

上海交通大学

计算机视觉

教师: 赵旭

班级: AI4701

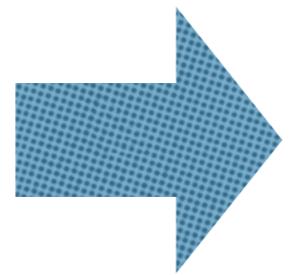
2024 春

12. 视觉词袋表示

内容

- ❖ 直方图特征表示
- ❖ 视觉词袋模型
- ❖ 应用：图像检索

如何表示一张图像？

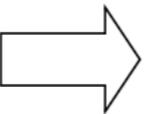


$$\mathbf{x} = (x_1, x_2, \dots, x_n)^T$$

如何表示一张图像？

❖ 元像素

把所有像素值顺序连起来形成1维向量



Credit: The Face Research Lab

如何表示一张图像？

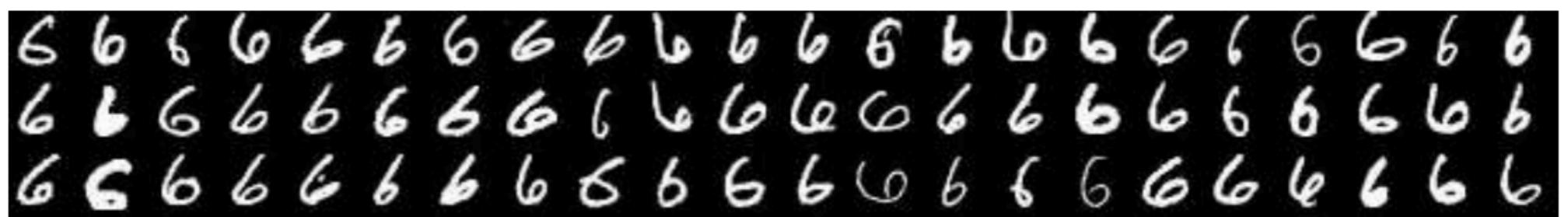
❖ 元像素

把所有像素值顺序连起来形成1维向量

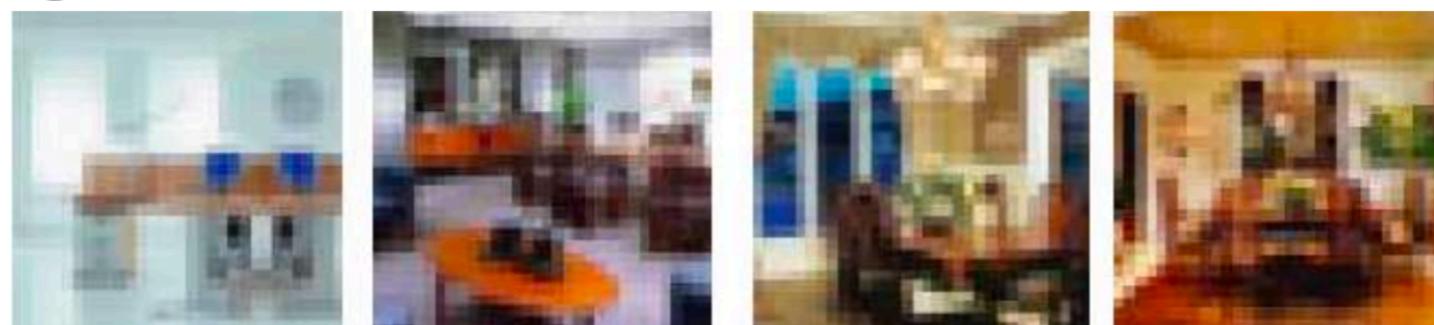
人脸识别



手写数字识别



小图像识别



office

waiting area

dining room

dining room

如何表示一张图像？

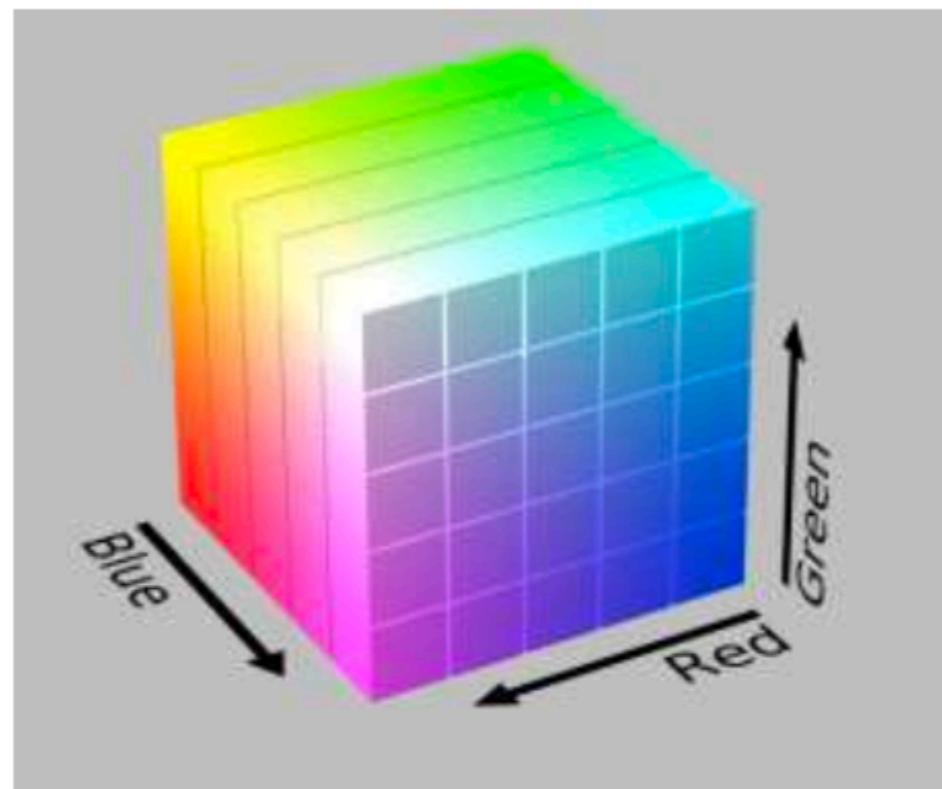
- ❖ 元像素表示的问题
 - ❖ 要严格对齐
 - ❖ 图像尺度相似
 - ❖ 无法处理遮挡
 - ❖ 视角变化敏感

如何表示一张图像？

- ❖ 颜色直方图表示



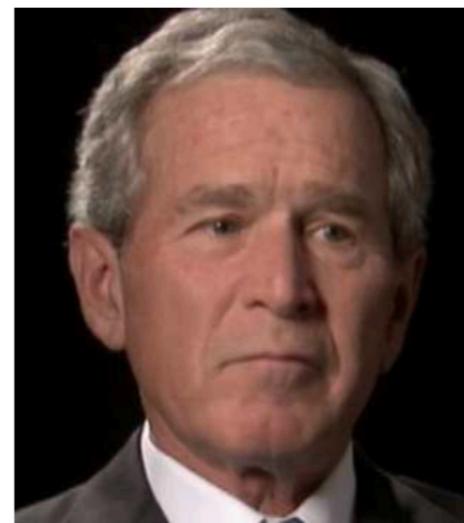
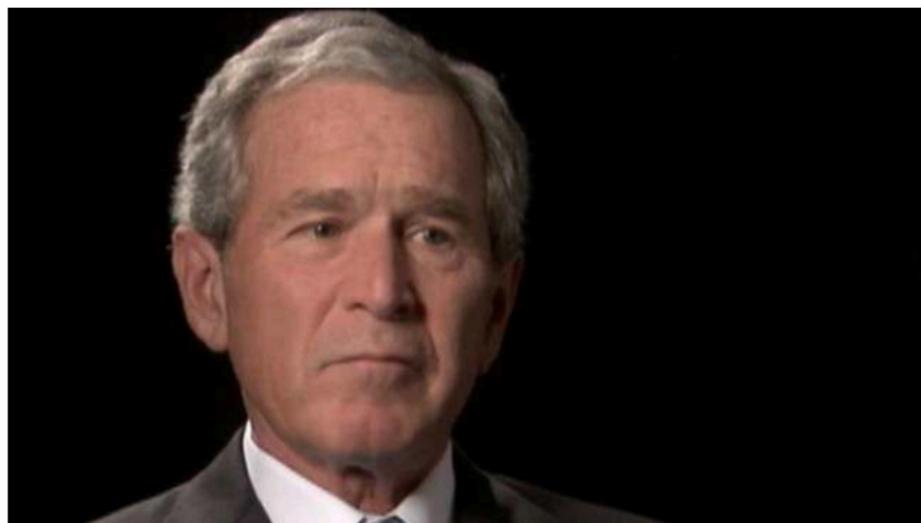
$$\begin{pmatrix} r \\ g \\ b \end{pmatrix}$$
A curved arrow points from the top right corner of the thumbnail image to a vector equation representing the color components of a pixel.



颜色直方图

如何表示一张图像？

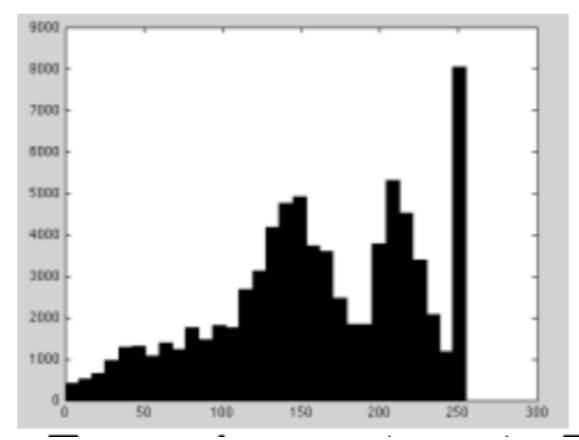
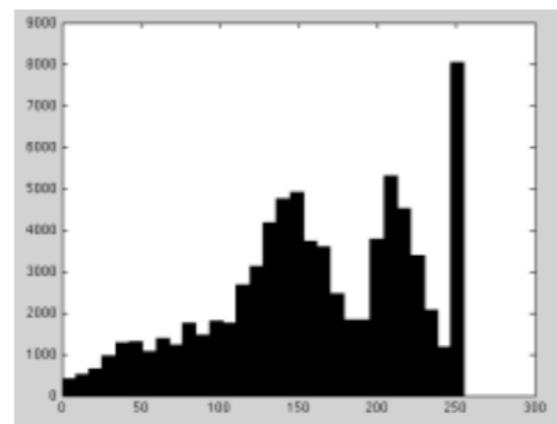
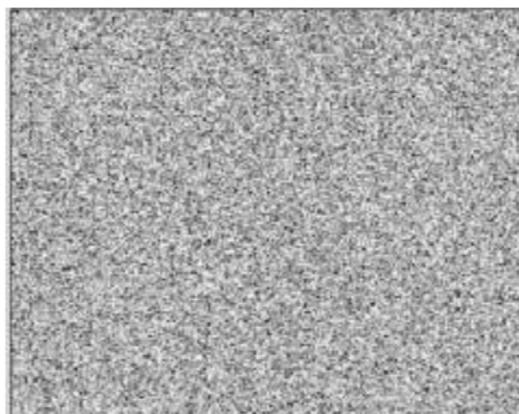
- ❖ 直方图表示的优点
 - ❖ 不再对对齐 (alignment) 、尺度变化、全局旋转等敏感



相似的颜色直方图（归一化后）

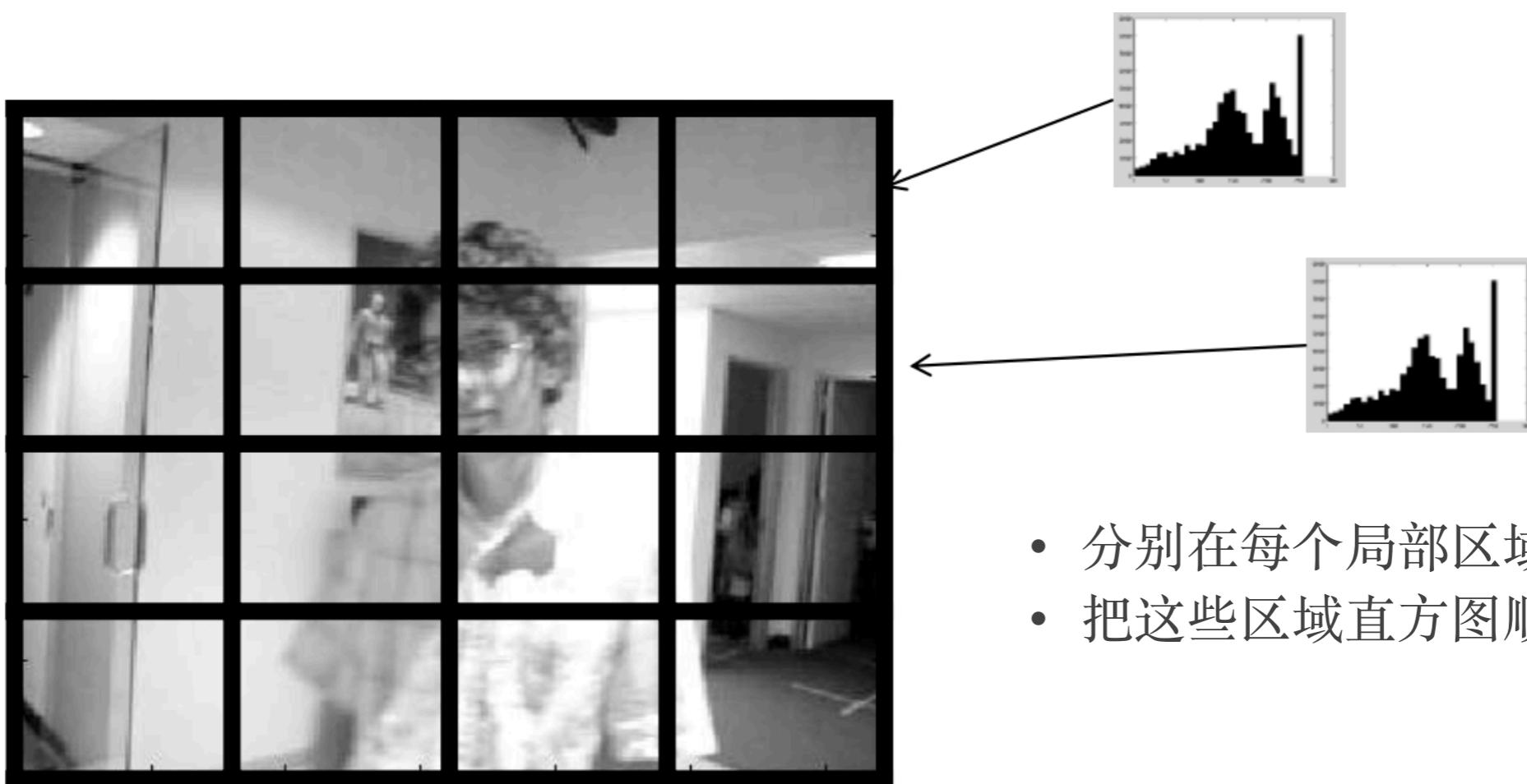
如何表示一张图像？

- ❖ 全局直方图表示的局限
 - ❖ 空间位置信息缺失



如何表示一张图像？

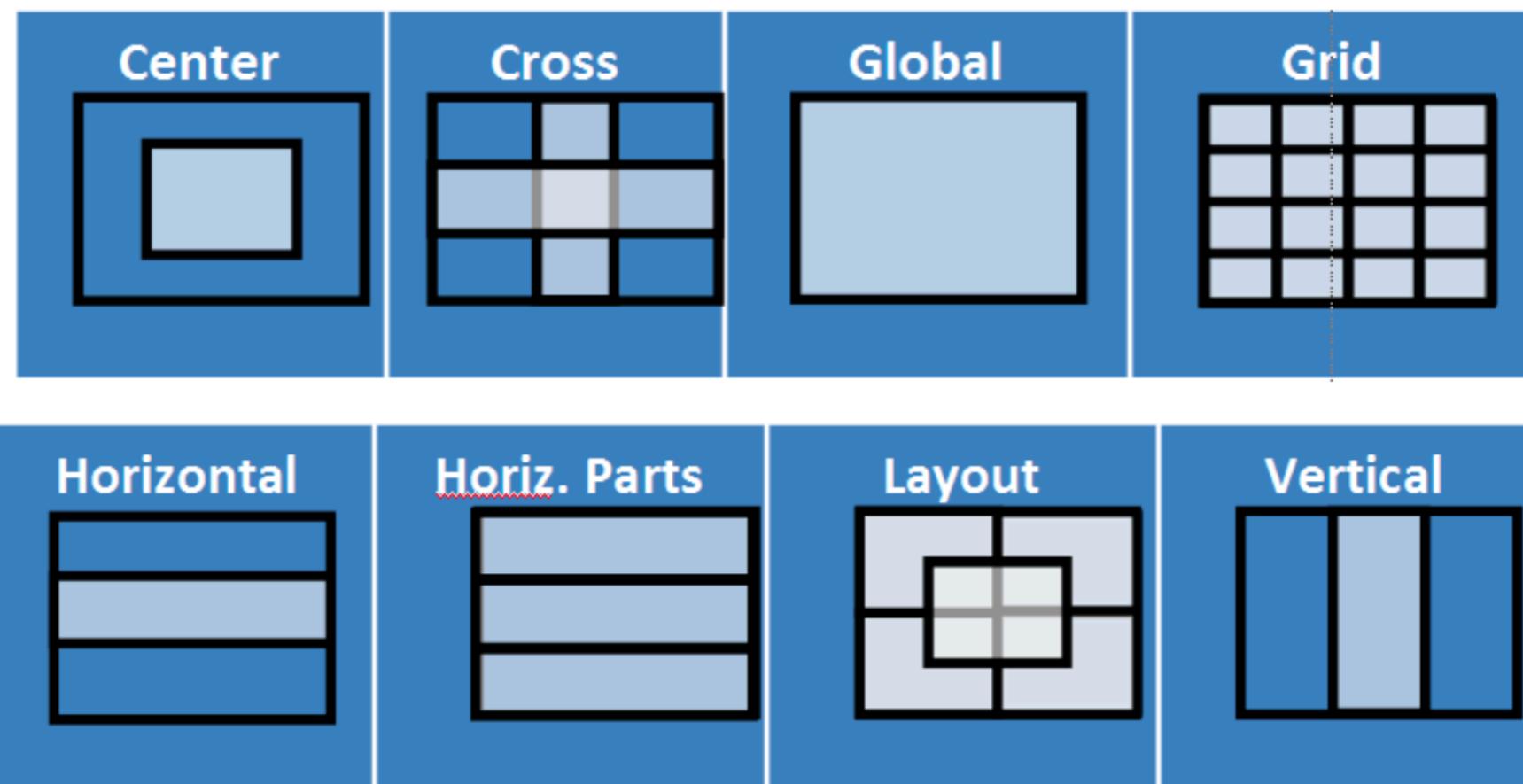
- ❖ 为了克服全局直方图表示的局限
 - ❖ 引入空间位置信息



如何表示一张图像？

- ❖ 为了克服全局直方图表示的局限
 - ❖ 引入空间位置信息

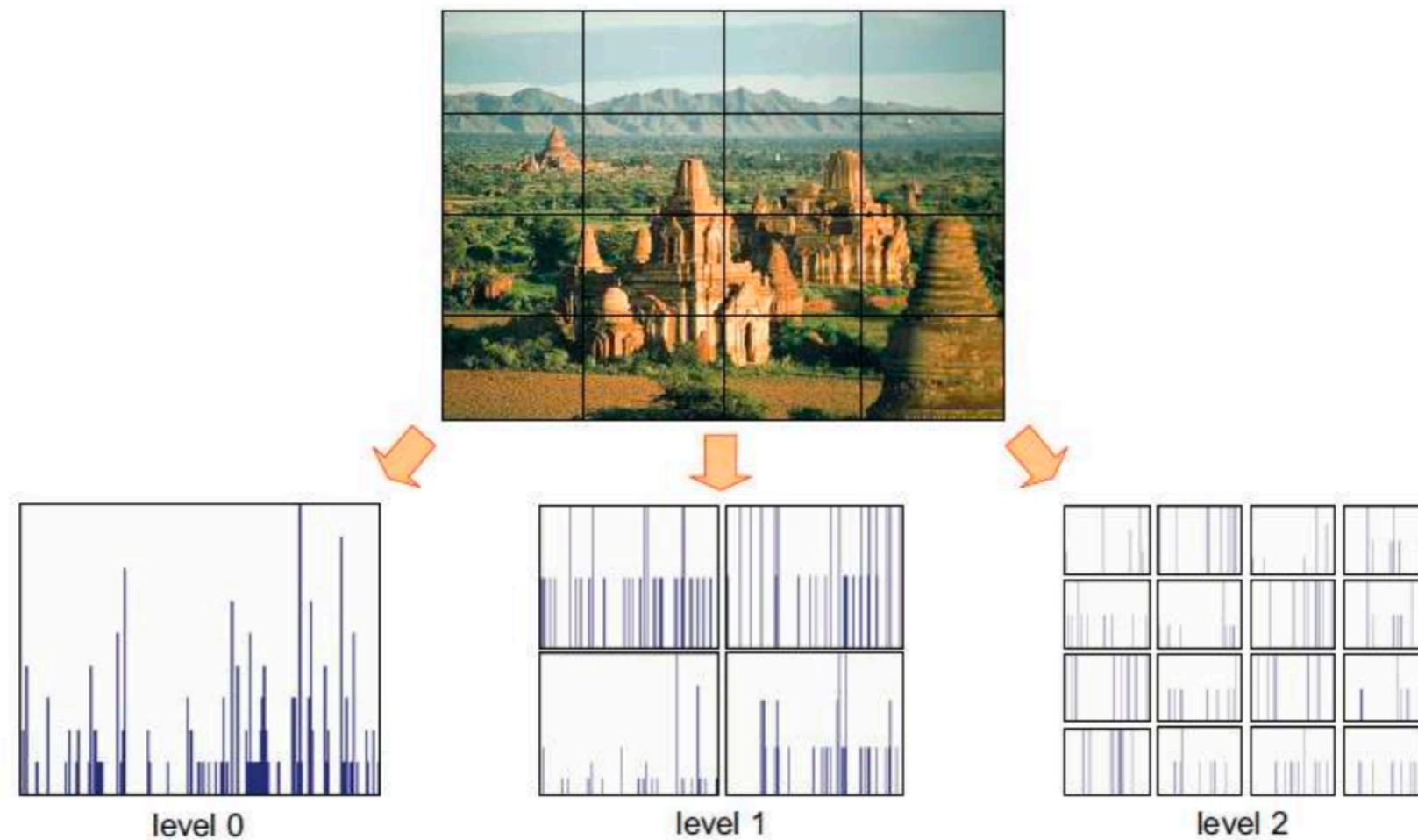
IBM IMARS Spatial Gridding



如何表示一张图像？

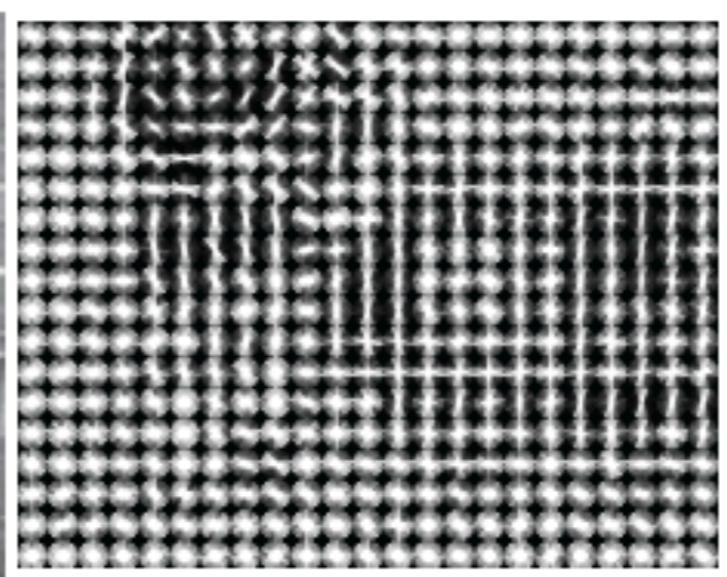
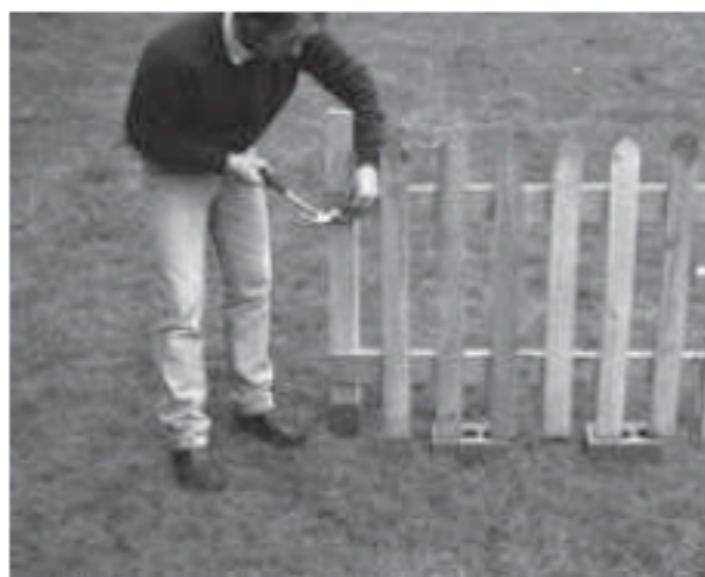
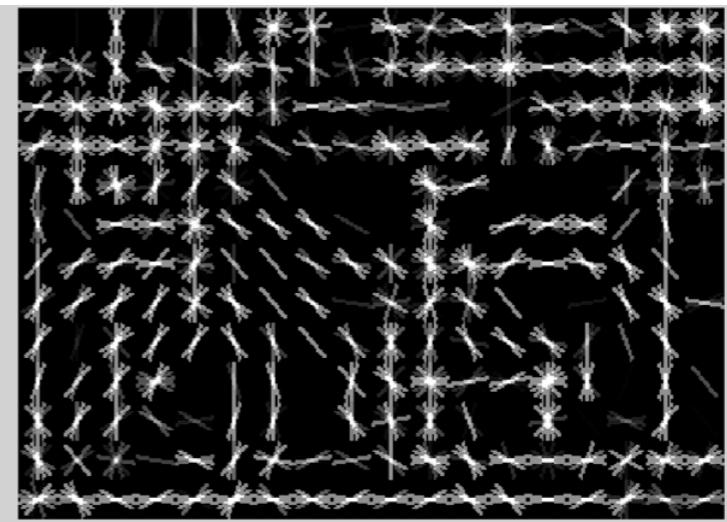
- ❖ 为了克服全局直方图表示的局限，引入空间位置信息
- ❖ 空间金字塔（Spatial Pyramid）

Lazebnik, Schmid and Ponce, CVPR'06



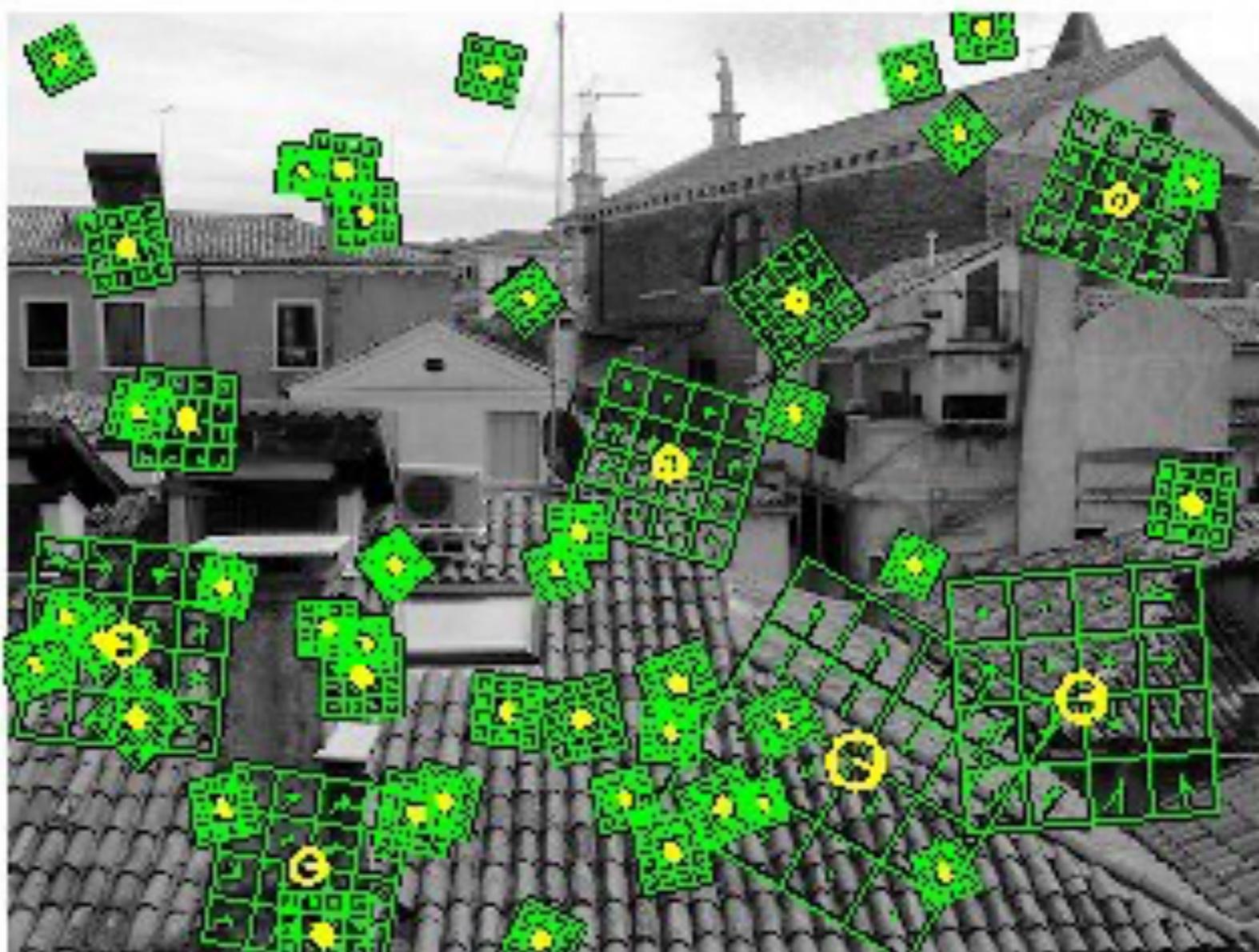
如何表示一张图像？

- ❖ 比元像素（颜色）更好的特征：梯度
- ❖ HOG



如何表示一张图像？

- ❖ 如果是稀疏采样图像块的情形，如何表示？



词袋模型 (Bag of Words)

2007-01-23: State of the Union Address

George W. Bush (2001-)

abandon accountable affordable afghanistan africa aided ally anbar armed army **baghdad** bless **challenges** chamber chaos
choices civilians coalition commanders **commitment** confident confront congressman constitution corps debates deduction
deficit deliver **democratic** deploy dikembe diplomacy disruptions earmarks **economy** einstein **elections** eliminates
expand **extremists** failing faithful families **freedom** fuel **funding** god haven ideology immigration impose
insurgents iran **iraq** islam julie lebanon love madam marine math medicare moderation neighborhoods nuclear offensive
palestinian payroll province pursuing **qaeda** radical **regimes** resolve retreat rieman sacrifices science sectarian senate
september **shia** stays strength students succeed sunni **tax** territories **terrorists** threats uphold victory
violence violent **war** washington weapons wesley

顺序无关的文本表示：字典中词汇出现的频率

词袋模型 (Bag of Words)

1962-10-22: Soviet Missiles in Cuba

John F. Kennedy (1961-63)

abandon achieving adversaries aggression agricultural appropriate armaments **arms** assessments atlantic ballistic berlin
buildup burdens cargo college commitment communist constitution consumers cooperation crisis **cuba** dangers
declined **defensive** deficit depended disarmament divisions domination doubled **economic** education
elimination emergence endangered equals **europe** expand exports fact false family forum **freedom** fulfill gromyko
halt hazards **hemisphere** hospitals ideals independent industries inflation labor latin limiting minister **missiles**
modernization neglect **nuclear** oas obligation observer **offensive** peril pledged predicted purchasing quarantine quote
recession rejection republics retaliatory safeguard sites solution **soviet** space spur stability standby **strength**
surveillance tax territory treaty undertakings **unemployment** war warhead **weapons** welfare western widen withdraw

顺序无关的文本表示：字典中词汇出现的频率

词袋模型 (Bag of Words)

1941-12-08: Request for a Declaration of War

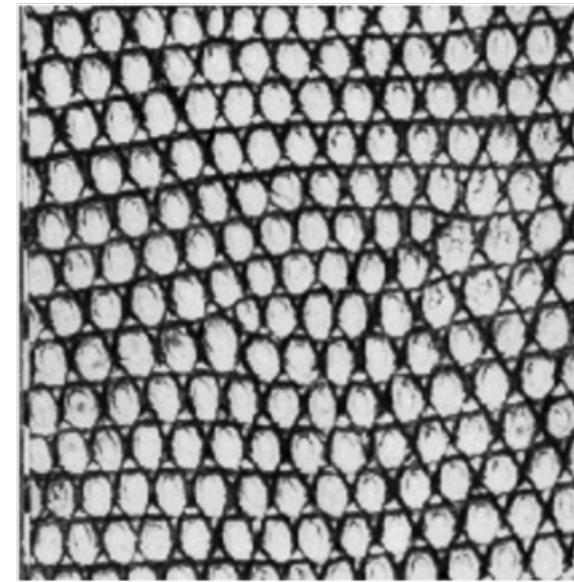
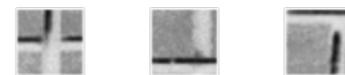
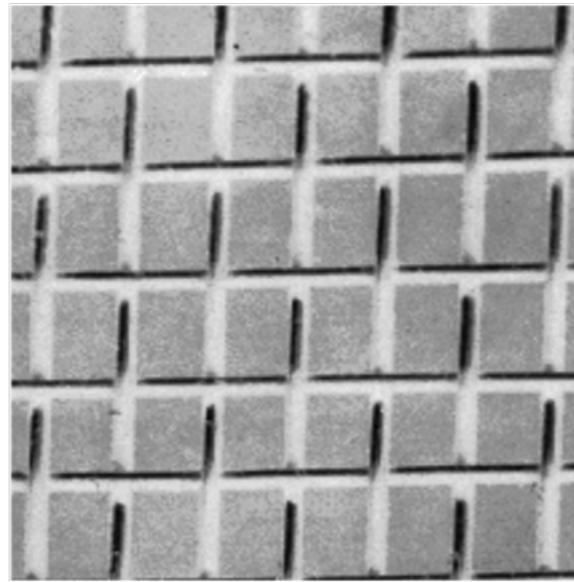
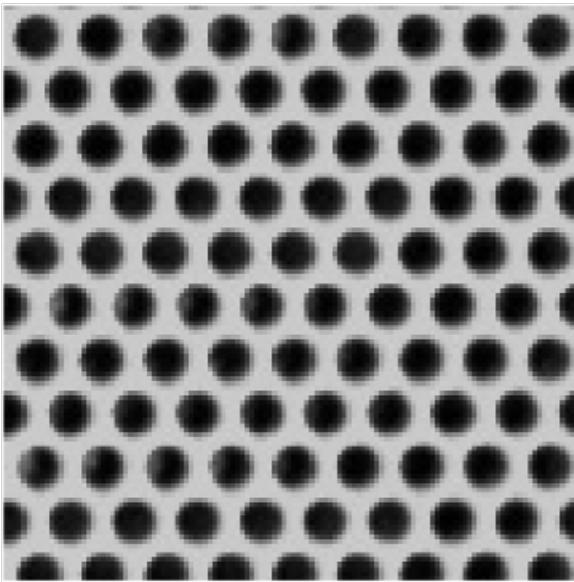
Franklin D. Roosevelt (1933-45)

abandoning acknowledge aggression aggressors airplanes armaments armed army assault assembly authorizations bombing
britain british cheerfully claiming constitution curtail december defeats defending delays democratic dictators disclose
economic empire endanger facts false forgotten fortunes france freedom fulfilled fullness fundamental gangsters
german germany god guam harbor hawaii hemisphere hint hitler hostilities immune improving indies innumerable
invasion islands isolate japanesee labor metals midst midway navy nazis obligation offensive
officially pacific partisanship patriotism pearl peril perpetrated perpetual philippine preservation privilege reject
repaired resisting retain revealing rumors seas soldiers speaks speedy stamina strength sunday sunk supremacy tanks taxes
treachery true tyranny undertaken victory war wartime washington

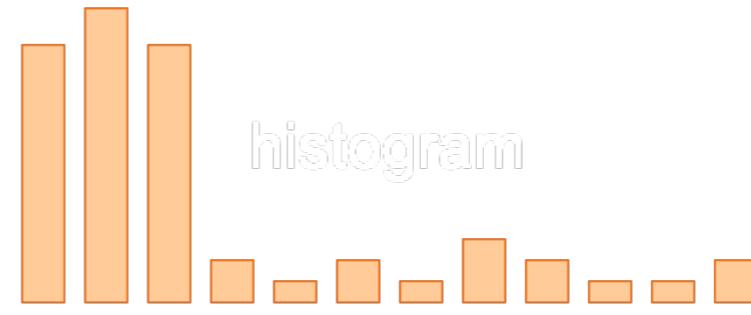
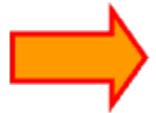
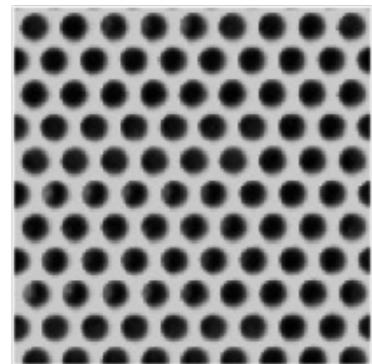
顺序无关的文本表示：字典中词汇出现的频率

纹理识别

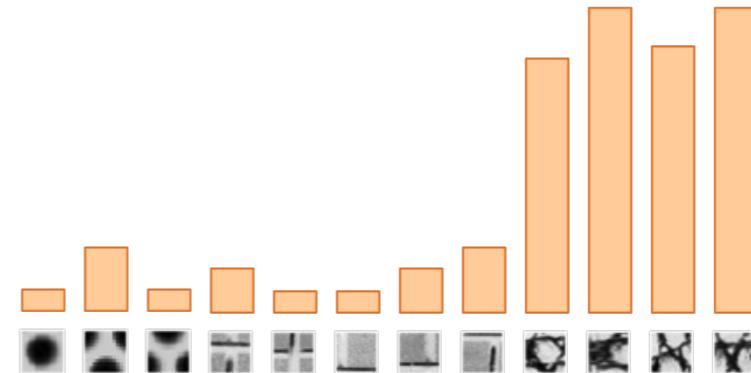
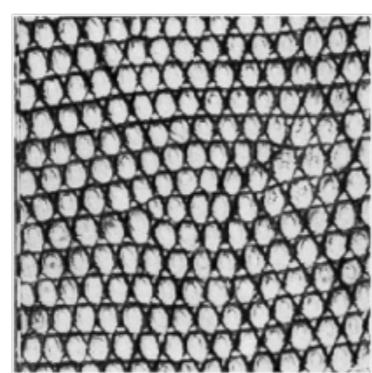
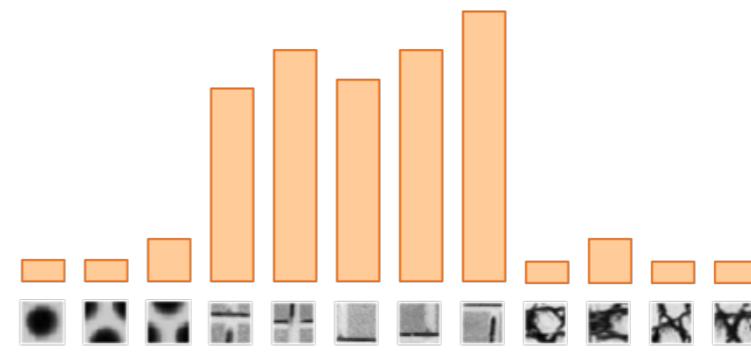
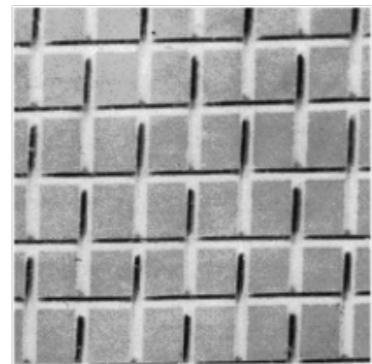
❖ 纹理元 (textons)



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

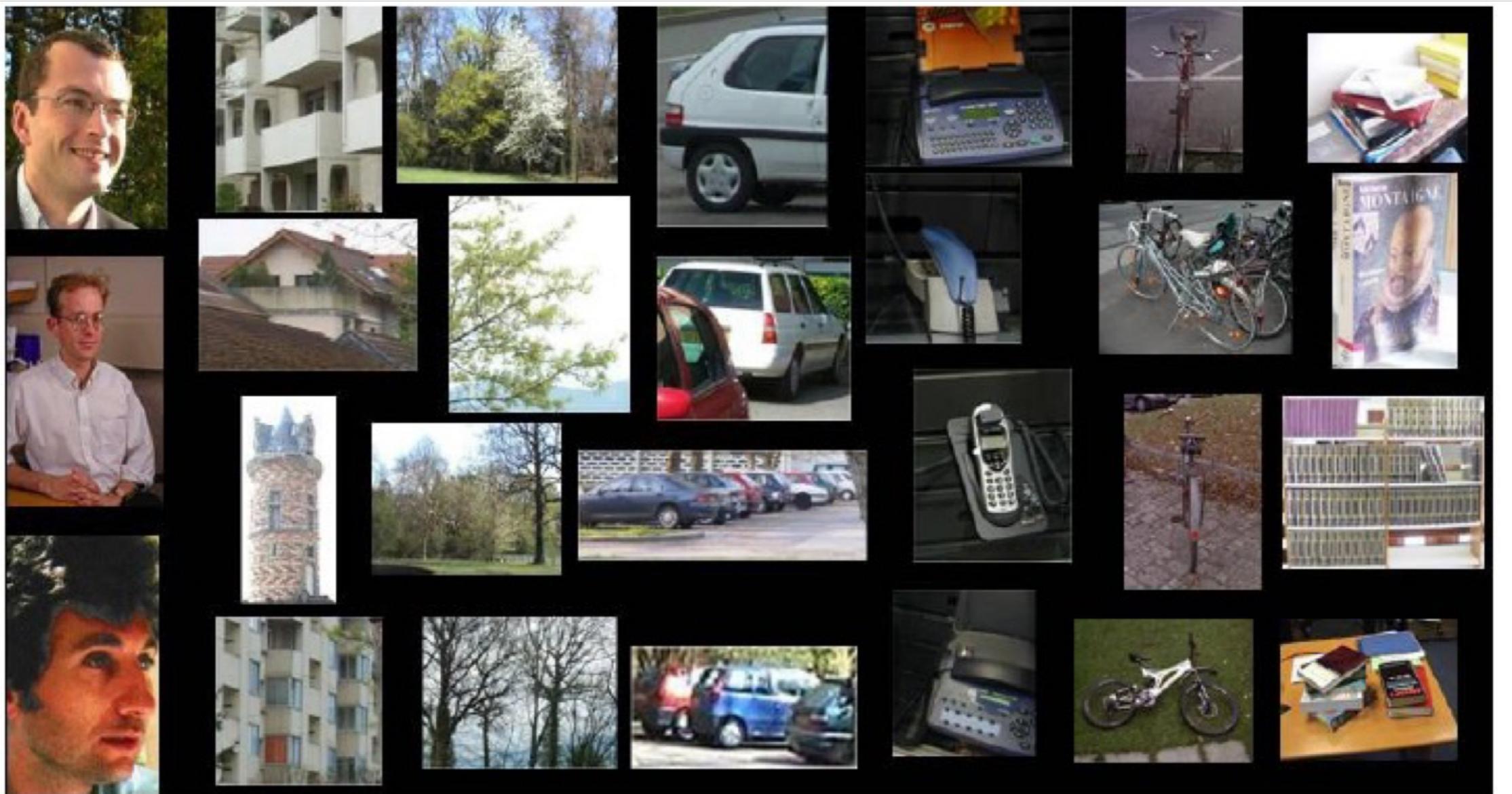


Universal texton dictionary



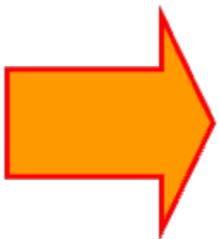
Julesz, 1981; Cula & Dana, 2001; Leung & Malik 2001; Mori, Belongie & Malik, 2001;
Schmid 2001; Varma & Zisserman, 2002, 2003; Lazebnik, Schmid & Ponce, 2003

视觉词袋模型



Our in-house database contains 1776 images in seven classes¹: faces, buildings, trees, cars, phones, bikes and books. Fig. 2 shows some examples from this dataset.

视觉词袋模型



视觉词袋模型：步骤

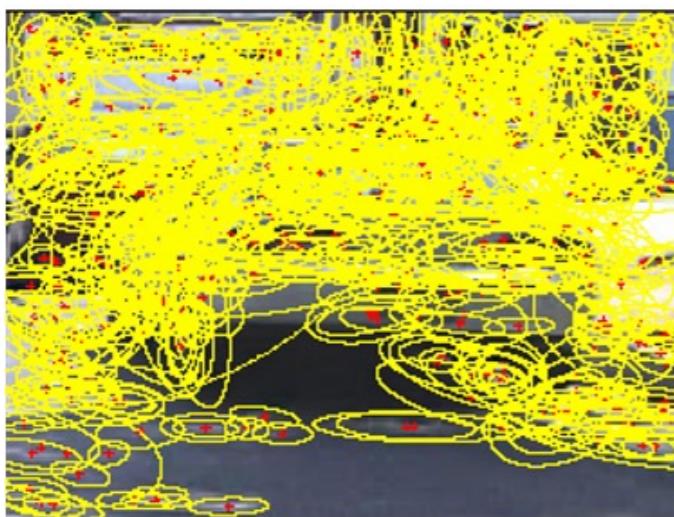
1. 视觉特征提取
2. 构建视觉词典
3. 特征矢量量化
4. 视觉词频表示

视觉词袋模型：步骤

1. 视觉特征提取：规则化的网格或特征点



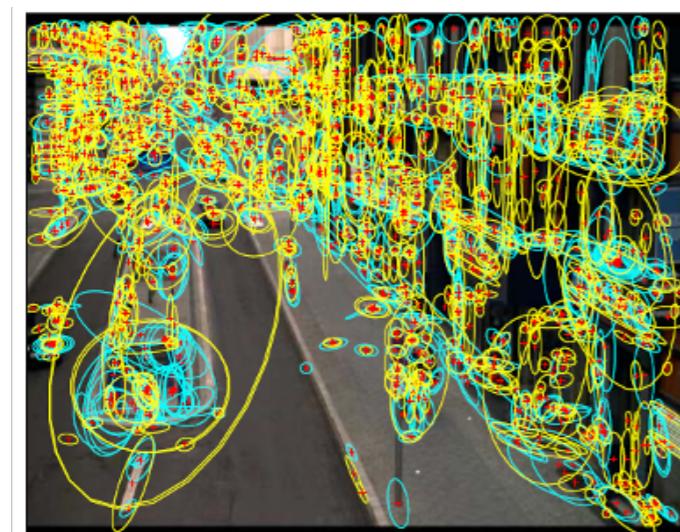
视觉词袋模型：采样策略



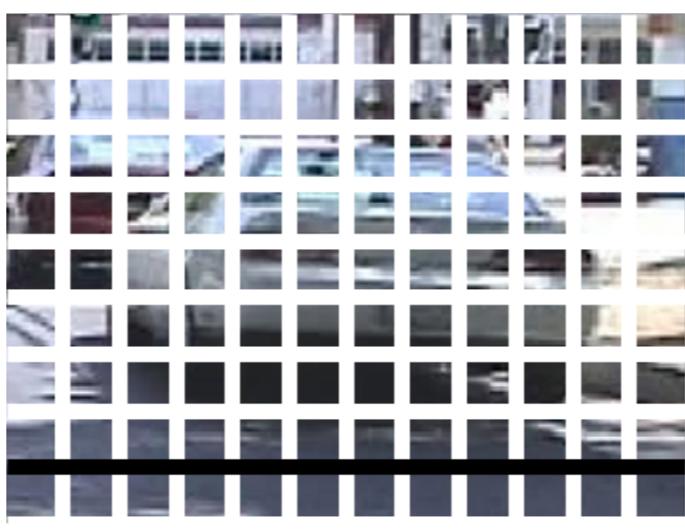
兴趣点采样



随机采样



多种兴趣点采样

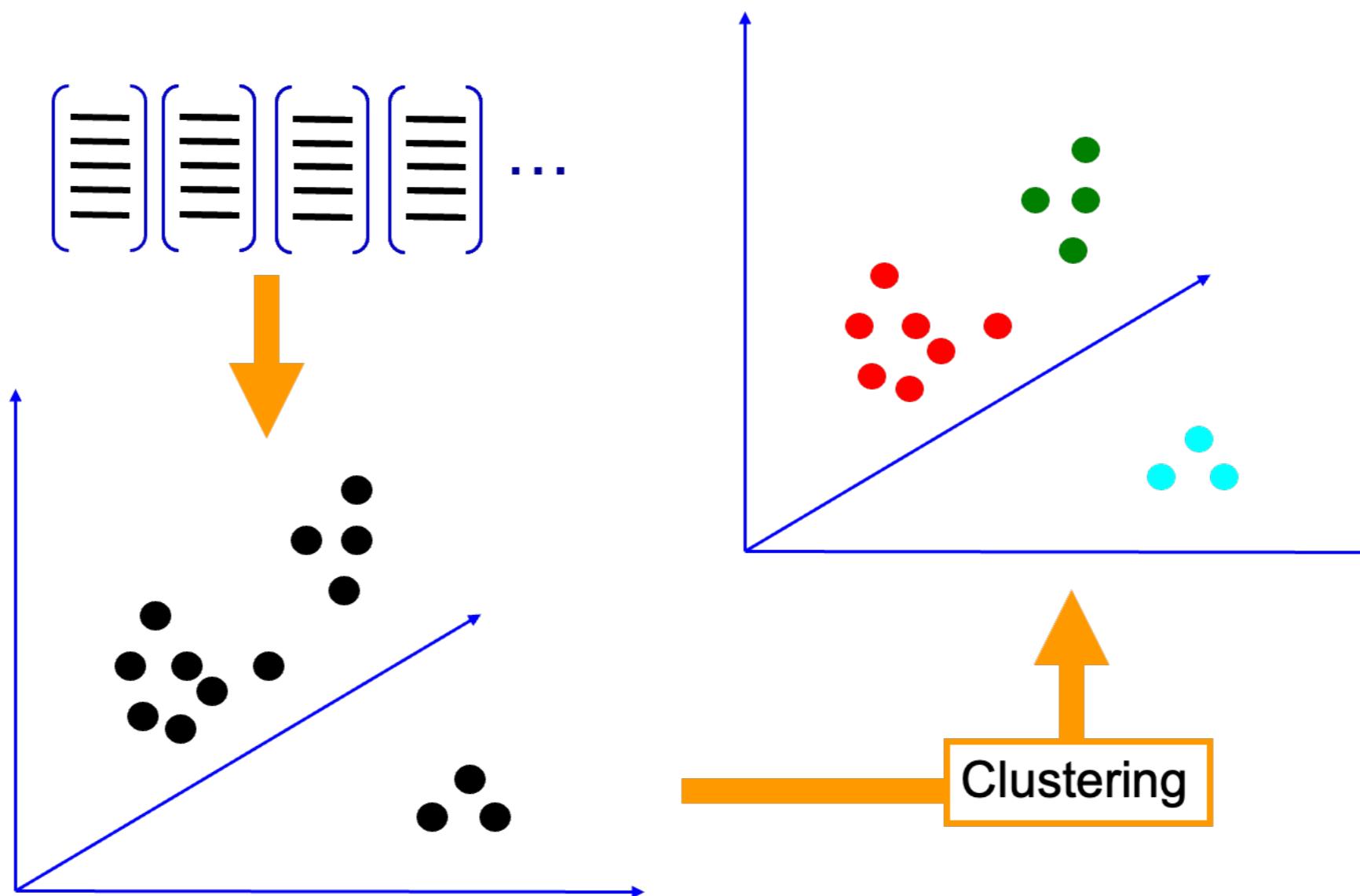


密集、均匀采样

1. 兴趣点采样对于识别具有特别纹理的物体更可靠
2. 多兴趣点可有效互补
3. 密集采样对于场景、物体分类具有更好的覆盖性

视觉词袋模型：步骤

2. 字典学习



视觉词袋模型：步骤

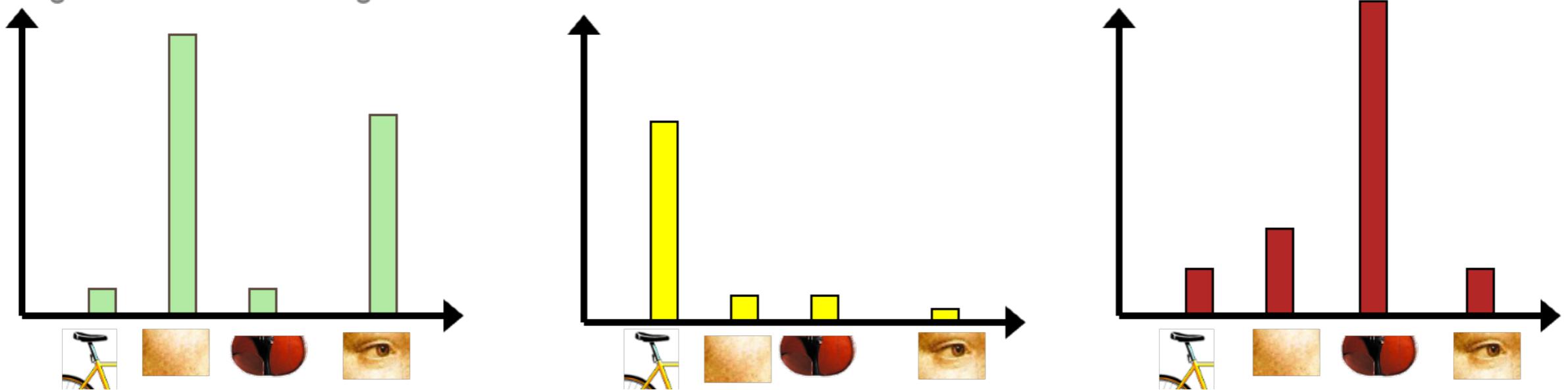
3. 矢量量化：对任何特征向量，可以在字典中找到一个与其最接近的字，则该字就是此特征向量的量化值
4. 对所有的特征，进行矢量量化，然后统计“视觉字”频率，形成直方图，直方图的维度就是字典的容量（包含有多少个字）



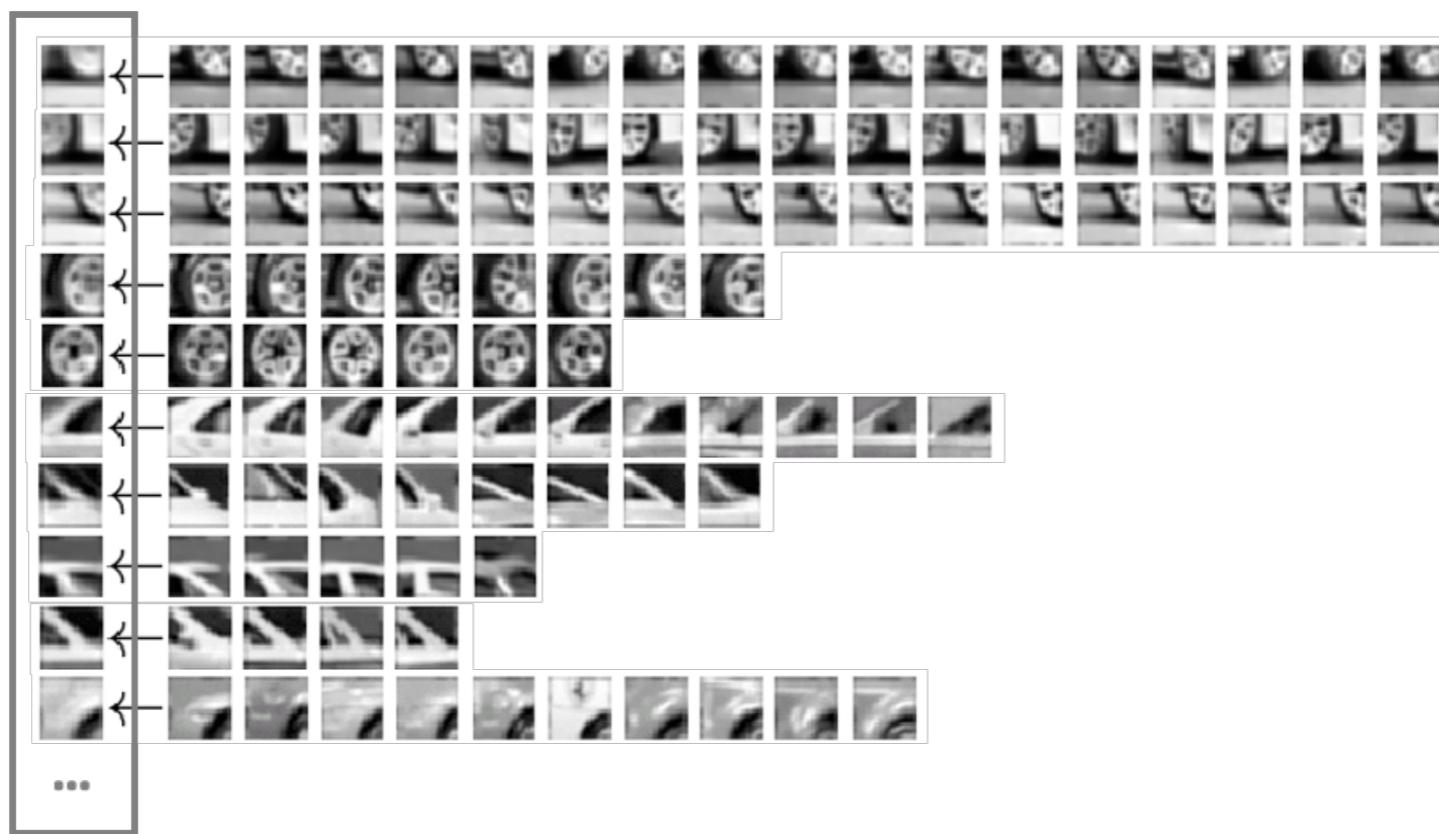
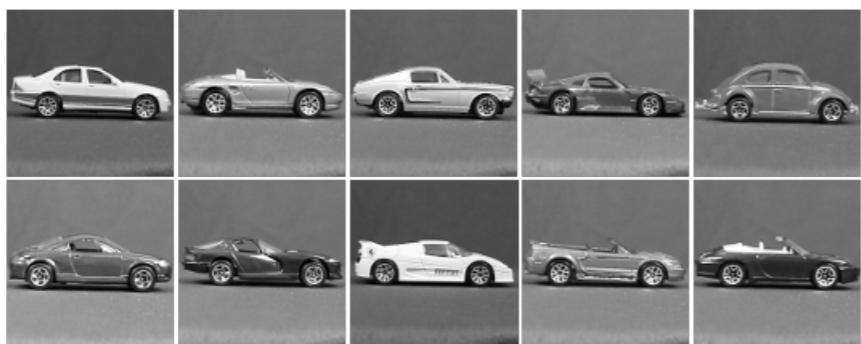
Visual words



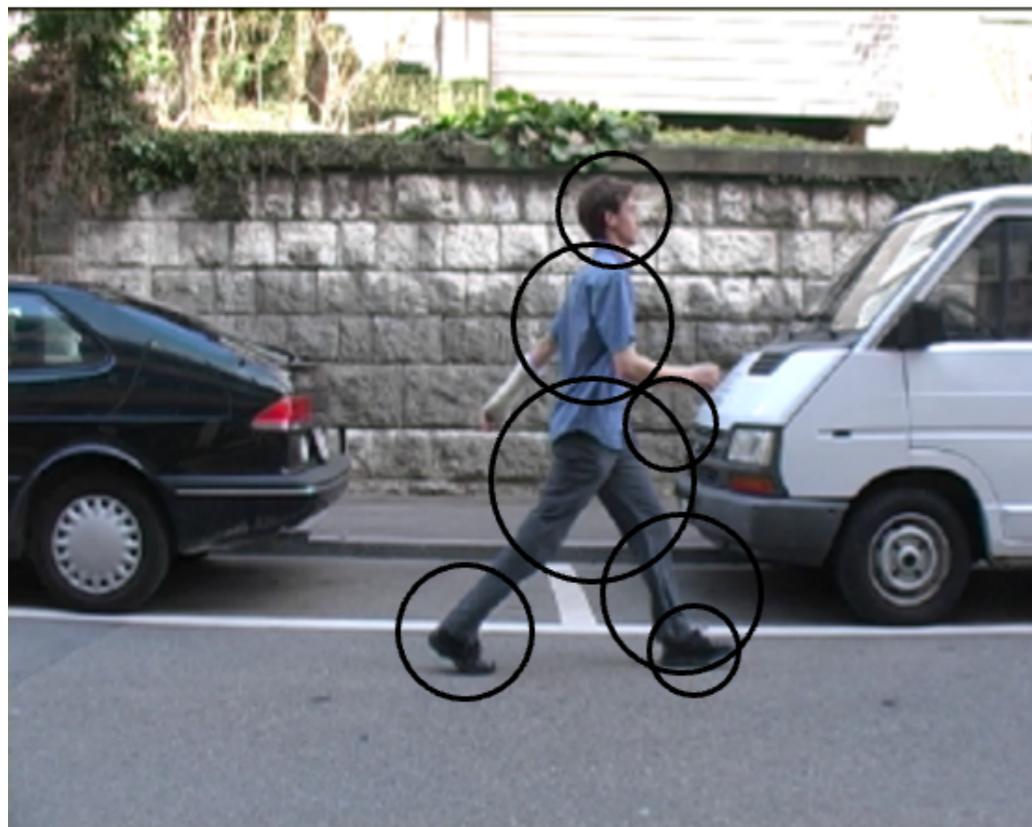
Bag of visual words histograms



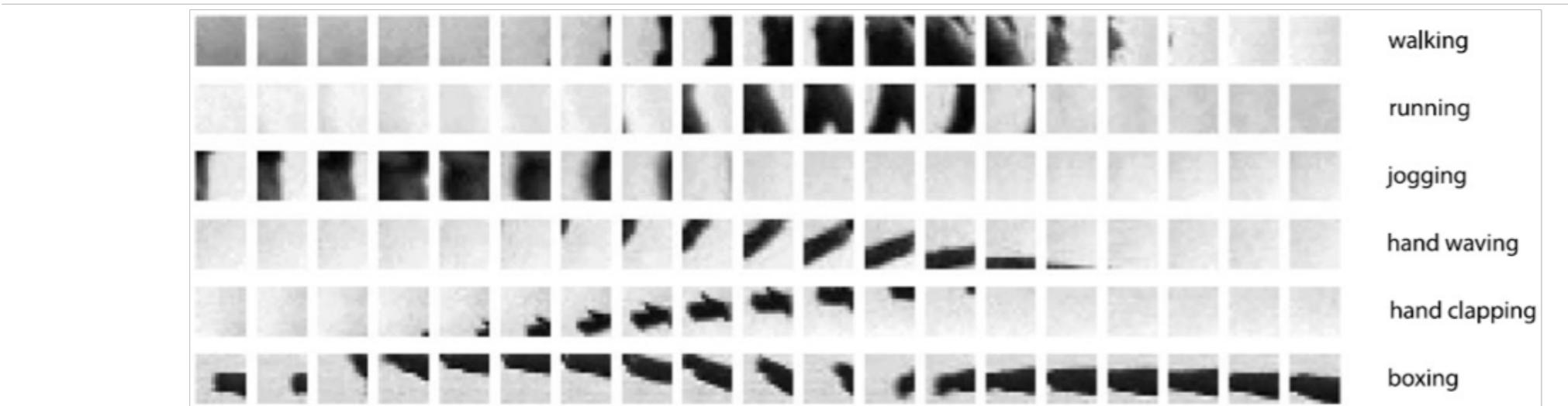
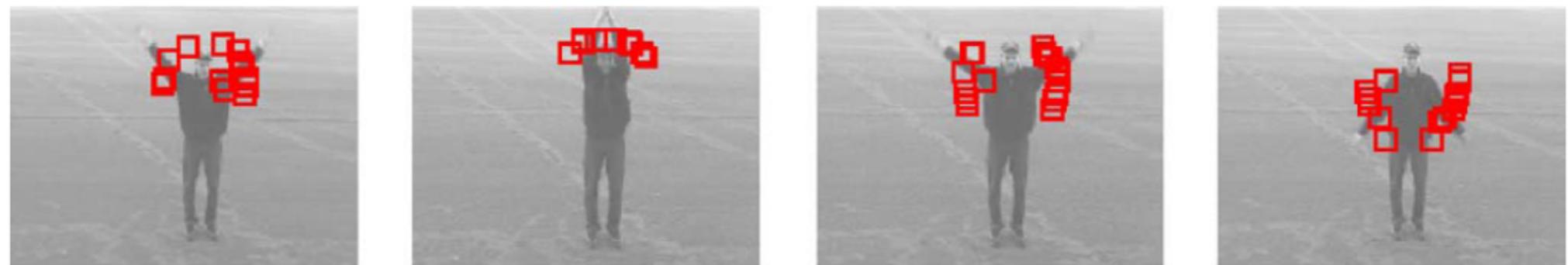
字典举例



字典举例



字典举例



视觉词袋表示：优缺点

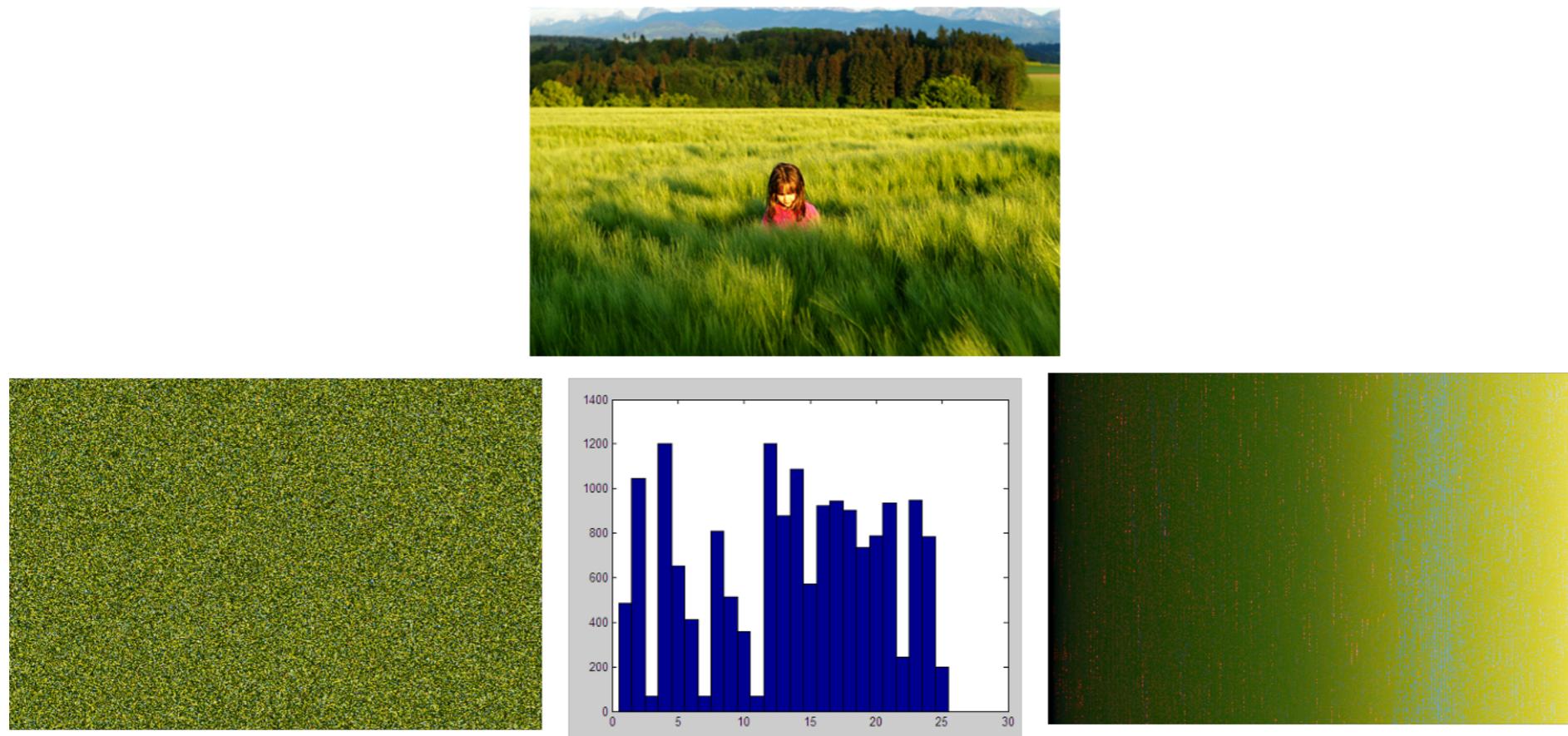
❖ 优点

1. 对几何、视角、变形具有鲁棒性
2. 对图像整体内容的紧凑总结性表示
3. 特征维度固定
4. 效果较好

❖ 缺点

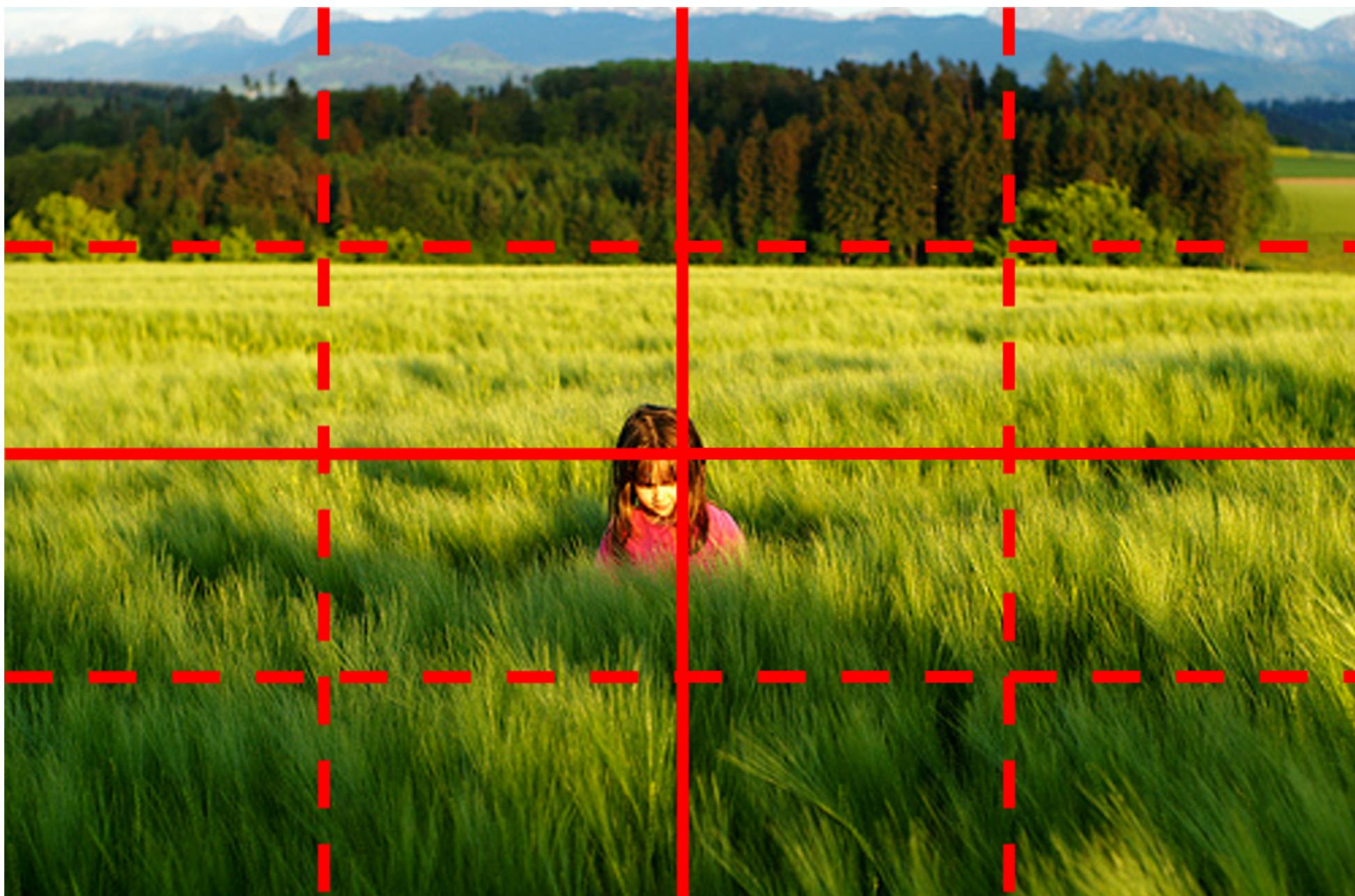
1. 字典大小？
2. 忽略了空间位置关系

升级版BoW (空间金字塔) : 动机

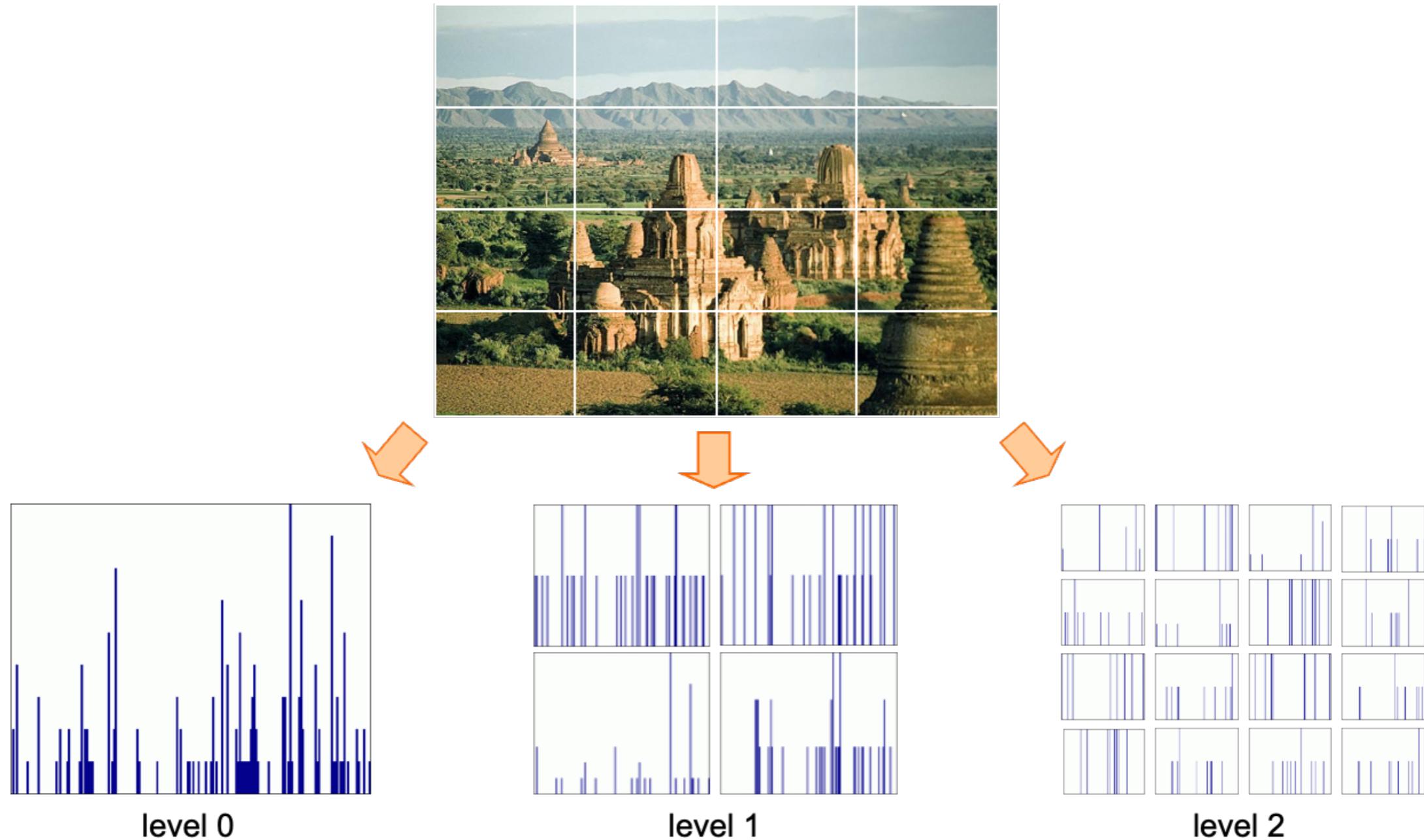


三张不同的图像，相似的直方图

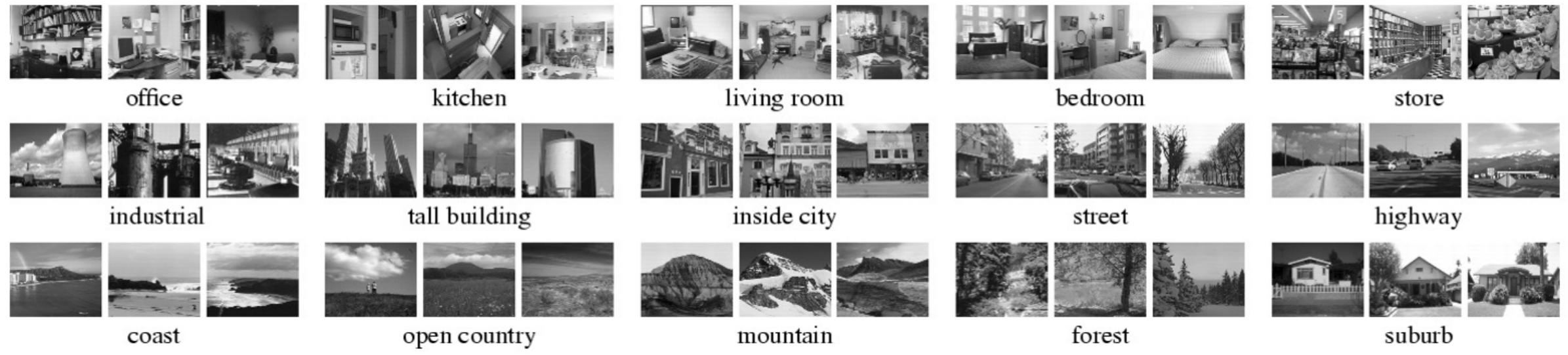
升级版BoW (空间金字塔)



升级版BoW (空间金字塔)



Scene category dataset

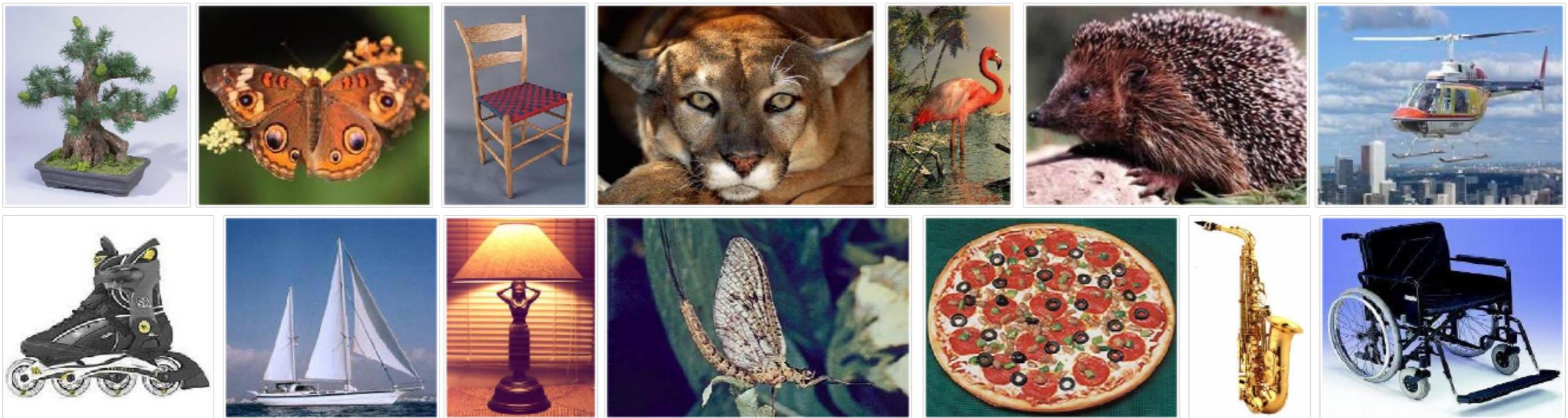


Multi-class classification results
(100 training images per class)

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

Caltech101 dataset

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



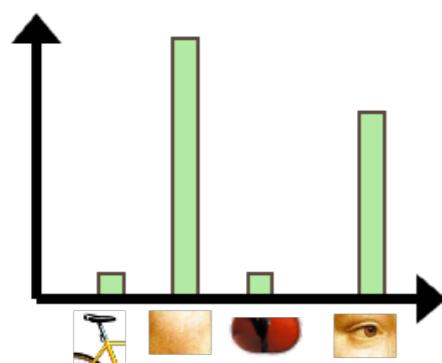
Multi-class classification results (30 training images per class)

	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

视觉词袋模型的应用：图像检索

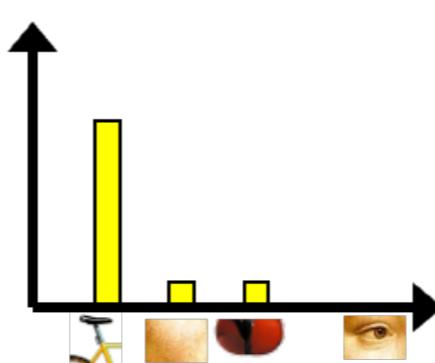
Database image

$$\vec{d}_j = [1 \quad 8 \quad 1 \quad 4]$$



Query

$$\vec{q} = [5 \quad 1 \quad 1 \quad 0]$$



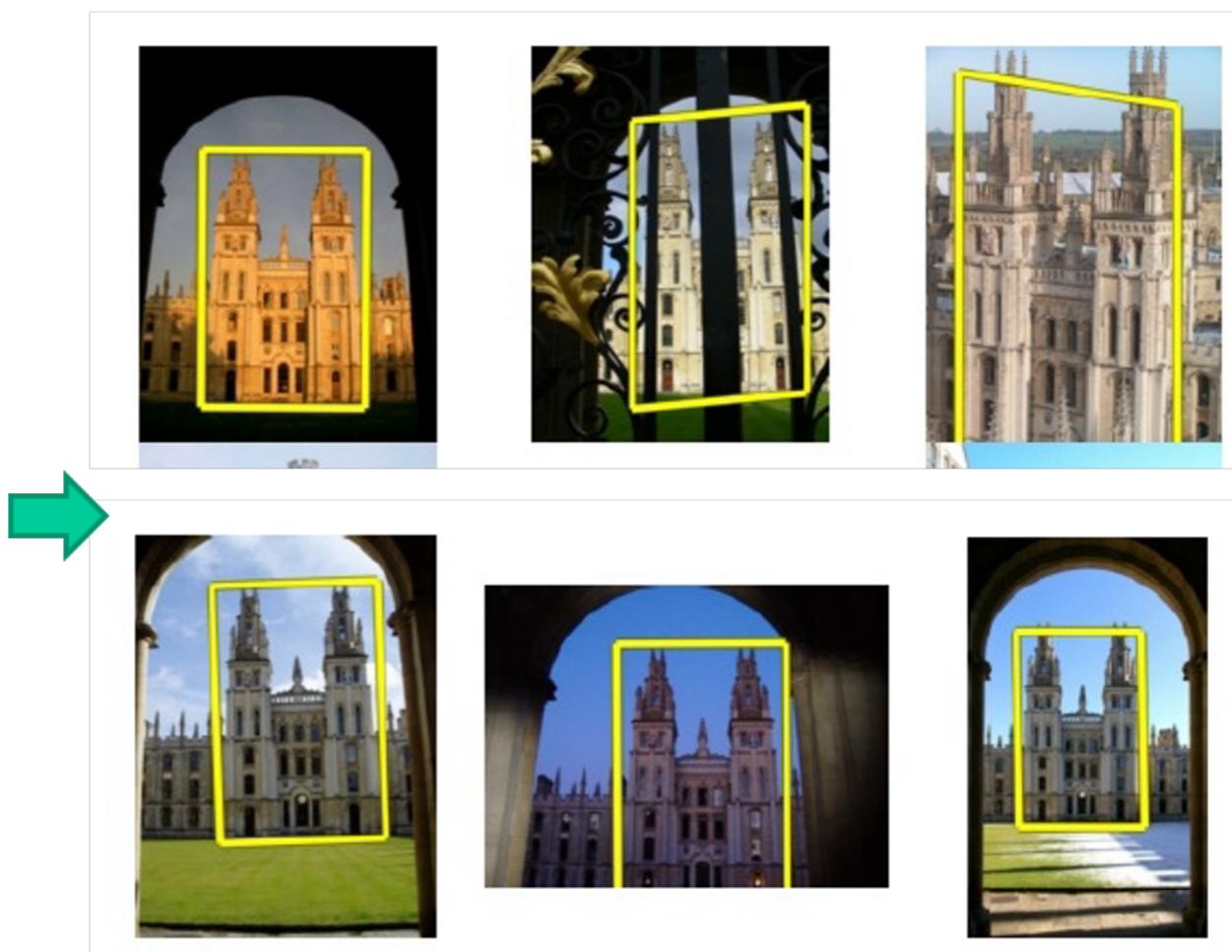
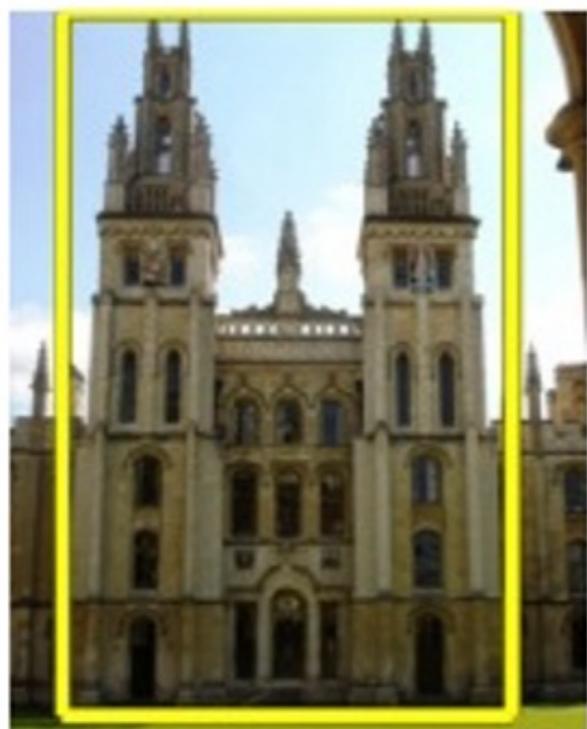
$$sim(d_j, q) = \frac{\langle d_j, q \rangle}{\|d_j\| \|q\|}$$

$$= \frac{\sum_{i=1}^V d_j(i) \times q(i)}{\sqrt{\sum_{i=1}^V d_j(i)^2} \times \sqrt{\sum_{i=1}^V q(i)^2}}$$

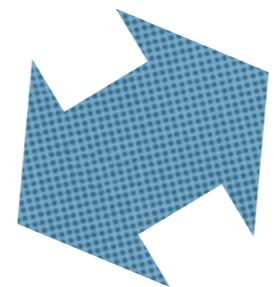
for vocabulary of V words

余弦距离：忽略绝对数量的差别，关注
“出现”而不是数量

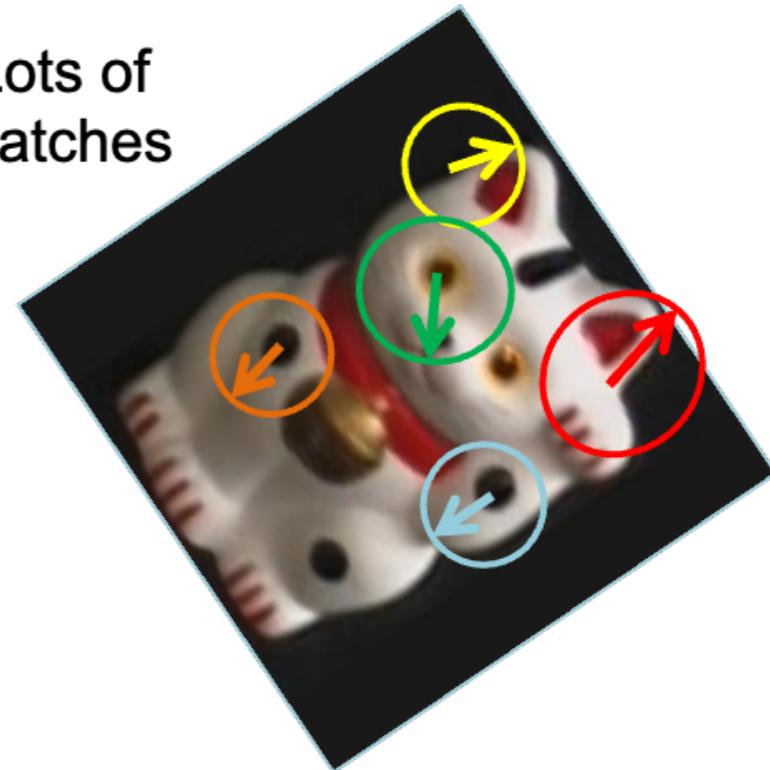
视觉词袋模型的应用：图像检索



视觉词袋模型的应用：图像检索



Lots of
Matches



Few or No
Matches

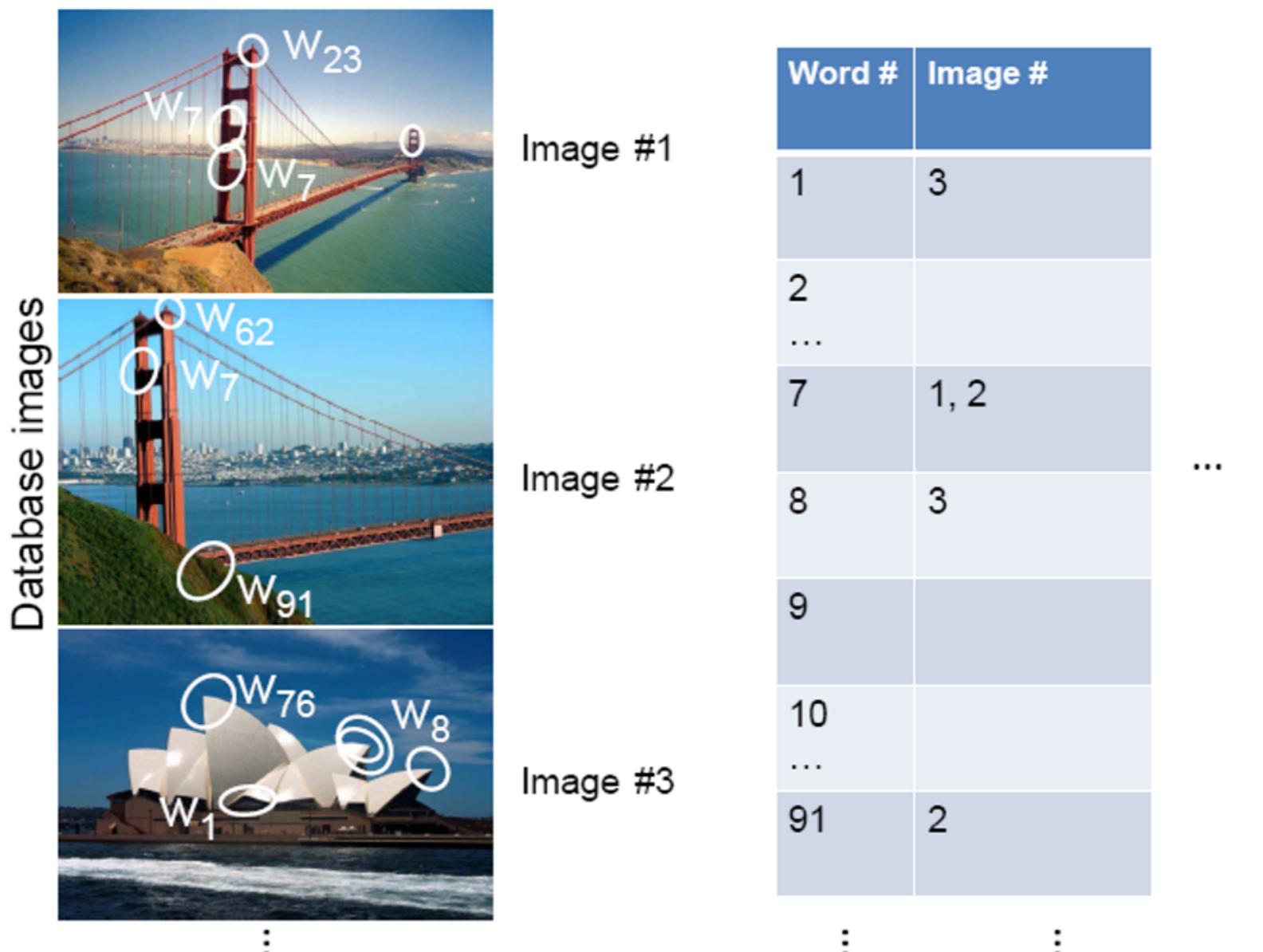


视觉词袋模型的应用：图像检索

Index		
"Along I-75," From Detroit to Florida; <i>inside back cover</i>	Butterfly Center, McGuire; 134	Driving Lanes; 85
"Drive I-95," From Boston to Florida; <i>inside back cover</i>	CAA (see AAA)	Duval County; 163
1929 Spanish Trail Roadway; 101-102,104	CCC, The; 111,113,115,135,142	Eau Gallie; 175
511 Traffic Information; 83	Ca d'Zan; 147	Edison, Thomas; 152
A1A (Barrier Isl) - I-95 Access; 86	Caloosahatchee River; 152	Eglin AFB; 116-118
AAA (and CAA); 83	Name; 150	Eight Reale; 176
AAA National Office; 88	Canaveral Natnl Seashore; 173	Ellenton; 144-145
Abbreviations,	Cannon Creek Airpark; 130	Emanuel Point Wreck; 120
Colored 25 mile Maps; cover	Canopy Road; 106,160	Emergency Callboxes; 83
Exit Services; 196	Cape Canaveral; 174	Epiphytes; 142,148,157,159
Travelogue; 85	Castillo San Marcos; 169	Escambia Bay; 119
Africa; 177	Cave Diving; 131	Bridge (I-10); 119
Agricultural Inspection Strns; 126	Cayo Costa, Name; 150	County; 120
Ah-Tah-Thi-Ki Museum; 160	Celebration; 93	Estero; 153
Air Conditioning, First; 112	Charlotte County; 149	Everglade; 90,95,139-140,154-160
Alabama; 124	Charlotte Harbor; 150	Draining of; 156,181
Alachua; 132	Chautauqua; 116	Wildlife MA; 160
County; 131	Chipley; 114	Wonder Gardens; 154
Alafia River; 143	Name; 115	Falling Waters SP; 115
Alapaha, Name; 126	Choctawatchee, Name; 115	Fantasy of Flight; 95
Alfred B Maclay Gardens; 106	Circus Museum, Ringling; 147	Fayer Dykes SP; 171
Alligator Alley; 154-155	Citrus; 88,97,130,136,140,180	Fires, Forest; 166
Alligator Farm, St Augustine; 169	CityPlace, W Palm Beach; 180	Fires, Prescribed; 148
Alligator Hole (definition); 157	City Maps,	Fisherman's Village; 151
Alligator, Buddy; 155	Ft Lauderdale Expwy; 194-195	Flagler County; 171
Alligators; 100,135,138,147,156	Jacksonville; 163	Flagler, Henry; 97,165,167,171
Anastasia Island; 170	Kissimmee Expwy; 192-193	Florida Aquarium; 186
Anhala; 108,109,146	Miami Expressways; 194-195	Florida,
Apalachicola River; 112	Orlando Expressways; 192-193	12,000 years ago; 187
Appleton Mus of Art; 136	Pensacola; 26	Cavern SP; 114
Aquifer; 102	Tallahassee; 191	Map of all Expressways; 2-3
Arabian Nights; 94	Tampa-St. Petersburg; 63	Mus of Natural History; 134
Art Museum, Ringling; 147	St. Augsutine; 191	National Cemetery ; 141
Aruba Beach Cafe; 183	Civil War; 100,108,127,138,141	Part of Africa; 177
Aucilla River Project; 106	Clearwater Marine Aquarium; 187	Platform; 187
Babcock-Web WMA; 151	Collier County; 154	Sheriff's Boys Camp; 126
Bahia Mar Marina; 184	Colonial Spanish Quarters; 168	Sports Hall of Fame; 130
Baker County; 99	Columbia County; 101,128	Sun 'n Fun Museum; 97
Barefoot Mallmen; 182	Coquina Building Material; 165	Supreme Court; 107
Barge Canal; 137	Corkscrew Swamp, Name; 154	Florida's Turnpike (FTP); 178,189
Bee Line Expy; 80	Cowboys; 95	25 mile Strip Maps; 66
Beltz Outlet Mall; 89	Crab Trap II; 144	Administration; 189
Bernard Castro; 136	Cracker, Florida; 88,95,132	Coin System; 190
Big "I"; 165	Crossstown Expy; 11,35,98,143	Exit Services; 189
Big Cypress; 155,158	Cuban Bread; 184	HEFT; 76,161,190
Big Foot Monster; 105	Dade Battlefield; 140	History; 189
Billie Swamp Safari; 160	Dade, Maj. Francis; 139-140,161	Names; 189
Blackwater River SP; 117	Dania Beach Hurricane; 184	Service Plazas; 190
Blue Angels	Daniel Boone, Florida Walk; 117	Spur SR91; 76
A4-C Skyhawk; 117	Daytona Beach; 172-173	Ticket System; 190
Atrium; 121	De Land; 87	Toll Plazas; 190
Blue Springs SP; 87	De Soto, Hernando,	Ford, Henry; 152
Blue Star Memorial Highway; 125	Amhaica; 108-109,146	Fort Barrancas; 122
Boca Ciega; 189	County; 149	Buried Alive; 123
Boca Grande; 150	Explorer; 146	Fort Caroline; 164
Boca Raton; 182	Landing; 146	Fort Clinch SP; 161
Bonnie Blue Flag; 124	Napitaca; 103	Fort De Soto & Egmont Key; 188
Boyd Hill Nature Trail; 188	National Park; 147	Fort Lauderdale; 161,182-184
Bradenton; 145-147	Tallahassee; 108	Fort Myers; 152-153
Breakers, The, Palm Beach; 181	DeFunik Springs; 116	Fort Pierce; 177-178
Brickell Point, Miami; 185	Name; 115	Farmers Market; 178
Britton Hill; 116	Delnor-Wiggins Pass SP; 155	Fountain of Youth; 170
Brogan Museum; 107	Denoel Cafe, St Augustine; 169	Frank Lloyd Wright Center; 97
Bromeliads (see Epiphytes)	Devil's Millhopper; 132	Gadsden County; 110
Broward County; 159,181	Dickson Azalea Park; 89	Gainesville; 99,104,131-135,146
Broward, Gov. Napoleon; 156	Dinosaur World; 98	Gamble Plantation; 145
Bulow Plantation Ruins; 171	Discovery Cove; 90	Garden of Eden; 112
Bush, Gov. Jeb; 100	Dixie Highway; 186	Gasparilla, Pirate; 150
	Don Garlits Drag Racing Mus; 138	Gatorade; 134
	Douglas, Marjory Stoneman ; 159	Gaylord Palms; 90
		Geology; 102-103,110,131-132

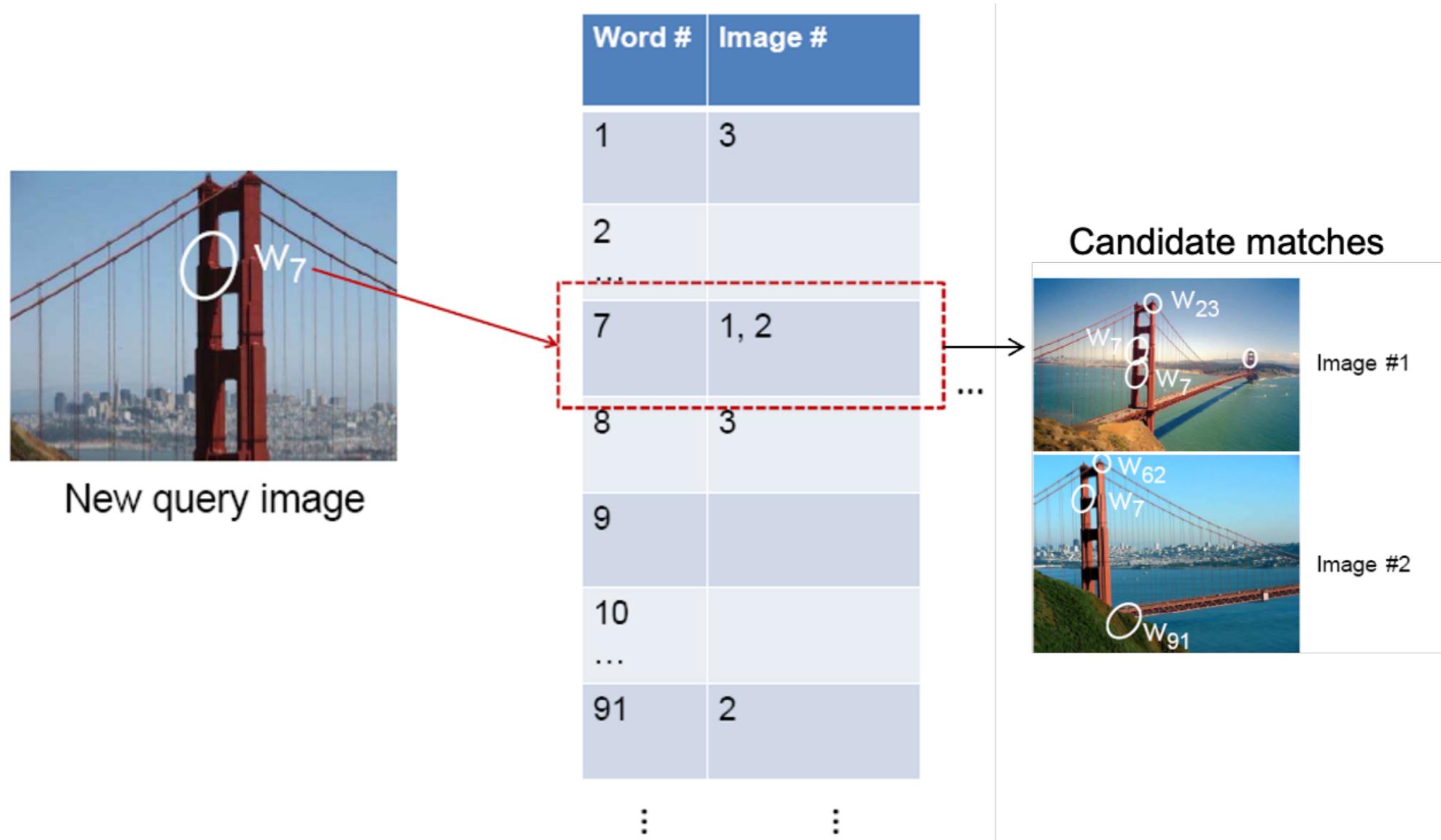
- 逆索引：某个词出现的所有页面
- 希望找到某个特征出现的所有图像

视觉词袋模型的应用：图像检索



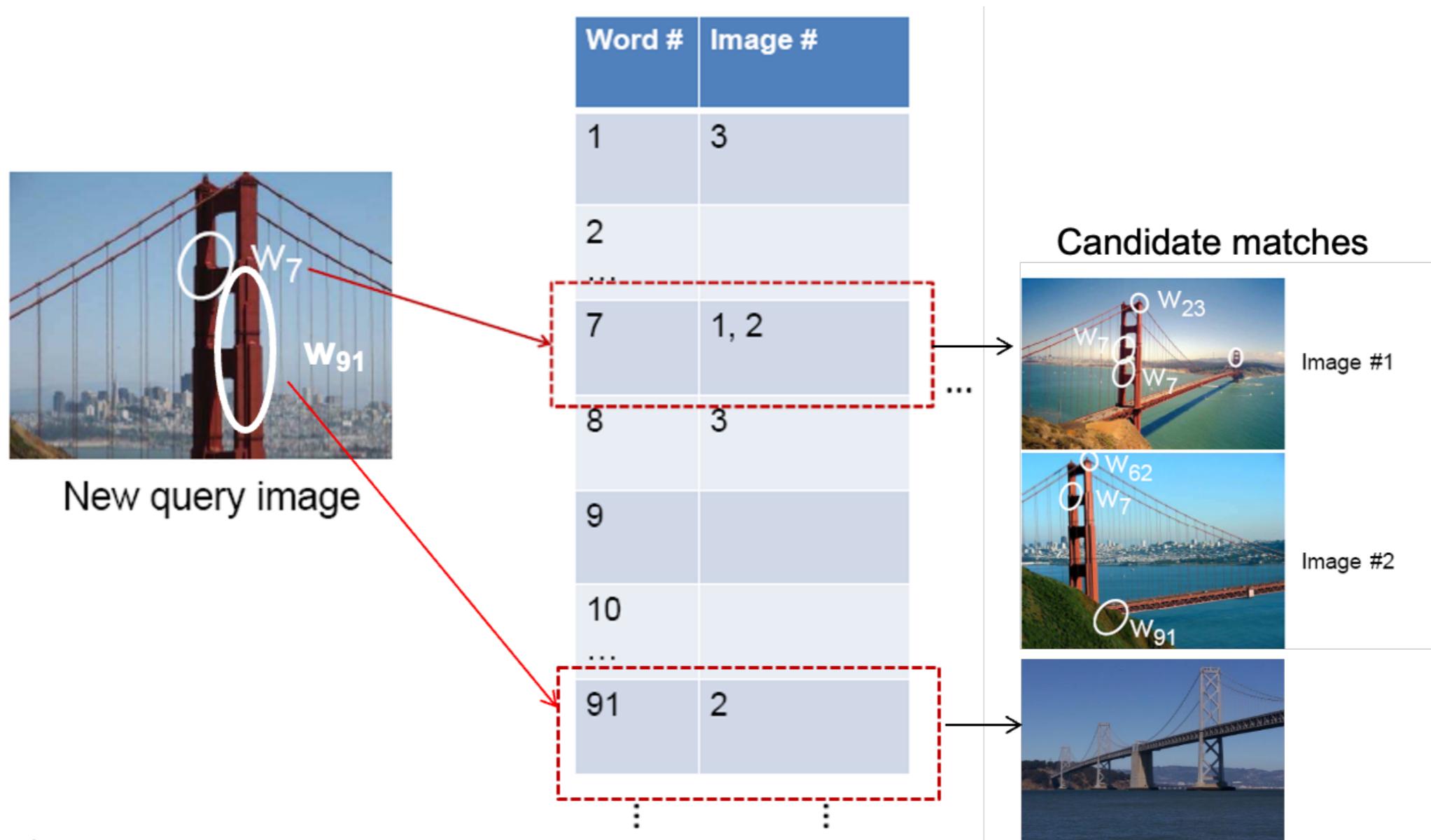
基于数据库建立逆索引

视觉词袋模型的应用：图像检索



查询逆索引

视觉词袋模型的应用：图像检索



查询逆索引，根据出现的词数排序

视觉词袋模型的应用：图像检索

- ❖ 关键点：稀疏性
- ❖ 如果大部分图像包含很多字，则计算类似贪婪搜索

视觉词袋模型的应用：图像检索

Precision and Recall

True positive (tp) – correct attribution

True negative (tn) – correct rejection

False positive (fp) – incorrect attribution

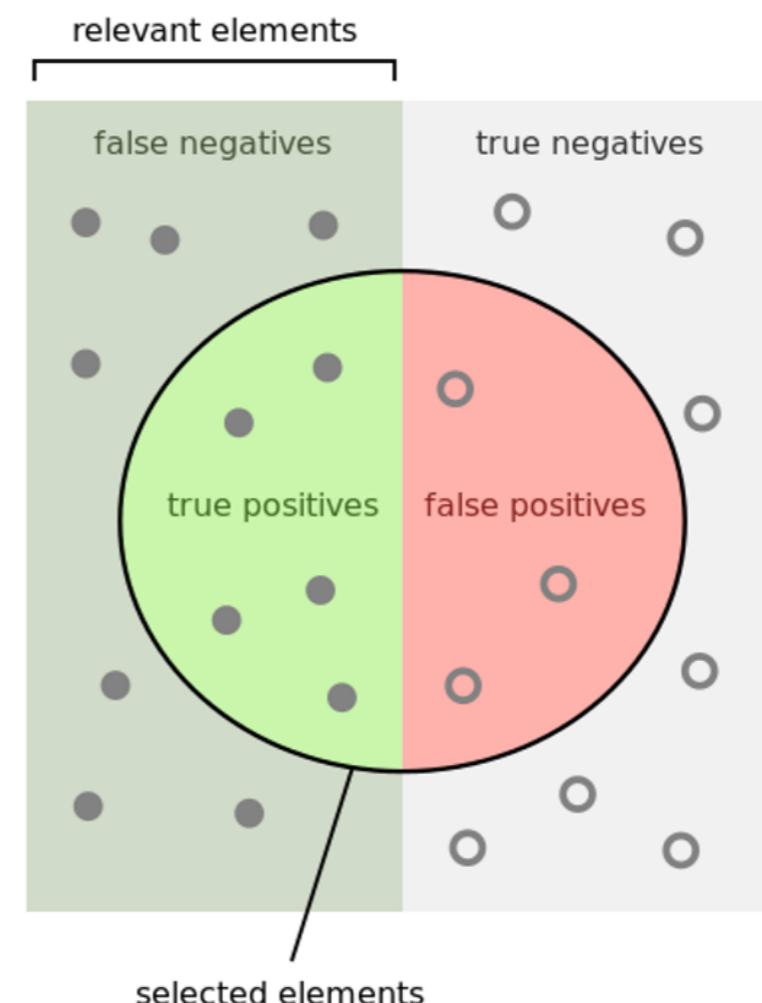
False negative (fn) – incorrect rejection

$$\text{Precision} = \frac{tp}{tp + fp}$$

Precision = #relevant / #returned

$$\text{Recall} = \frac{tp}{tp + fn}$$

Recall = #relevant / #total relevant



How many selected items are relevant?

$$\text{Precision} = \frac{\text{green}}{\text{green} + \text{red}}$$

How many relevant items are selected?

$$\text{Recall} = \frac{\text{green}}{\text{green} + \text{grey}}$$

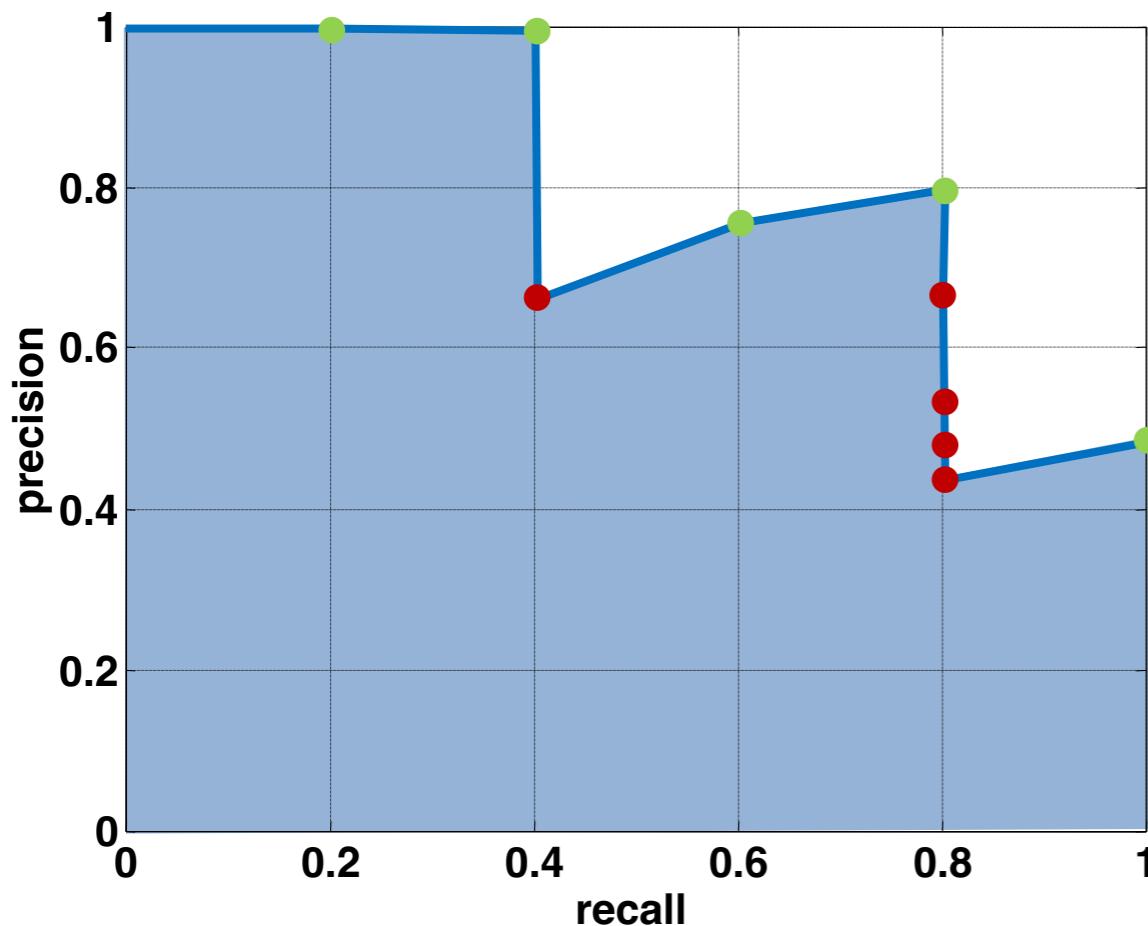
检索性能的评估



Query

Database size: 10 images
Relevant (total): 5 images

$$\text{precision} = \#\text{relevant} / \#\text{returned}$$
$$\text{recall} = \#\text{relevant} / \#\text{total relevant}$$



Results (ordered):

