: Fei-Fei Li

2020/4/26

Beijing University of Posts and Telecommunications

16

How many object categories are there?



Source: Fei-Fei Li

2020/4/26

Beijing University of Posts and Telecommunications

17

Challenges: viewpoint variation



Michelangelo 1475-1564

Source: Fei-Fei Li

2020/4/26

Beijing University of Posts and Telecommunications

18

Challenges: illumination



Source: Fei-Fei Li

2020/4/26

Beijing University of Posts and Telecommunications

19

Challenges: scale

and small things
from Apple.
(Actual size)



Source: Fei-Fei Li

2020/4/26

Beijing University of Posts and Telecommunications

20

Challenges: deformation



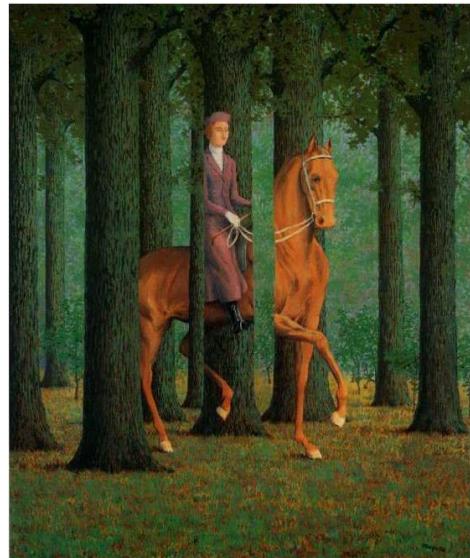
Source: Fei-Fei Li

2020/4/26

Beijing University of Posts and Telecommunications

21

Challenges: occlusion



Magritte, 1957

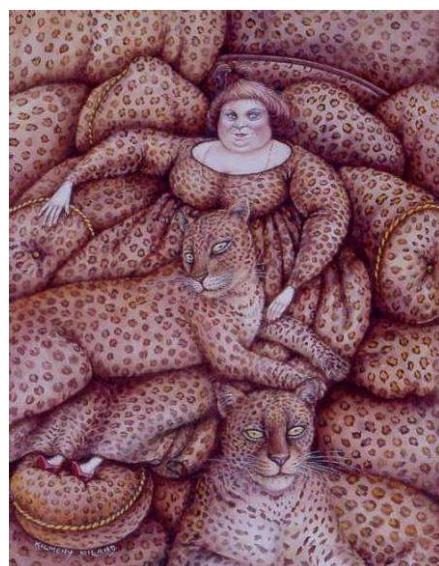
Source: Fei-Fei Li

2020/4/26

Beijing University of Posts and Telecommunications

22

Challenges: background clutter



Kilmeny Niland. 1995

Source: Fei-Fei Li

2020/4/26

Beijing University of Posts and Telecommunications

23

Challenges: intra-class variation



Source: Fei-Fei Li

2020/4/26

Beijing University of Posts and Telecommunications

24

Basic issues

- Representation
 - How to represent an object category; which classification scheme?
- Learning
 - How to learn the classifier, given training data
- Recognition
 - How the classifier is to be used on novel data

Source: Fei-Fei Li

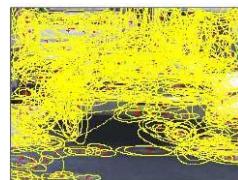
2020/4/26

Beijing University of Posts and Telecommunications

25

Representation

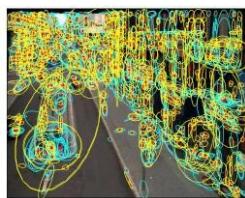
- Building blocks: Sampling strategies



Interest operators



Dense, uniformly



Multiple interest operators

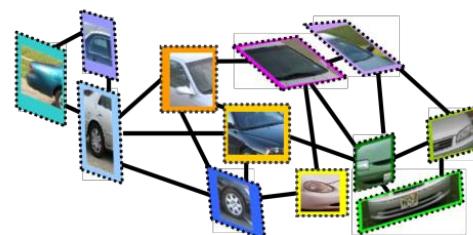
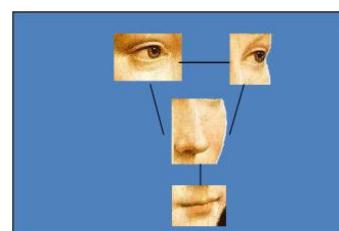
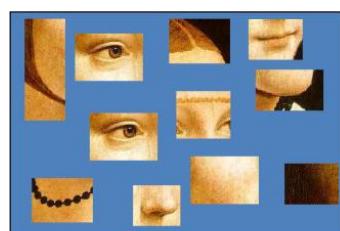


Randomly

Source: Fei-Fei Li

Representation

- Appearance only or location and appearance

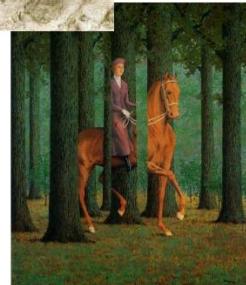


Source: Fei-Fei Li

Representation

Invariances

- View point
- Illumination
- Occlusion
- Scale
- Deformation
- Clutter
- etc.



Source: Fei-Fei Li

Representation

- To handle intra-class variability, it is convenient to describe an object categories using probabilistic models
- Object models: Generative vs Discriminative vs hybrid

关注同类样本内特征
关注样本间差异

Source: Fei-Fei Li

Object categorization: the statistical viewpoint



$p(\text{ zebra} \mid \text{image})$

VS.

$p(\text{ no zebra} \mid \text{image})$

$$\bullet \text{ Bayes rule: } P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

$$\frac{p(\text{zebra} \mid \text{image})}{p(\text{no zebra} \mid \text{image})} = \underbrace{\frac{p(\text{image} \mid \text{zebra})}{p(\text{image} \mid \text{no zebra})}}_{\text{likelihood ratio}} \underbrace{\frac{p(\text{zebra})}{p(\text{no zebra})}}_{\text{prior ratio}}$$

Source: Fei-Fei Li

Object categorization: the statistical viewpoint

- Discriminative methods model posterior
- Generative methods model likelihood and prior

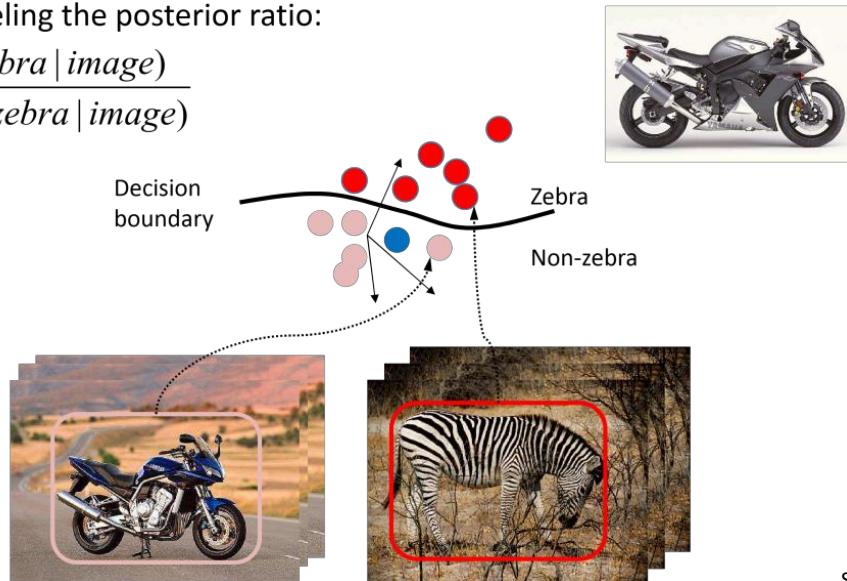
$$\frac{p(\text{zebra} \mid \text{image})}{p(\text{no zebra} \mid \text{image})} = \underbrace{\frac{p(\text{image} \mid \text{zebra})}{p(\text{image} \mid \text{no zebra})}}_{\text{likelihood ratio}} \underbrace{\frac{p(\text{zebra})}{p(\text{no zebra})}}_{\text{prior ratio}}$$

Source: Fei-Fei Li

Discriminative models

- Modeling the posterior ratio:

$$\frac{p(\text{zebra} \mid \text{image})}{p(\text{no zebra} \mid \text{image})}$$



Source: Fei-Fei Li

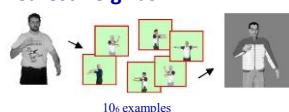
2020/4/26

Beijing University of Posts and Telecommunications

32

Discriminative models

Nearest neighbor



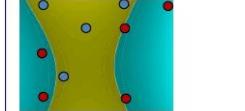
Shakhnarovich, Viola, Darrell 2003
Berg, Berg, Malik 2005...

Neural networks



LeCun, Bottou, Bengio, Haffner 1998
Rowley, Baluja, Kanade 1998
...

Support Vector Machines



Guyon, Vapnik, Heisele,
Serre, Poggio...

Latent SVM Structural SVM



Felzenszwalb 00
Ramanan 03...

Boosting



Viola, Jones 2001,
Torralba et al. 2004,
Opelt et al. 2006,...

Source: Vittorio Ferrari, Kristen Grauman, Antonio Torralba

2020/4/26

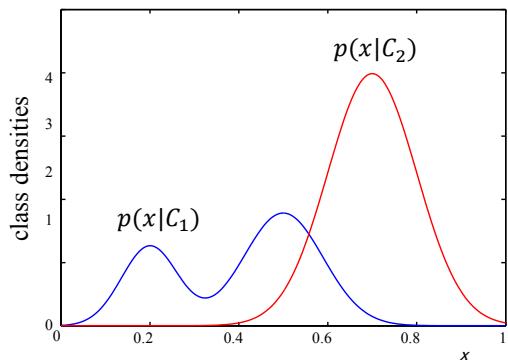
Beijing University of Posts and Telecommunications

33

Generative models

- Modeling the likelihood ratio:

$$\frac{p(\text{image} \mid \text{zebra})}{p(\text{image} \mid \text{no zebra})}$$



Source: Fei-Fei Li

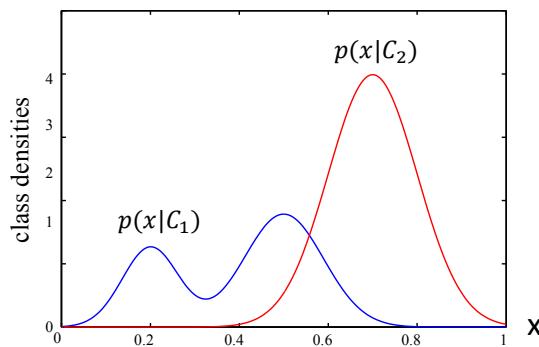
2020/4/26

Beijing University of Posts and Telecommunications

34

Generative models

$P(\text{image} \mid \text{zebra})$	$P(\text{image} \mid \text{no zebra})$
High	Low
Low	High



2020/4/26

Beijing University of Posts and Telecommunications

35

Generative models

- **Naïve Bayes classifier**
 - Csurka Bray, Dance & Fan, 2004
- **Hierarchical Bayesian topic models (e.g. pLSAand LDA)**
 - Object categorization: Sivic et al. 2005, Sudderth et al. 2005
 - Natural scene categorization: Fei-Fei et al. 2005
- **2D Part based models**
 - Constellation models: Weber et al 2000; Fergus et al 200
 - Star models: ISM (Leibe et al 05)
- **3D part based models:**
 - multi-aspects: Sun, et al, 2009

Source: Fei-Fei Li

Basic issues

- Representation
 - How to represent an object category; which classification scheme?
- Learning
 - How to learn the classifier, given training data
- Recognition
 - How the classifier is to be used on novel data

Source: Fei-Fei Li

Learning

- **Learning parameters: What are you maximizing?**

Likelihood (Gen.) or performances on train/validation set (Disc.)

- **Level of supervision**

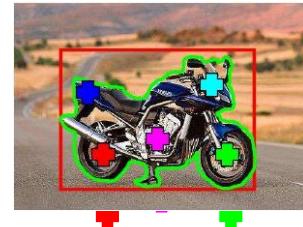
- Manual segmentation; bounding box; image labels; noisy labels

- **Batch/incremental**

- **Priors**

- **Training images:**

- Issue of overfitting
- Negative images for discriminative methods



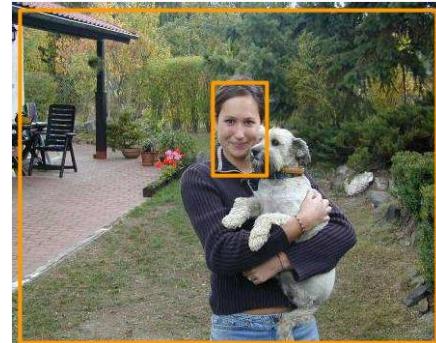
Source: Fei-Fei Li

Basic issues

- Representation
 - How to represent an object category; which classification scheme?
- Learning
 - How to learn the classifier, given training data
- Recognition
 - How the classifier is to be used on novel data

Recognition

- Recognition task: classification, detection, etc. .



Source: Fei-Fei Li

2020/4/26

Beijing University of Posts and Telecommunications

40

Recognition

- Recognition task: classification, detection, etc. .
- Search strategy: Sliding Windows
 - Simple
 - Computational complexity (x, y, S, θ, N of classes)
 - BSW by Lampert et al 08
 - Also, Alexe, et al 10



Source: Fei-Fei Li

2020/4/26

Beijing University of Posts and Telecommunications

41

Recognition

- Recognition task: classification, detection, etc. .
- Search strategy: Sliding Windows
 - Simple
 - Computational complexity (x, y, S, θ, N of classes)
 - BSW by Lampert et al 08
 - Also, Alexe, et al 10
 - Localization
 - Objects are not boxes
 - Prone to false positive



Source: Fei-Fei Li

2020/4/26

Beijing University of Posts and Telecommunications

42

Recognition

- Recognition task: classification, detection, etc. .
- Search strategy: Sliding Windows
 - Simple
 - Computational complexity (x, y, S, θ, N of classes)
 - BSW by Lampert et al 08
 - Also, Alexe, et al 10
 - Localization
 - Objects are not boxes
 - Prone to false positive

Non max suppression:

Canny '86

....

Desai et al , 2009



Source: Fei-Fei Li

2020/4/26

Beijing University of Posts and Telecommunications

43

Basic issues

- Representation
 - How to represent an object category; which classification scheme?
- Learning
 - How to learn the classifier, given training data
- Recognition
 - How the classifier is to be used on novel data

2020/4/26

Beijing University of Posts and Telecommunications

44

Bag-of-features models 词袋模型



Source: Lazebnik

2020/4/26

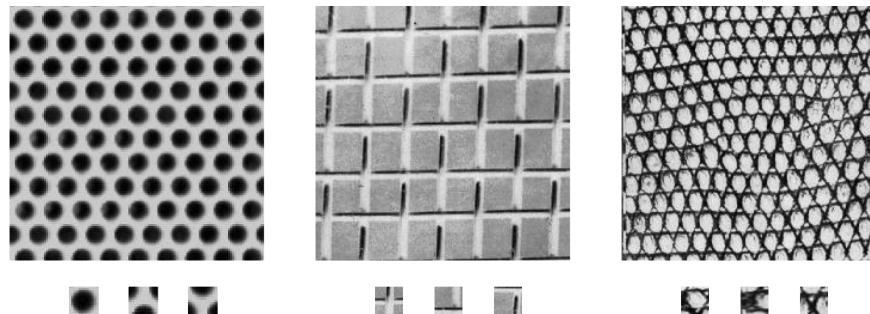
Beijing University of Posts and Telecommunications

45

Origin 1: Texture recognition

Texture is characterized by the repetition of basic elements or *textons*

For stochastic textures, it is the identity of the textons, not their spatial arrangement, that matters



Julesz, 1981; Cula & Dana, 2001; Leung & Malik 2001; Mori, Belongie & Malik, 2001;
Schmid 2001; Varma & Zisserman, 2002, 2003; Lazebnik, Schmid & Ponce, 2003

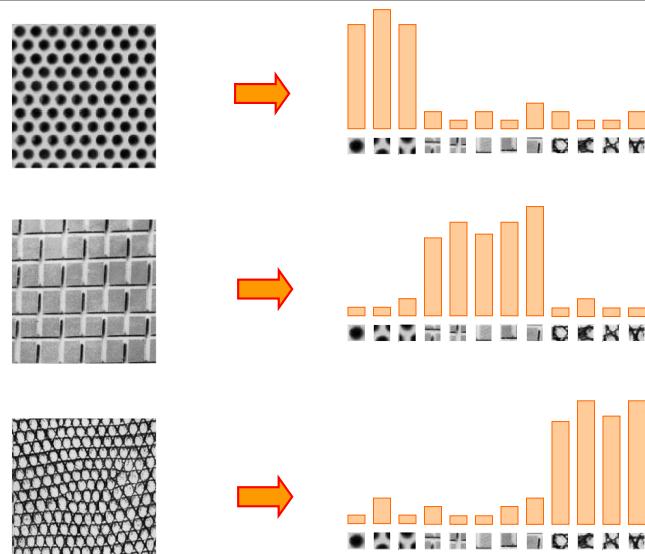
Source: Lazebnik

2020/4/26

Beijing University of Posts and Telecommunications

46

Origin 1: Texture recognition



Julesz, 1981; Cula & Dana, 2001; Leung & Malik 2001; Mori, Belongie & Malik, 2001;
Schmid 2001; Varma & Zisserman, 2002, 2003; Lazebnik, Schmid & Ponce, 2003

Source: Lazebnik

2020/4/26

Beijing University of Posts and Telecommunications

47

Origin 2: Bag-of-words models

Orderless document representation: frequencies of words from a dictionary Salton & McGill (1983)

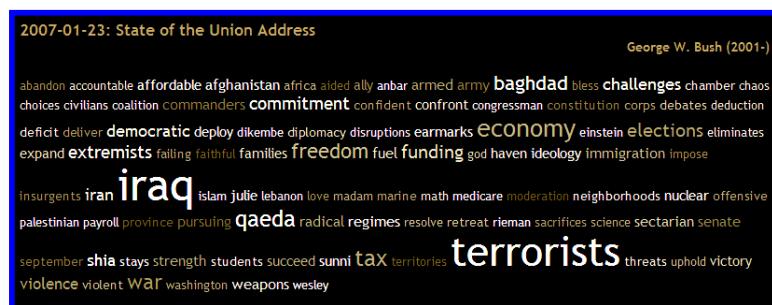
2020/4/26

Beijing University of Posts and Telecommunications

48

Origin 2: Bag-of-words models

Orderless document representation: frequencies of words from a dictionary Salton & McGill (1983)



US Presidential Speeches Tag Cloud
<http://chir.ag/phernalia/preztags/>

Source: Lazebnik

2020/4/26

Beijing University of Posts and Telecommunications

49

Origin 2: Bag-of-words models

Orderless document representation: frequencies of words from a dictionary Salton & McGill (1983)



US Presidential Speeches Tag Cloud
<http://chir.ag/phernalia/preztags/>

Source: Lazebnik

2020/4/26

Beijing University of Posts and Telecommunications

50

Origin 2: Bag-of-words models

Orderless document representation: frequencies of words from a dictionary Salton & McGill (1983)



US Presidential Speeches Tag Cloud
<http://chir.ag/phernalia/preztags/>

Source: Lazebnik

2020/4/26

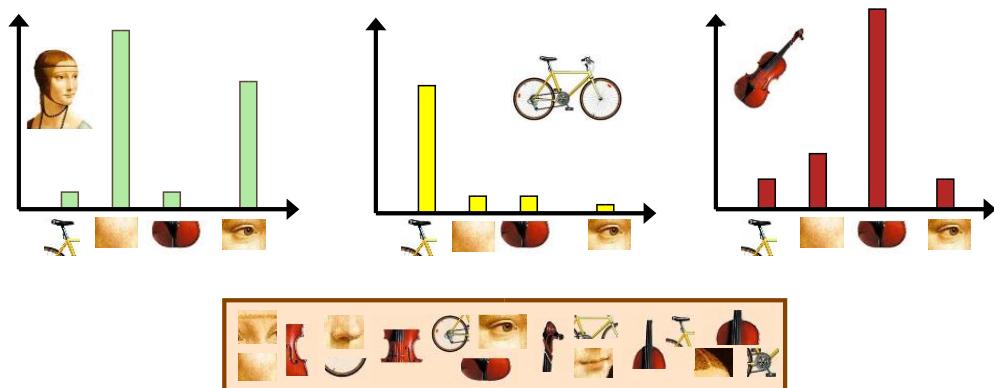
Beijing University of Posts and Telecommunications

51

Bag-of-features steps

1. Extract features
2. Learn “visual vocabulary”
3. Represent images by frequencies of “visual words”

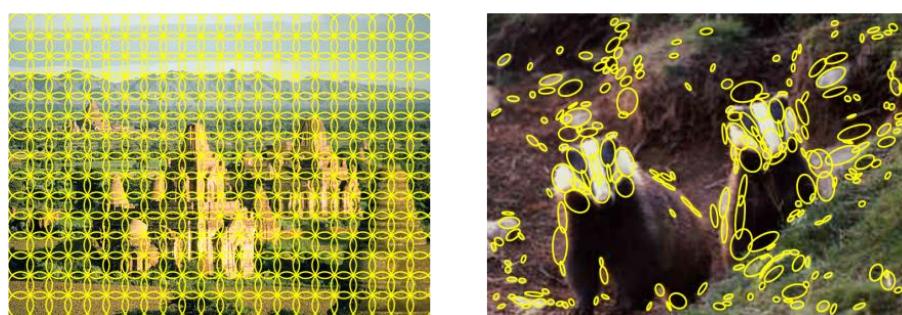
特征维度 = 词典大小



Source: Lazebnik

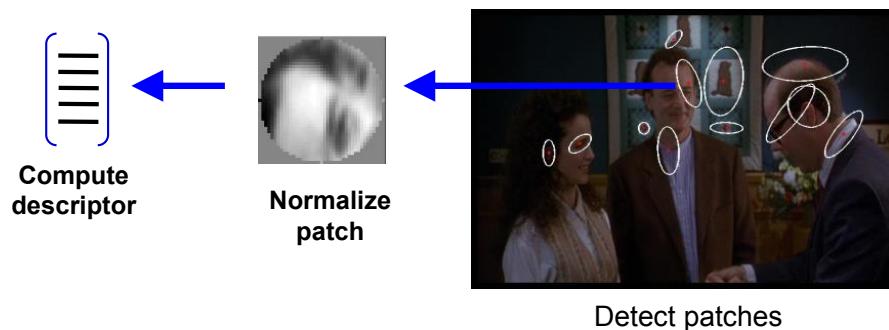
1. Feature extraction

Regular grid or interest regions



Source: Lazebnik

1. Feature extraction



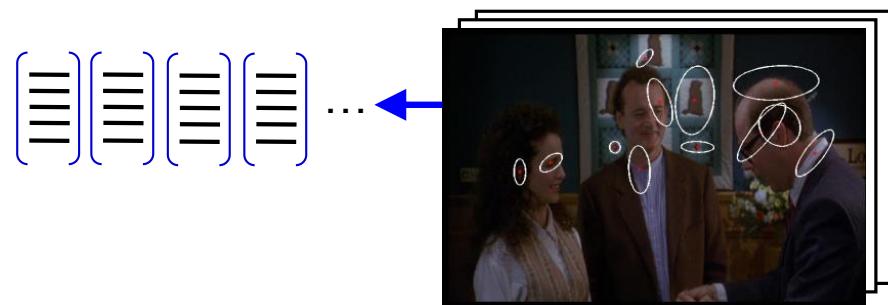
Source: Josef Sivic

2020/4/26

Beijing University of Posts and Telecommunications

54

1. Feature extraction



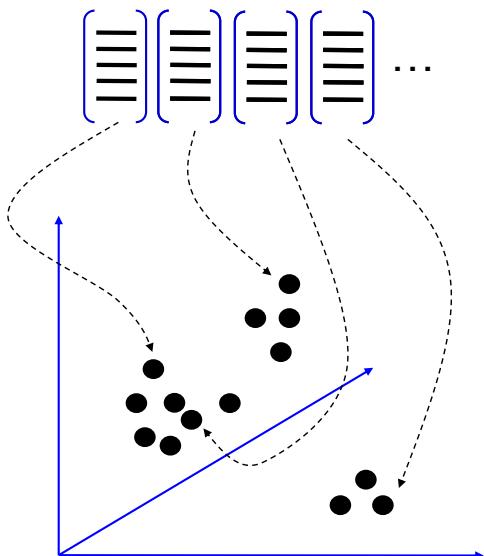
Source: Josef Sivic

2020/4/26

Beijing University of Posts and Telecommunications

55

2. Learning the visual vocabulary



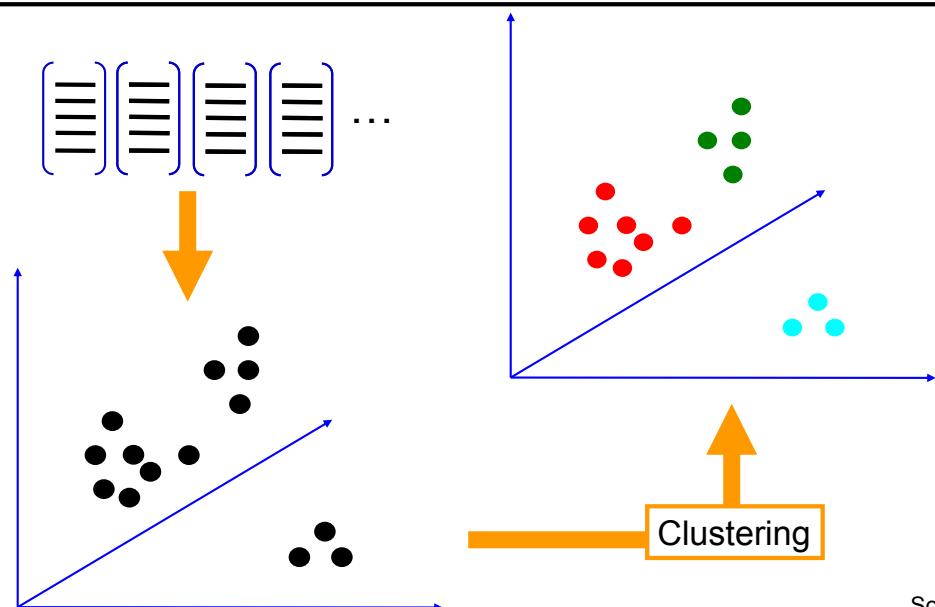
Source: Josef Sivic

2020/4/26

Beijing University of Posts and Telecommunications

56

2. Learning the visual vocabulary



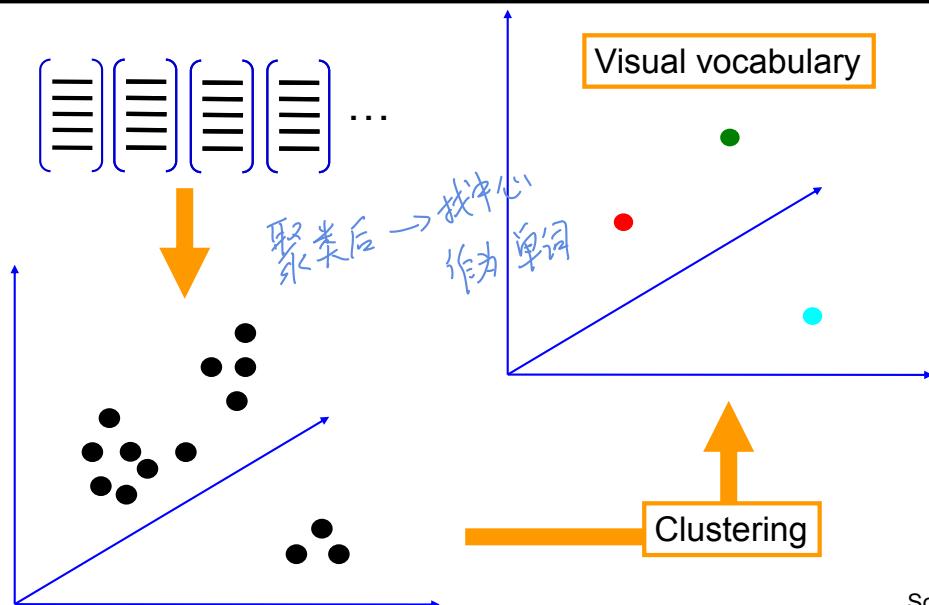
Source: Josef Sivic

2020/4/26

Beijing University of Posts and Telecommunications

57

2. Learning the visual vocabulary



Source: Josef Sivic

2020/4/26

Beijing University of Posts and Telecommunications

58

K-means clustering

- Want to minimize sum of squared Euclidean distances between points x_i and their nearest cluster centers m_k

$$D(X, M) = \sum_{\text{cluster } k} \sum_{\text{point } i \text{ in cluster } k} (x_i - m_k)^2$$

Algorithm:

- Randomly initialize K cluster centers
- Iterate until convergence:
 - Assign each data point to the nearest center
 - Recompute each cluster center as the mean of all points assigned to it

Source: Lazebnik

2020/4/26

Beijing University of Posts and Telecommunications

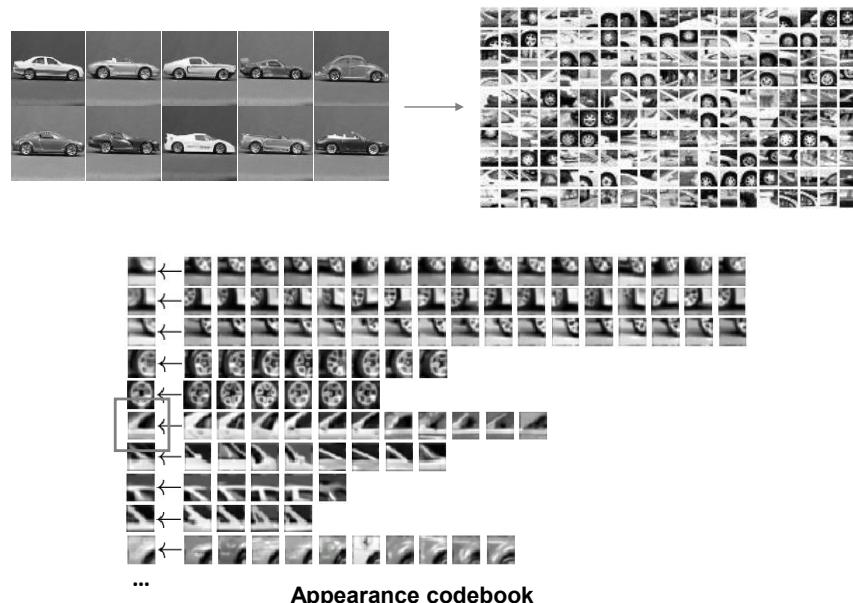
59

Clustering and vector quantization

- Clustering is a common method for learning a visual vocabulary or codebook
 - Unsupervised learning process
 - Each cluster center produced by k-means becomes a codevector
 - Codebook can be learned on separate training set
 - Provided the training set is sufficiently representative, the codebook will be “universal”
- The codebook is used for quantizing features
 - A *vector quantizer* takes a feature vector and maps it to the index of the nearest codevector in a codebook
 - Codebook = visual vocabulary
 - Codevector = visual word

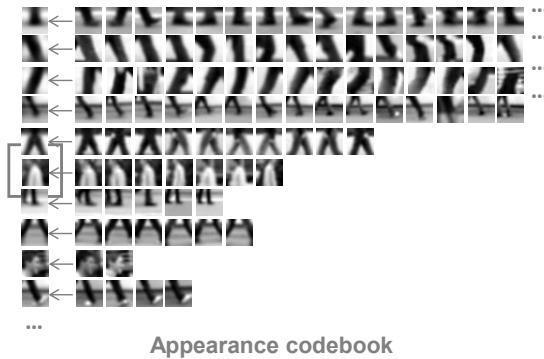
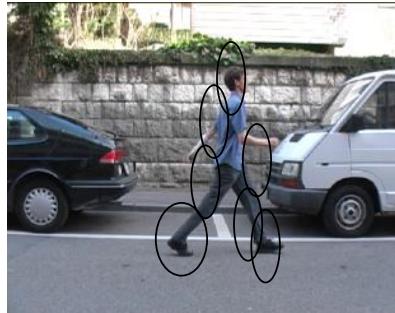
Source: Lazebnik

Example codebook



Source: B. Leibe

Another codebook



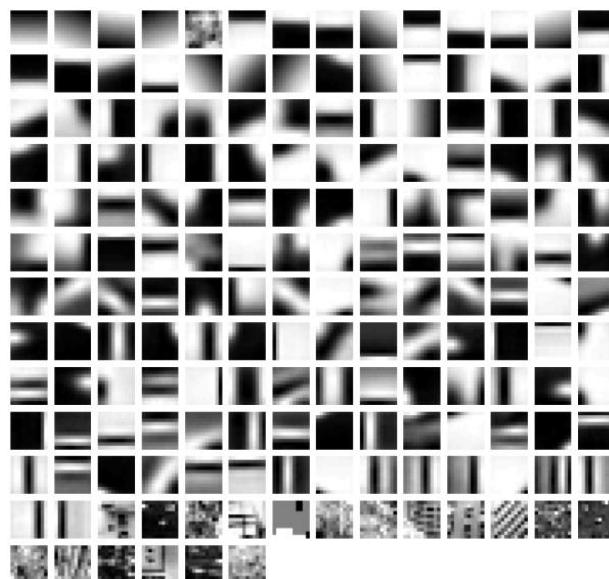
Source: B. Leibe

2020/4/26

Beijing University of Posts and Telecommunications

62

Yet another codebook



Source: Fei-Fei Li

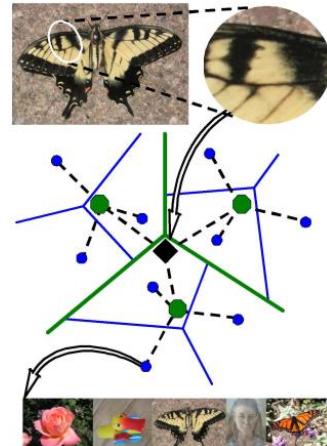
2020/4/26

Beijing University of Posts and Telecommunications

63

Visual vocabularies: Issues

- How to choose vocabulary size?
 - Too small: visual words not representative of all patches
 - Too large: quantization artifacts, overfitting
- Computational efficiency
 - Vocabulary trees
(Nister & Stewenius, 2006)



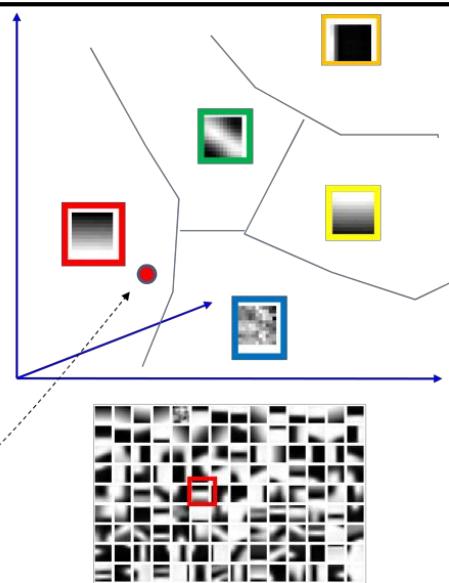
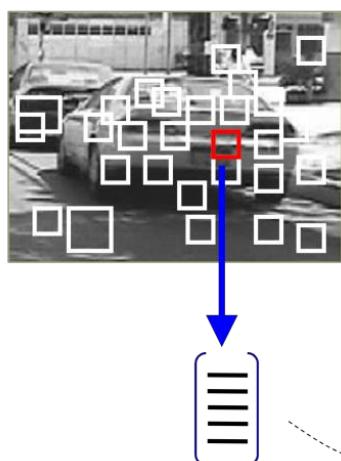
Source: Lazebnik

2020/4/26

Beijing University of Posts and Telecommunications

64

3. Bag of word representation



- Nearest neighbors assignment
- K-D tree search strategy

Codewords dictionary

Source: Fei-Fei Li

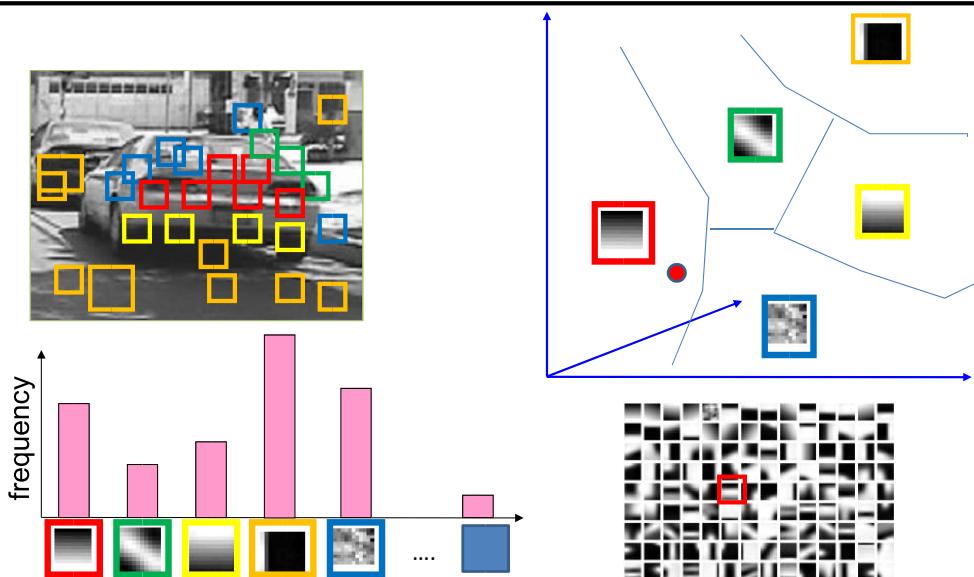
2020/4/26

Beijing University of Posts and Telecommunications

65

词典有了之后，将需被分类的图片进行裁取(扫描)，把结果与词典内的单词作比较
统计出直方图，这个直方图就是
这张图的特征向量

3. Bag of word representation



Source: Fei-Fei Li

2020/4/26

Beijing University of Posts and Telecommunications

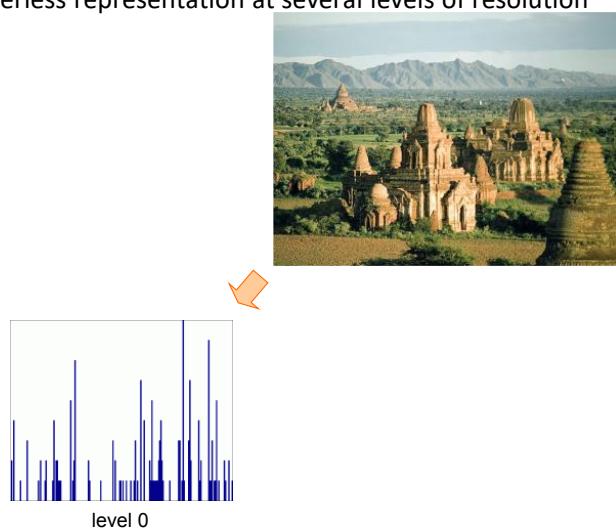
66

Spatial pyramid representation

Lazebnik, Schmid & Ponce (CVPR 2006)

Extension of a bag of features

Locally orderless representation at several levels of resolution



2020/4/26

Beijing University of Posts and Telecommunications

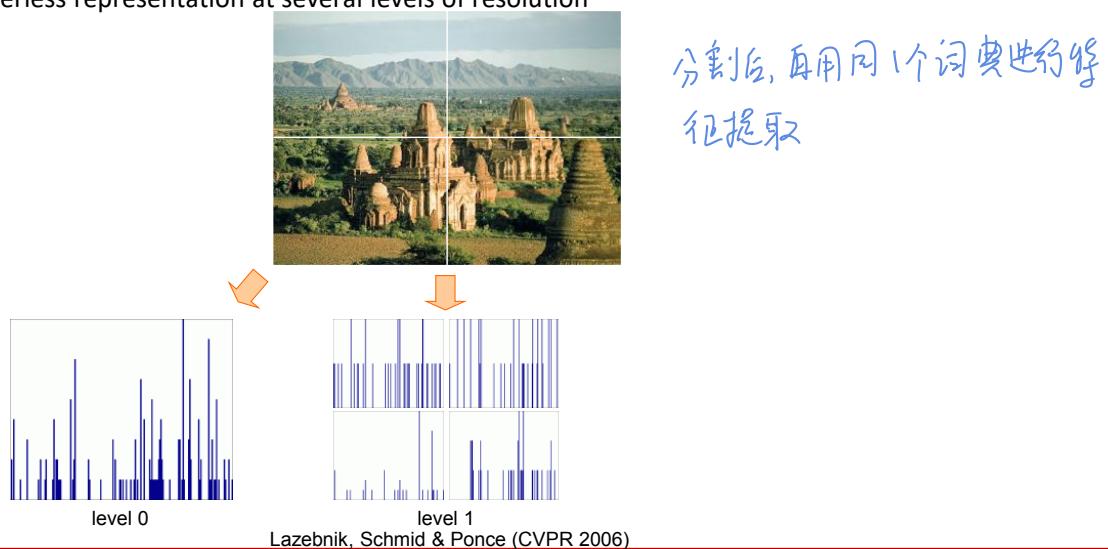
67

Spatial pyramid representation

Lazebnik, Schmid & Ponce (CVPR 2006)

Extension of a bag of features

Locally orderless representation at several levels of resolution



2020/4/26

Beijing University of Posts and Telecommunications

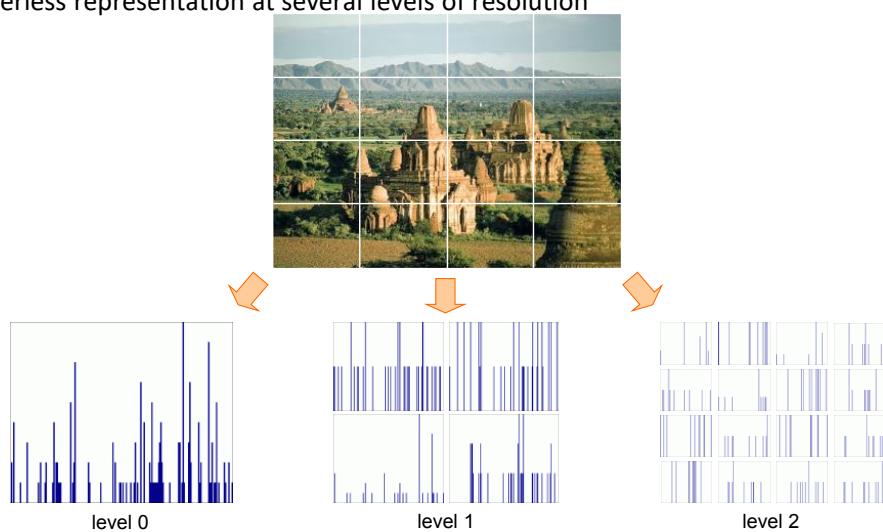
68

Spatial pyramid representation

Lazebnik, Schmid & Ponce (CVPR 2006)

Extension of a bag of features

Locally orderless representation at several levels of resolution



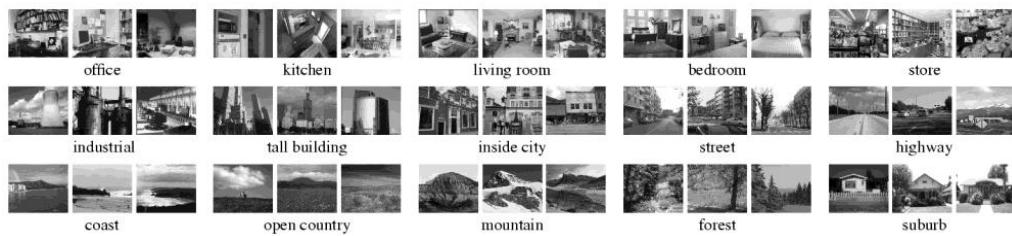
2020/4/26

Beijing University of Posts and Telecommunications

69

Scene category dataset

Lazebnik, Schmid & Ponce (CVPR 2006)



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

Level	Weak features (vocabulary size: 16)		Strong features (vocabulary size: 200)	
	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

2020/4/26

Beijing University of Posts and Telecommunications

70

Caltech101 dataset

Lazebnik, Schmid & Ponce (CVPR 2006)



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

Level	Weak features (16)		Strong features (200)	
	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

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

2020/4/26

Beijing University of Posts and Telecommunications

71