

第四章：图像表达-特征描述

授课老师：李厚强，胡洋，周文罡，李礼

图像特征描述的必要性

□ 同一类或同一个物体在不同图像中像素差异大

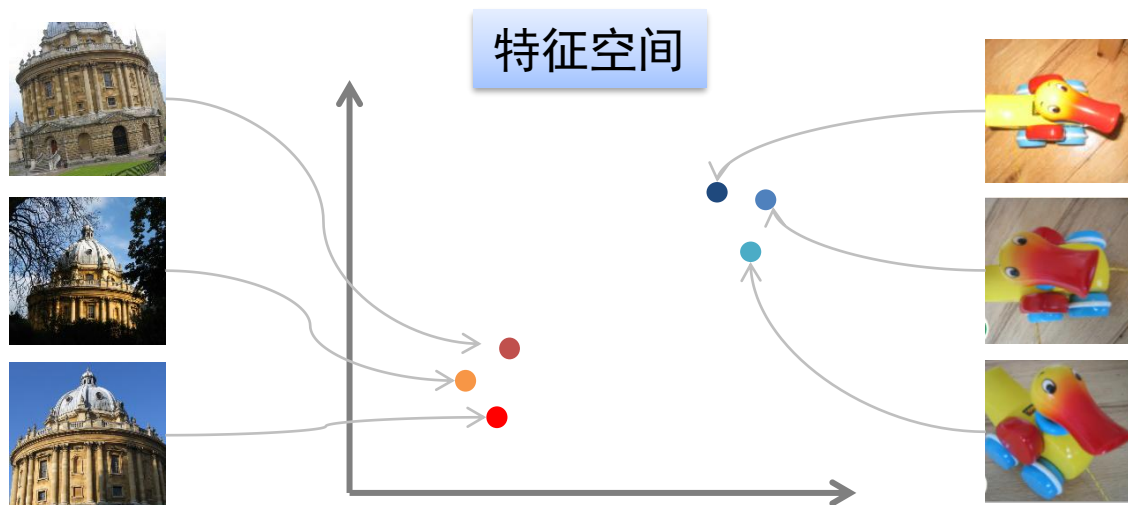
■ 空间变换

- ✓ 常见变换：平移，旋转，缩放
- ✓ 复杂的几何畸变

■ 颜色变换

□ 图像特征特征的目标

- 以不变应万变，即对于相同语义目标的不同图像，其特征表达是相同的或相似的





图像表达

□ 基于全局特征的图像表达

- 灰度特征: Histogram
- 颜色特征: Color Name
- 纹理特征: GIST

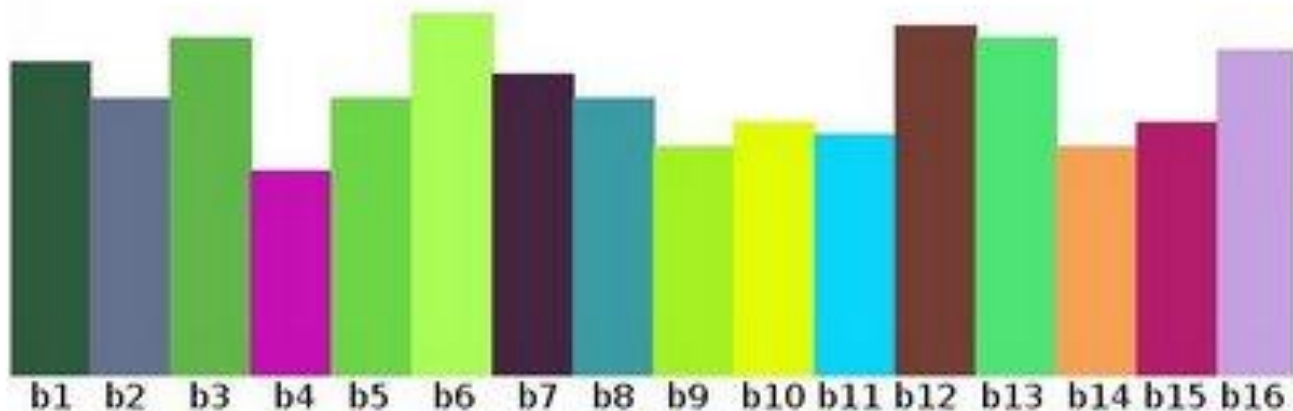
□ 基于局部特征的图像表达

- 简单的局部视觉特征
 - ✓ 局部二值模式 (LBP)
 - ✓ 梯度方向直方图 (HOG)
 - ✓ 形状上下文 (Shape Context)
- 基于关键点的局部视觉特征
 - ✓ 局部视觉特征描述: SIFT, SURF, LIOP, ORB
- 特征编码与聚合
- 乘积量化

直方图 (Histogram)

□ 什么是直方图

- 将数据的取值空间划分为一系列的子空间，对每个子空间进行编号 (bin)，然后统计落在各个子空间的样本数量（或比例）
- 直方图就是对数据进行统计的一种方法，并且将统计值组织到一系列事先定义好的 bin 当中
- 如下图所示：bin 即直方图中的纵向条带，其数值是从数据中计算出的特征统计量，这些数据可以是诸如梯度方向、色彩或任何其他特征

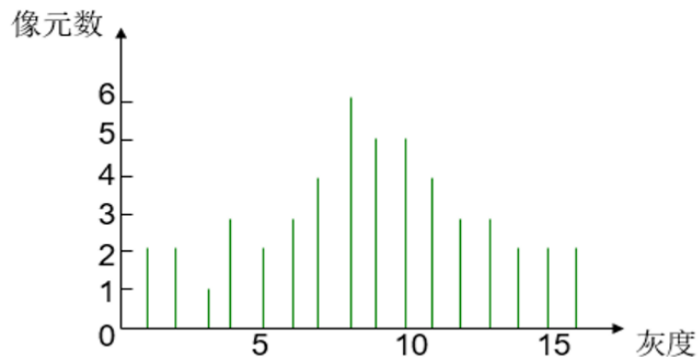


灰度直方图

- 一幅图像由不同灰度值的像素组成，图像中灰度的分布情况是该图像的一个重要特征。
- 灰度直方图对图像几何变换具有不变性

1	7	8	9	10	11	14
5	2	6	7	14	12	15
3	4	7	8	6	9	11
2	1	4	7	8	8	9
8	4	5	9	11	12	10
8	10	11	15	16	10	13
13	6	9	16	13	12	10

数字图像



灰度直方图



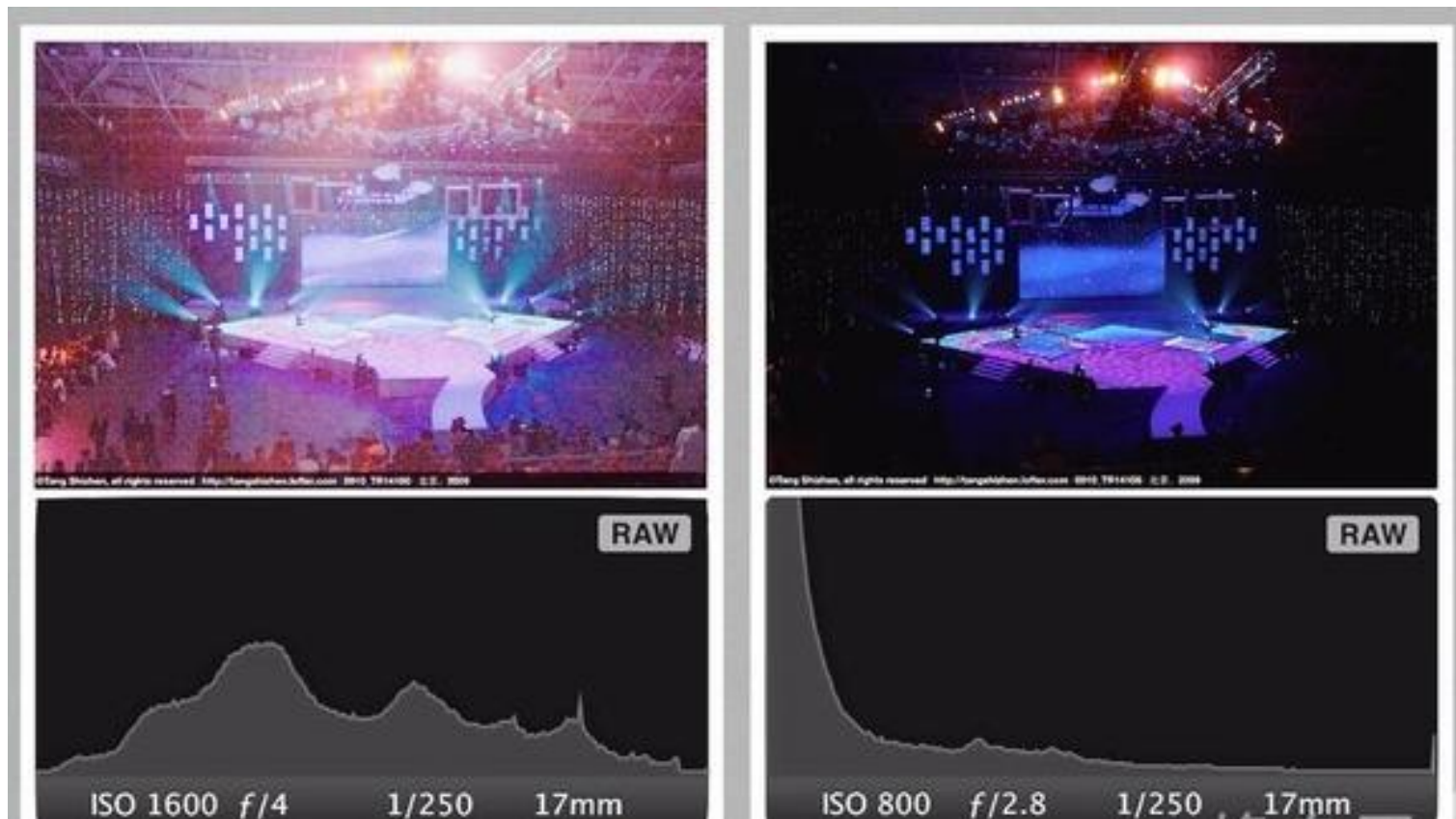
直方图统计表



灰度级	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
像元数	2	2	1	3	2	3	4	6	5	5	4	3	3	2	2	2
比例值	0.04	0.04	0.02	0.06	0.04	0.06	0.08	0.12	0.10	0.10	0.08	0.06	0.06	0.04	0.04	0.04

灰度直方图

□ 根据直方图的形态可以大致推断图像质量的好坏



灰度直方图的比较计算

□ 直方图的相关系数（事先对直方图进行 L2 归一化）

- 值越高则越匹配

$$s(H_1, H_2) = \frac{\sum_I (H_1(I) - \bar{H}_1)(H_2(I) - \bar{H}_2)}{\sqrt{\sum_I (H_1(I) - \bar{H}_1)^2 \sum_I (H_2(I) - \bar{H}_2)^2}}$$

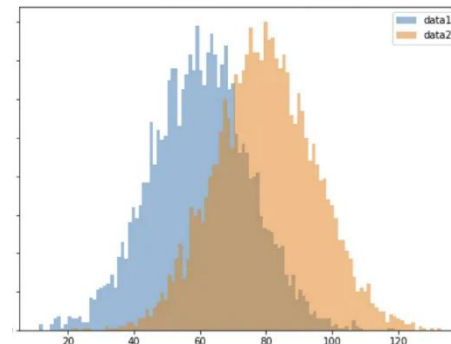
- 其中 $\bar{H}_k = \frac{1}{N} \sum_J H_k(J)$ N为直方图 bins 的个数



□ 直方图相交(intersection)，度量其相似性

- 事先对直方图进行L1归一化

$$s(H_1, H_2) = \sum_n \min(H_1(n), H_2(n))$$



Color Name: 一种颜色描述的方法



□ 定义：人们交流时用于描述颜色的词汇

■ e.g. 蓝色、红色、绿色 ...

□ Basic color terms (word): 基本颜色术语

■ 无统一可操作的定义，一般可指

✓ 可应用于描述绝大部分种类的物体

✓ 不包含于其它基本颜色

➤ 深红(crimson), 鲜红(scarlet)不是基本颜色，同属于红色(red)

✓ 该语言的绝大部分使用者对这些词汇的使用都是一致的

✓ 不同的语言间基本颜色的描述通常是共享的

➤ 中文里有红色，英语也有 red

✓ ...

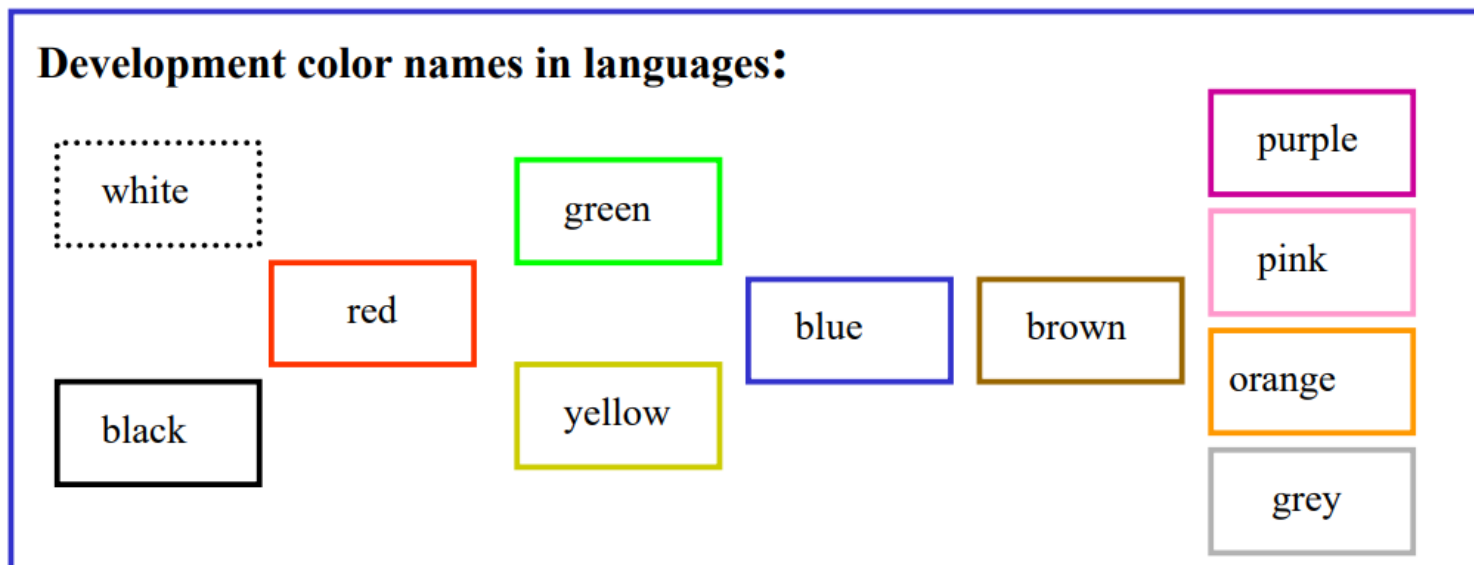
□ 不同语言的 Basic color terms 的数量是不一致的

■ 原始：黑白2种

■ Russian: 12种, English: 11种

Color Name

- 语言研究表明：不同语言中 Color Name 的发展都遵循一种相似的模式



- 英语中包含 11 种基本颜色术语
 - black, blue, brown, grey, green, orange, pink, purple, red, white, and yellow

Computational Color Naming



□ 学习如何将像素值映射为color name标签



□ 问题：大量图片中物体的颜色并未明确标注，是否可以自动对这些图片用 color name 进行标注？

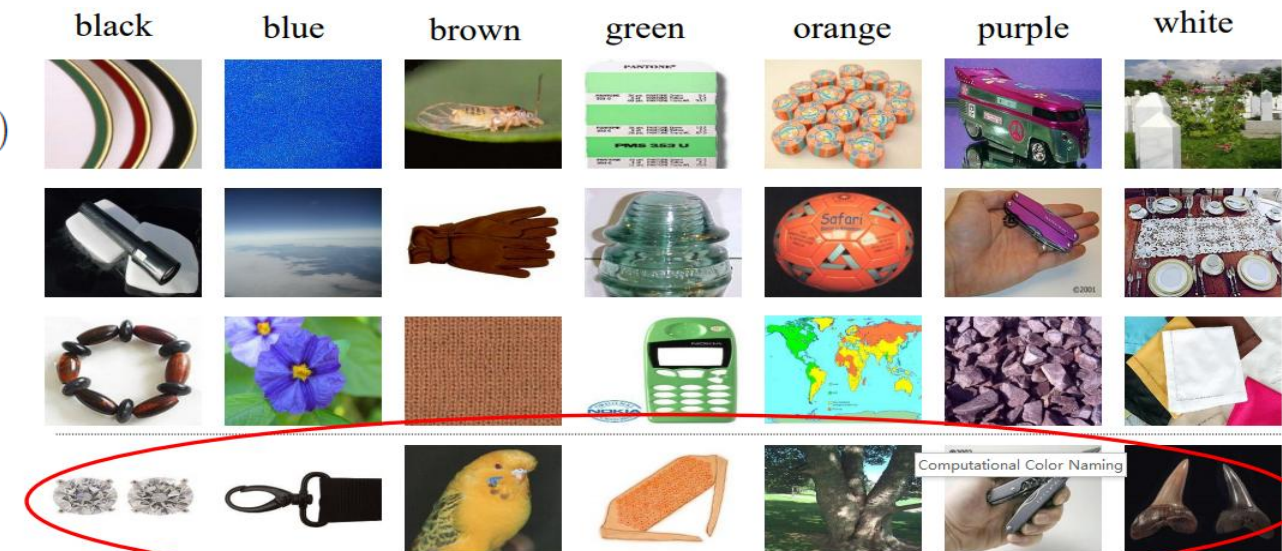
■ Find me all yellow cars?



Color Names学习

- 使用搜索引擎收集一组弱标记图片(weakly labeled)
- 利用话题模型 (topic model) 学习图像color name特征表达
 - 话题模型: Probabilistic Latent Semantic Analysis (PLSA)
 - Document (图像)->Topic (color name)->Word (RGB value)

$$P(w|d) = \sum_{z \in Z} P(w|z)P(z|d)$$



false positives

使用 Google图片 检索出的图片

- Van De Weijer J, Schmid C, Verbeek J. Learning color names from real-world images[C]//2007 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2007: 1-8.

- 在图像处理中，Gabor函数是一个用于边缘提取的线性滤波器
- 二维 Gabor 滤波器：正弦平面波调制的高斯核函数

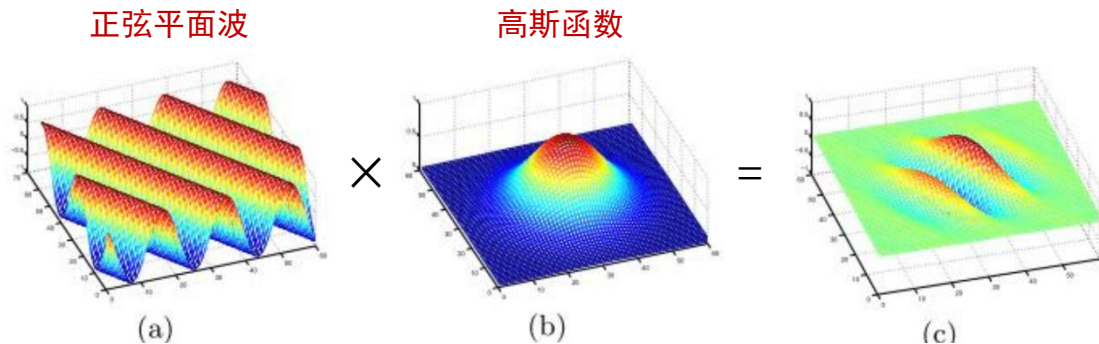
- 加窗短时 Fourier 变换

$$g(x, y) = s_c(x, y)w_r(x, y)$$

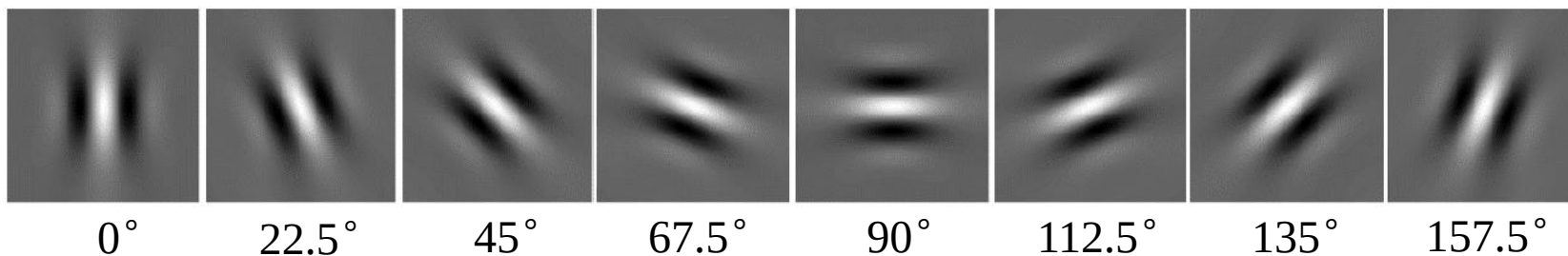
其中 $s_c(x, y) = e^{j(2\pi(u_0x+v_0y)+P)}$ $w_r(x, y) = Ke^{-\pi(a^2(x-x_0)^2+b^2(y-y_0)^2)}$

- ✓ u_0 和 v_0 分别表示空间中水平和垂直方向的频率
- ✓ P 是正弦函数的相位
- ✓ x_0 和 y_0 分别表示高斯函数的中心

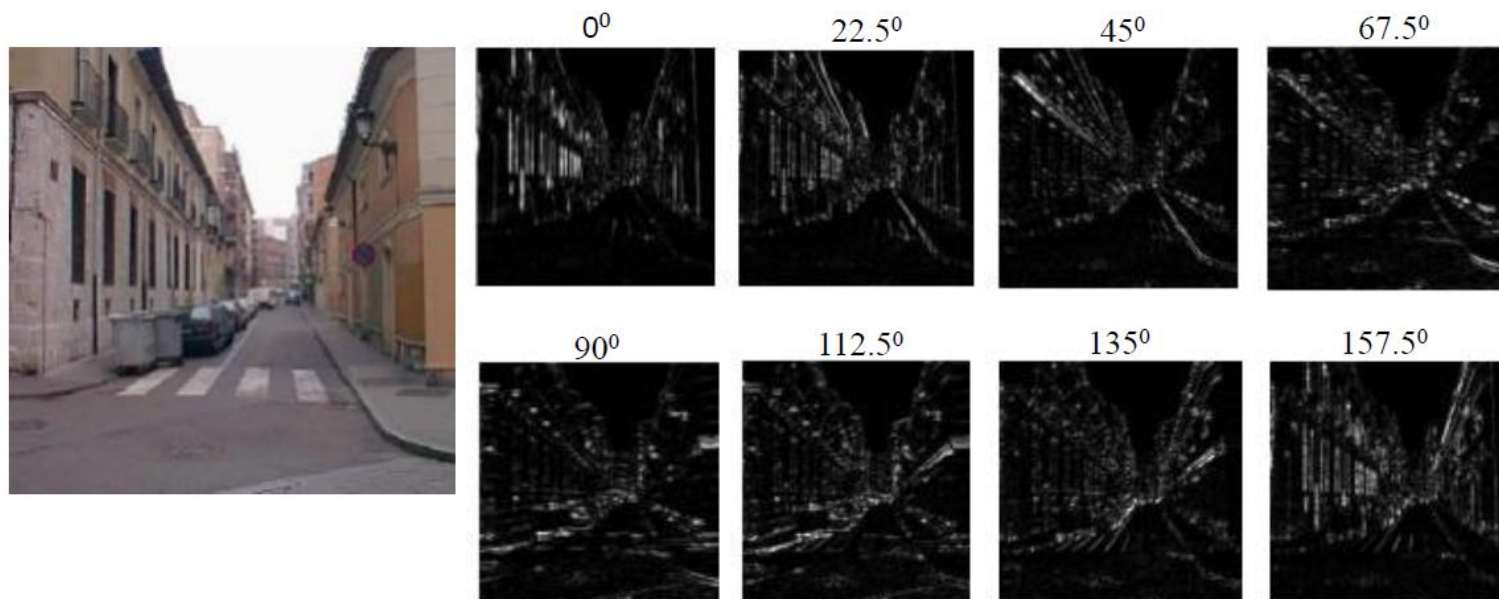
- Gabor滤波器的频率和方向表达同人类视觉系统类似
- Gabor滤波器十分适合纹理表达和分离



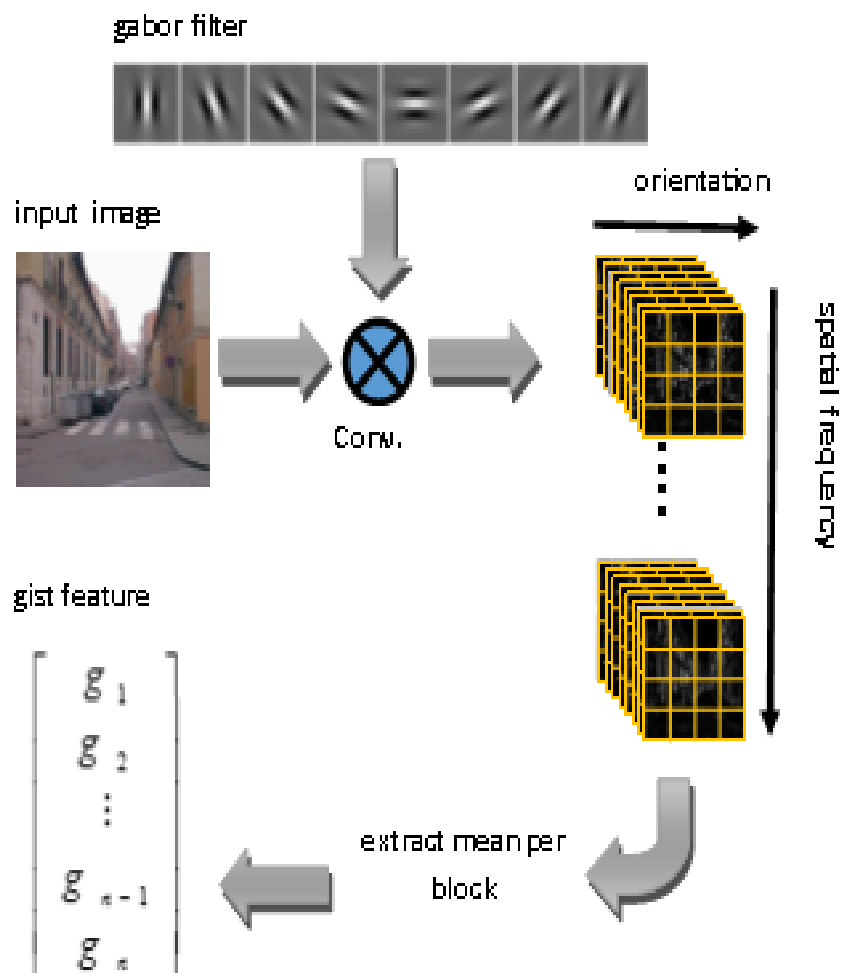
□ 不同方向的 Gabor 滤波器



□ Gabor 滤波示例

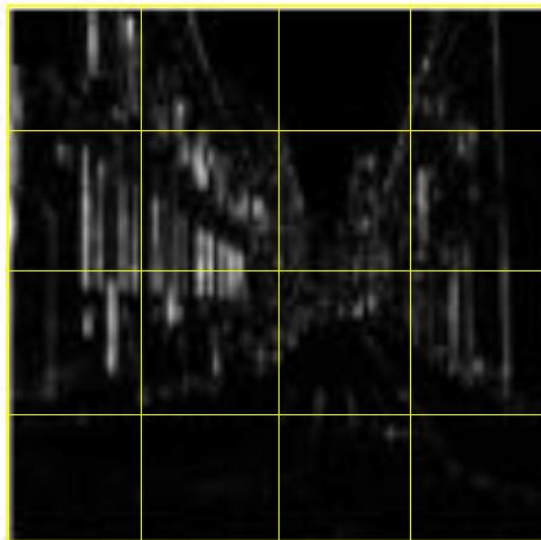


□ GIST特征提取一般流程



□ GIST-512

- 32个 gabor 滤波器：8个方向，4个尺度
 - ✓ 得到 32 个与原图等大的 feature map。
- 将每个 feature map 分为 4x4 block，计算每个block内的均值
- 将 32 个 feature map 的所有 block 的均值合并
 - ✓ 共 16x32=512 个均值，即 512 维 GIST 特征



$$\bar{x} = \frac{\sum x}{n}$$

feature map 中每个 block 计算均值

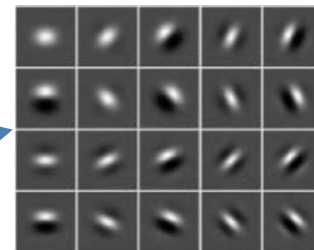
□ GIST 特征提取算法

Algorithm : Extraction algorithm of global GIST feature

Input: Image $I(x, y)$ with the size of $M \times N$

Out put: Feature vector G^G

- 1: for $j=1, \dots, m$ do
- 2: for $k=1, \dots, n$ do
- 3: Calculate $m \times n$ Gabor filters $g_{jk}(x, y)$
- 4: end for
- 5: end for
- 6: Divide $I(x, y)$ into $n_g = n_b \times n_b$ blocks and label those different blocks with B_i
- 7: for $i=1, \dots, n_g$ do
- 8: Calculate $G_i^B(x, y)$
- 9: Calculate $\overline{G_i^B}$
- 10: end for
- 11: $G^G = \{\overline{G_1^B}, \overline{G_2^B}, \dots, \overline{G_{n_g}^B}\}$



$$G_i^B(x, y) = \text{cat}(I(x, y) * g_{mn}(x, y)), (x, y) \in B_i$$

$$\overline{G_i^B} = \frac{1}{M' \times N'} \sum_{(x, y) \in B_i} G_i^B(x, y)$$

block size



图像表达

□ 基于全局特征的图像表达

- Histogram
- Color Name
- GIST

□ 基于局部特征的图像表达

- 简单的局部视觉特征
 - ✓ 局部二值模式 (LBP)
 - ✓ 梯度方向直方图 (HOG)
 - ✓ 形状上下文 (Shape Context)
- 基于关键点的局部视觉特征
 - ✓ 局部视觉特征描述: SIFT, SURF, LIOP, ORB
- 特征编码与聚合
- 乘积量化

局部二值模式 (Local Binary Pattern)

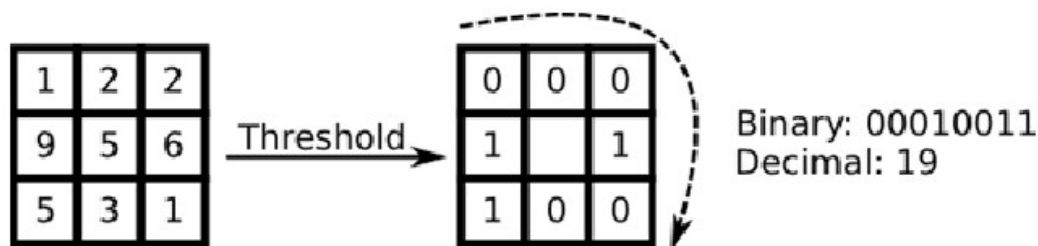


□ 图像特点

- 在不同光照条件下，同一场景的图像像素亮度变化显著，但局部区域的像素亮度相对大小关系是稳定的

□ LBP特征描述

- 在3x3的窗口内，以中心像素为阈值，将相邻8个像素的灰度值与中心像素比较
 - ✓ 周围像素值 $\geq$ 中心像素值：标记为 1
 - ✓ 周围像素值 $<$ 中心像素值：标记为 0
- 邻域8个点经比较可产生8位二进制数
 - ✓ 种类：256种 (2^8)
 - ✓ 转换为十进制数：LBP值



局部二值模式 (Local Binary Pattern)

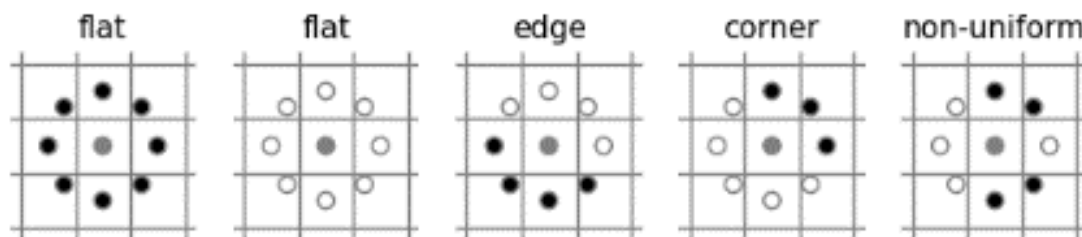


□ 函数表示

- LBP操作可定义为
- $LBP(x_c, y_c) = \sum_{p=0}^{P-1} 2^p \cdot s(i_p - i_c)$
- 其中 (x_c, y_c) 为中心像素，其亮度为 i_c ， i_p 是周围像素亮度， s 为符号函数。

□ 小结

- LBP可以很好地捕捉图像细节，其值可反映不同的纹理

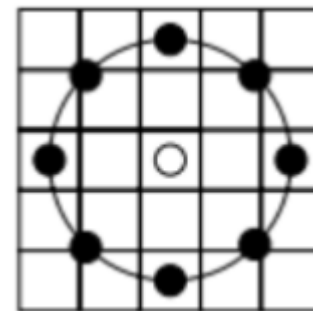


- 但由于邻域区域大小固定，不能满足不同尺度和频率纹理需要
- 对图像的亮度变化具有较好的鲁棒性/不变性
 - ✓ 不变性(invariance): 以不变应万变

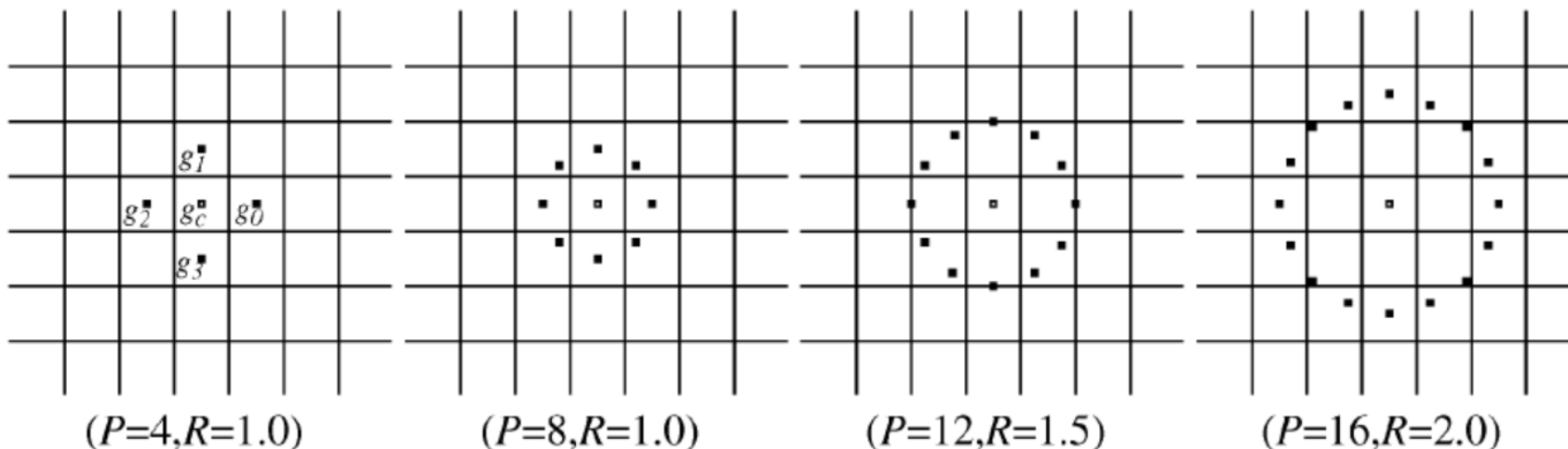
LBP的扩展（环状LBP）

□ 改进

- 3x3邻域扩展为任意邻域
- 圆形邻域代替正方形邻域
- 非整数点像素的灰度值可通过插值得到



□ LBP_P^R : 表示在半径为 R 的圆内有 P 个像素点



LBP的扩展（环状LBP）

□ 对于一个给定的点 (x_c, y_c) ，其近邻点为 (x_p, y_p) ， $p \in P$

■ 可以由如下公式计算坐标

$$x_p = x_c + R \cos\left(\frac{2\pi p}{P}\right) \quad y_p = y_c - R \sin\left(\frac{2\pi p}{P}\right)$$

✓ 其中R 圆的半径，P 为样本点的数目

■ 如果近邻点不在图像的整数坐标上，如何确定该点的像素值？

✓ 双线性插值：

$$f(x, y) \approx \begin{bmatrix} 1-x & x \end{bmatrix} \begin{bmatrix} f(0,0) & f(0,1) \\ f(1,0) & f(1,1) \end{bmatrix} \begin{bmatrix} 1-y \\ y \end{bmatrix}$$



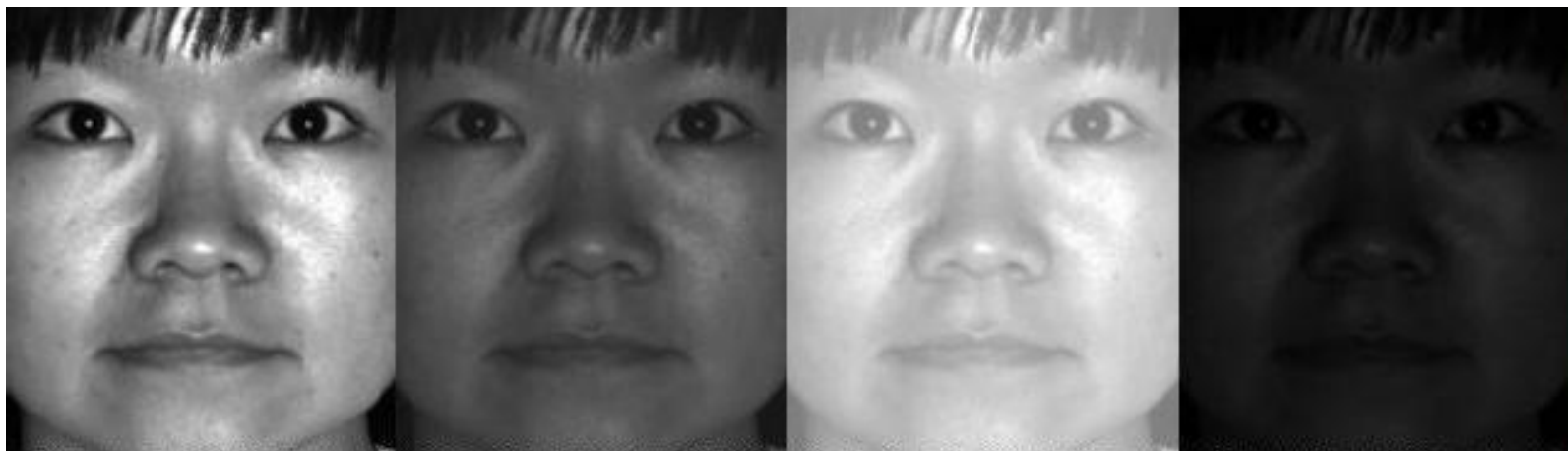
半径越小，纹理越细

原始LBP 圆形 LBP_8^1 圆形 LBP_8^3 圆形 LBP_6^3

局部二值模式：光照不变性

□ 由定义可见，LBP 对光照变化非常鲁棒

原图

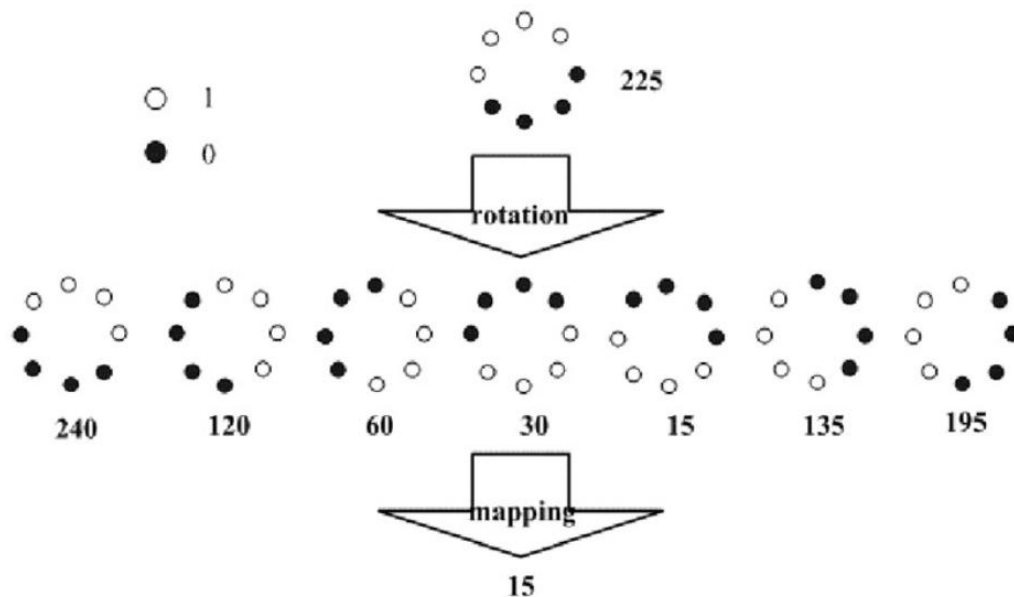


LBP图



局部二值模式：旋转不变性

- 旋转不变性：图像发生旋转变换后，图像特征保持不变
- LBP本身不是旋转不变的，同一幅图像，进行旋转后，其特征会有很大的差别
- 改进：不断旋转圆形邻域得到一系列初始定义的LBP值，取最小值作为该邻域的值
 - 下图 8 种LBP模式，对应旋转不变的LBP模式均为 00001111
 - 代价：牺牲了特征对不同模式的区分性



LBP等价模式 (Uniform Pattern LBP)



□ 问题：二进制模式过多

- 一个LBP算子可以产生不同的二进制模式，对于半径为R的圆形区域内含有P个采样点的LBP算子 (LBP_P^R)将会产生 2^P 种模式。（3x3邻域内采8个点，则有 2^8 种模式，7x7邻域内采36个点，则有 2^{36} 种模式）

□ 解决方法：等价模式

- 作用：对LBP算子的模式种类进行归并
- 定义：当某个LBP所对应的循环二进制数从0到1或从1到0最多有两次跳变时，该LBP所对应的二进制就称为一个等价模式类

□ 示例

- 等价模式类：
 - ✓ 00000000 (0次跳变), 00000111 (2次跳变), 10001111 (2次跳变), ...
- 除等价模式类以外的模式均归为一类，称作混合模式类：
 - ✓ 10010111 (4次跳变)

LBP等价模式 (Uniform Pattern LBP)



□ 模式数量降维

- 等价模式类数量: $2^P \rightarrow P(P - 1) + 2$

□ 示例

- 3x3邻域8个采样点
 - ✓ 总模式由 256 种减少为 59 种
 - ✓ 等价模式的值从小到大编码为 1-58, 混合模式类编码为 0
- 等价模式LBP特征图整体偏暗, 特征向量维数更少, 还可以减少高频噪声带来的影响。



圆形
 LBP_8^3

旋转不变
 LBP_8^3

等价模式
 LBP_8^3



LBP的应用：LBP Histograms

□ LBP特征统计直方图(LBPH)

- 将LBP特征与图像的空间信息结合。
- 首先，将LBP特征图像分成m个局部块，并提取每个局部块的直方图
- 然后，将这些直方图依次连接在一起形成LBP特征的统计直方图，即LBPH。

□ 应用：OpenCV人脸识别

- 计算图像的LBP图像
- 将LBP特征图像进行分块（如分成8行8列64块区域）
- 计算每块区域特征图像的直方图，将直方图进行归一化，直方图大小为： $1 \times NumPatterns$
- 将每块区域的直方图按空间顺序依次排列成一行，形成LBP特征向量，大小为： $1 \times (NumPatterns \times 64)$
- 用机器学习的方法对LBP特征向量进行训练，用于检测和识别目标



LBP: 小结

□ 优点

- 一定程度上消除了光照变化的影响
- 通过改进，可具有旋转不变性
- 纹理特征维度低，计算速度快

□ 缺点

- 当光照变化不均匀导致局部邻域像素间的灰度大小关系被破坏,对应的LBP结果可能也发生变化
- 通过引入旋转不变的定义，使LBP算子更具鲁棒性。但这也使得LBP结果丢失了方向信息



图像表达

□ 基于全局特征的图像表达

- Histogram
- Color Name
- GIST

□ 基于局部特征的图像表达

- 简单的局部视觉特征
 - ✓ 局部二值模式 (LBP)
 - ✓ 梯度方向直方图 (HOG)
 - ✓ 形状上下文 (Shape Context)
- 基于关键点的局部视觉特征
 - ✓ 局部视觉特征描述: SIFT, SURF, LIOP, ORB
- 特征编码与聚合
- 乘积量化



HOG: 梯度方向直方图

□ HOG (Histogram of Oriented Gradient, HOG)

- 统计图像局部区域的梯度方向直方图
- 主要思想
 - ✓ 在一副图像中，局部目标的外观和形状（appearance and shape）能够被梯度或边缘的方向分布描述。
 - ✓ 本质：梯度的统计信息，而梯度主要存在于边缘的地方

□ 主要步骤

- 梯度计算
- 直方图统计
- 块(block)描述子

梯度方向直方图：梯度计算

□ 图像预处理（可选）

- 灰度化：将彩色图像变换为灰度图像
- 伽马校正：调节图像的对比度，降低光照因素的影响

□ 梯度计算

- 计算水平梯度 g_x 和竖直梯度 g_y
- 计算总的梯度强度和梯度方向
 - ✓ 有符号梯度： $0 \sim 360^\circ$ ；无符号梯度： $0 \sim 180^\circ$

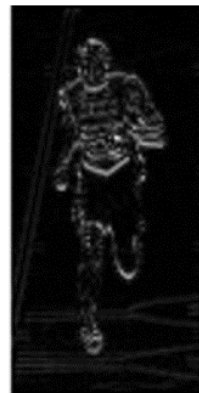
-1	0	1
----	---	---

-1
0
1

$$g = \sqrt{g_x^2 + g_y^2} \quad \theta = \arctan \frac{g_x}{g_y}$$



水平梯度



竖直梯度

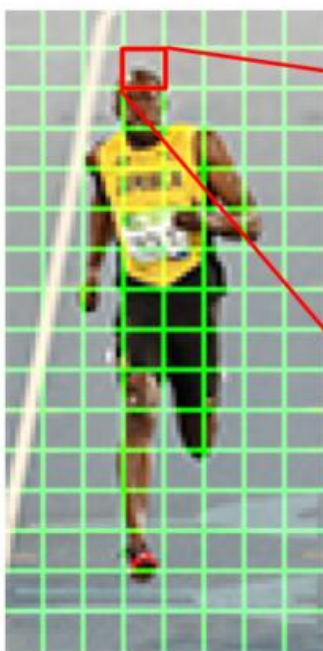


梯度强度

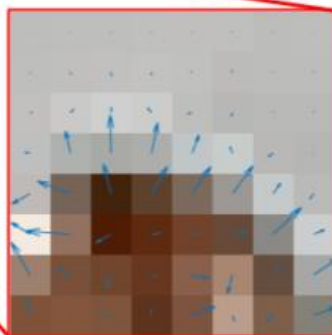
梯度方向直方图：直方图统计-I

□ 对每个 cell 统计梯度直方图

- Cell size: 8x8
- 对于灰度图，每个 8x8 的 cell 包含 64 个像素，每个像素包含 2 个值（梯度大小和方向）
- 使用 9-bins 的直方图对梯度进行统计



8x8 cells of HOG



2	3	4	4	3	4	2	2
5	11	17	13	7	9	3	4
11	21	23	27	22	17	4	6
23	99	165	135	85	32	26	2
91	155	133	136	144	152	57	28
98	196	76	38	26	60	170	51
165	60	60	27	77	85	43	136
71	13	34	23	108	27	48	110

Gradient Magnitude

梯度大小

80	36	5	10	0	64	90	73
37	9	9	179	78	27	169	166
87	136	173	39	102	163	152	176
76	13	1	168	159	22	125	143
120	70	14	150	145	144	145	143
58	86	119	98	100	101	133	113
30	65	157	75	78	165	145	124
11	170	91	4	110	17	133	110

Gradient Direction

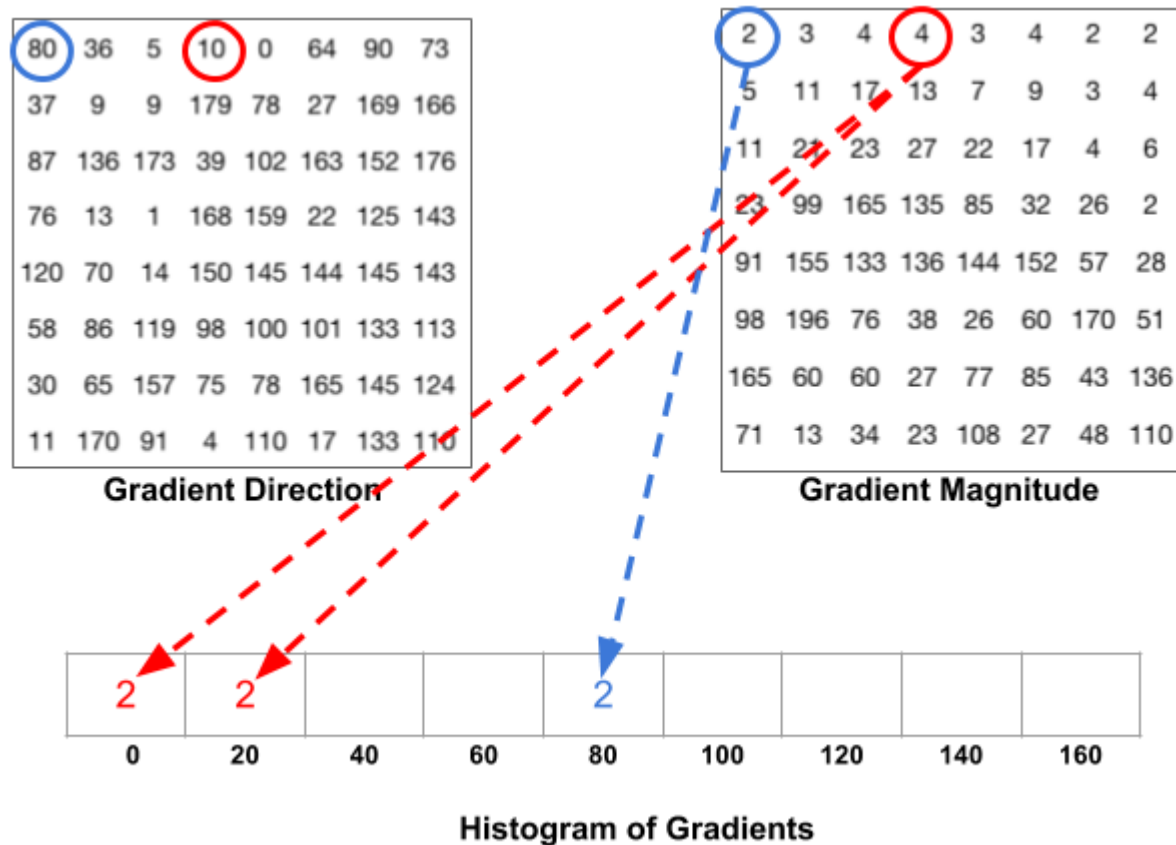
梯度方向
(0~180°)

梯度方向直方图：直方图统计-II

□ 将 0-180 度分为 9 个 bins

■ 统计：[0, 20, 40, ..., 160]

■ 将梯度幅值线性分配到与梯度方向最近的两个bin中

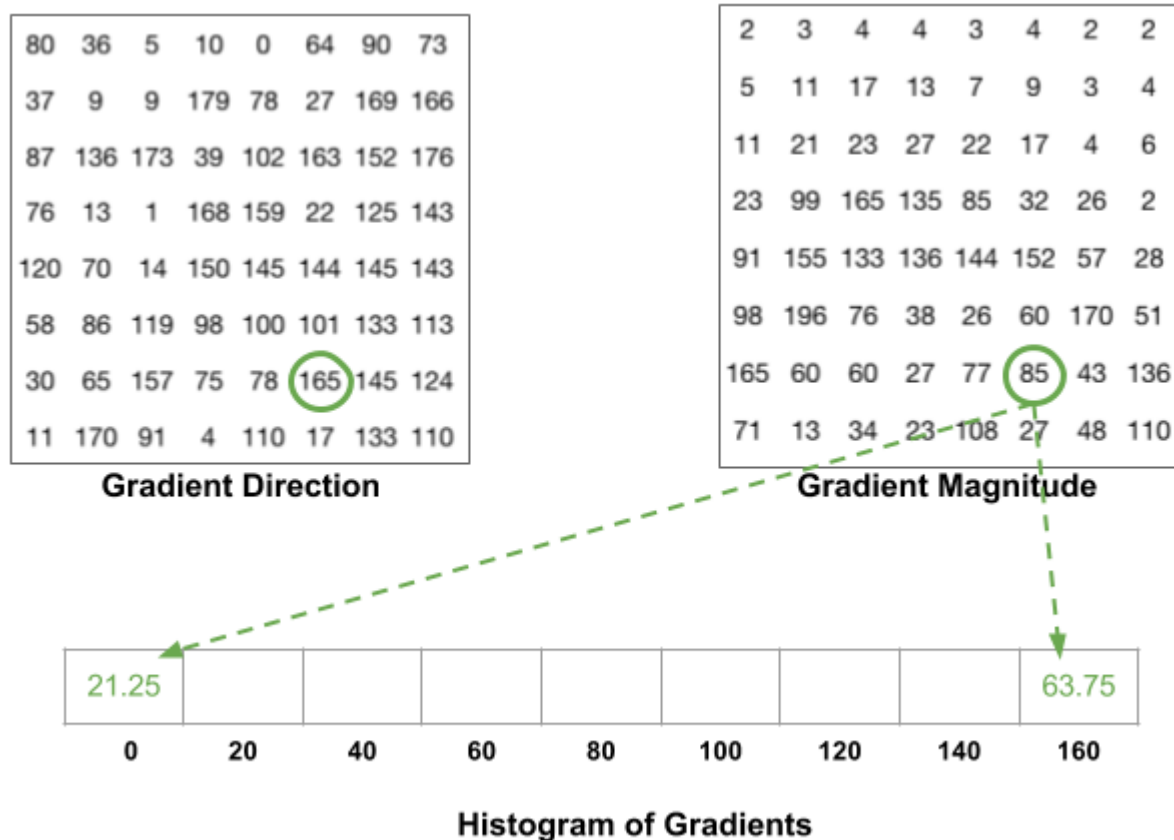


梯度方向直方图：直方图统计-III

□ 将 0-180 度分为 9 个 bins

■ 统计：[0, 20, 40, ..., 160]

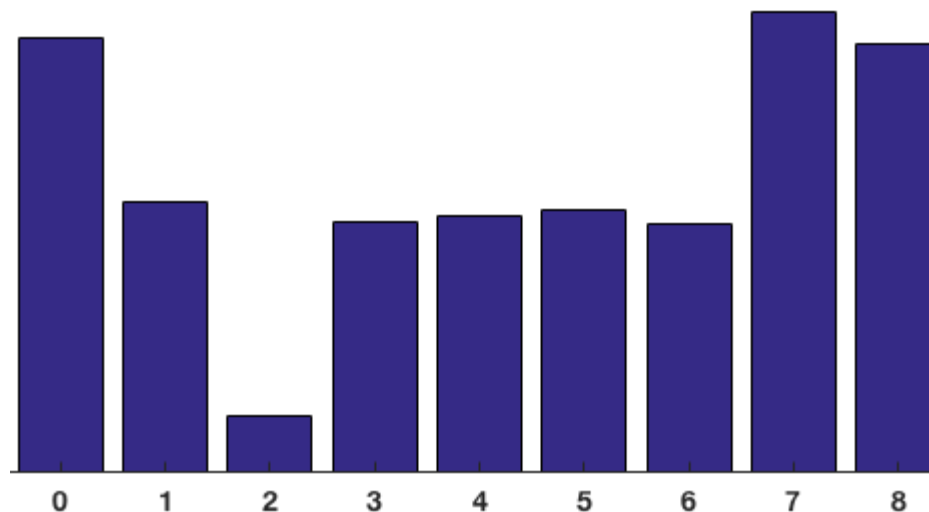
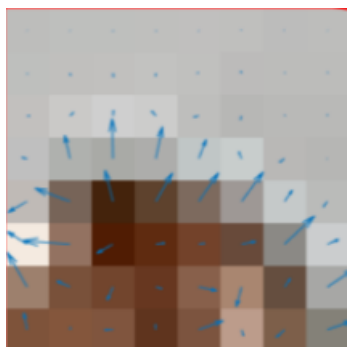
■ 将梯度幅值线性分配到与梯度方向最近的两个bin中



梯度方向直方图：直方图统计-IV

□ 每个 8x8 cell 的梯度方向直方图

- 所有像素贡献的累加
- 此处，0 度表示y轴方向，在 140-160-0(180) 度附近直方图的值较大，说明在这个 cell 中梯度方向朝上下的分量较多，有着较为明显的横向边缘。



块描述 (block descriptor)

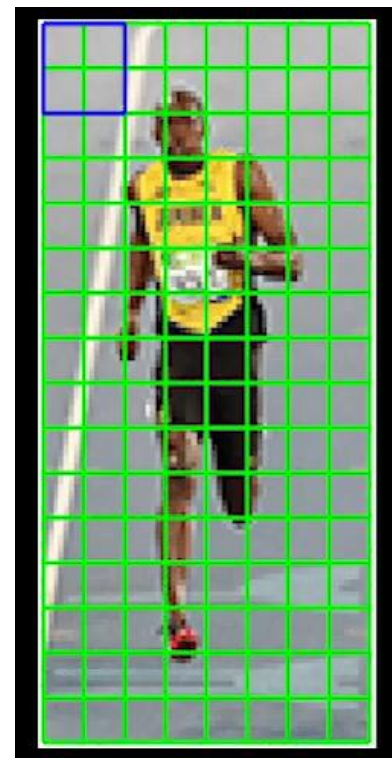
□ block 归一化

- 大小：16x16 (包含4个cell)
- 目的：降低光照的影响
- 每个block由4个维度为9的向量组成，对block进行归一化即对该36维向量进行归一化
- block间有重叠

□ 获取HOG特征向量

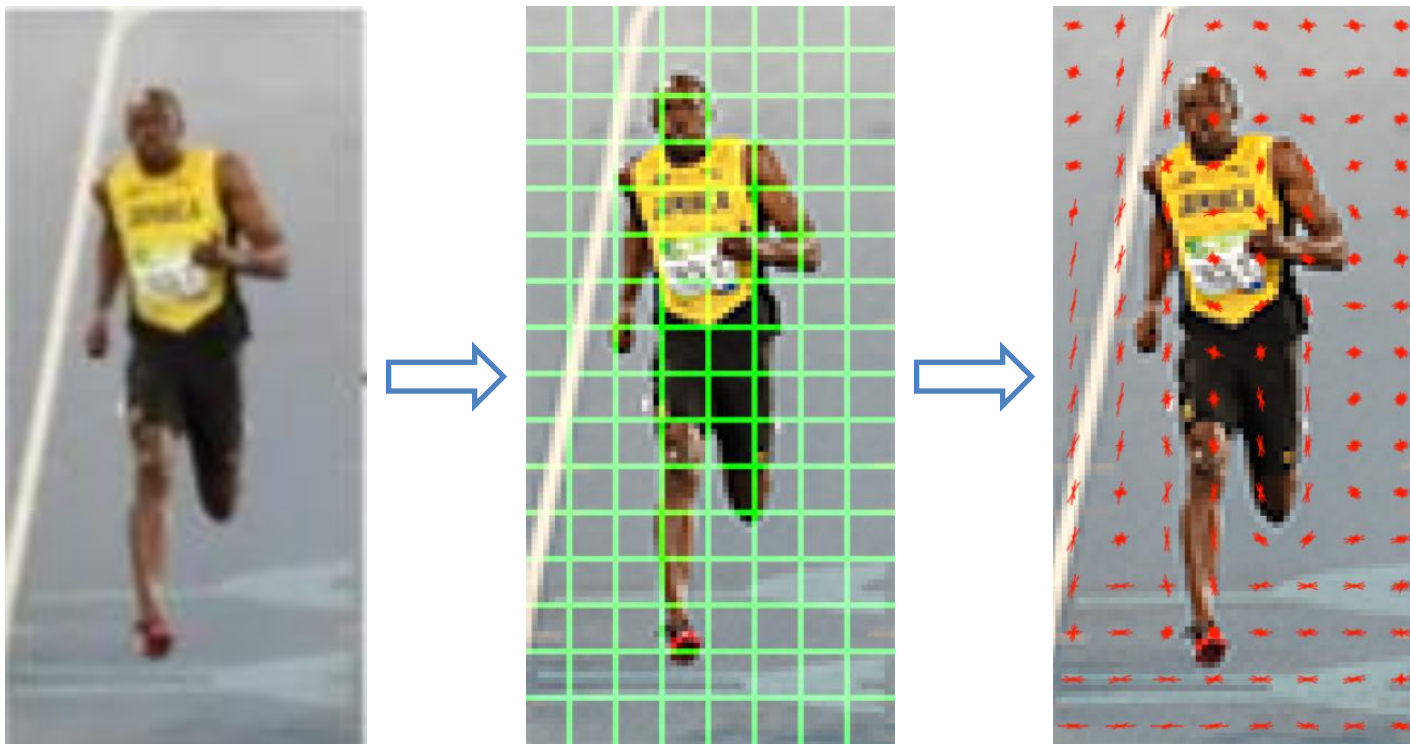
- 拼接所有block的向量即可
- 思考：图像(64x128)的HOG特征维度？
 - ✓ block 数：7x15
 - ✓ HOG特征维度：7x15x36=3780

思考：若一个block中所有像素的灰度值都经历相同参数 (a, b) 的线性变换： $f(x) = a \cdot x + b$ ，变换前后的HOG特征是否有变化？



HOG特征可视化

- 对每个 cell，画出经过block归一化后的直方图





图像表达

□ 基于全局特征的图像表达

- Histogram
- Color Name
- GIST

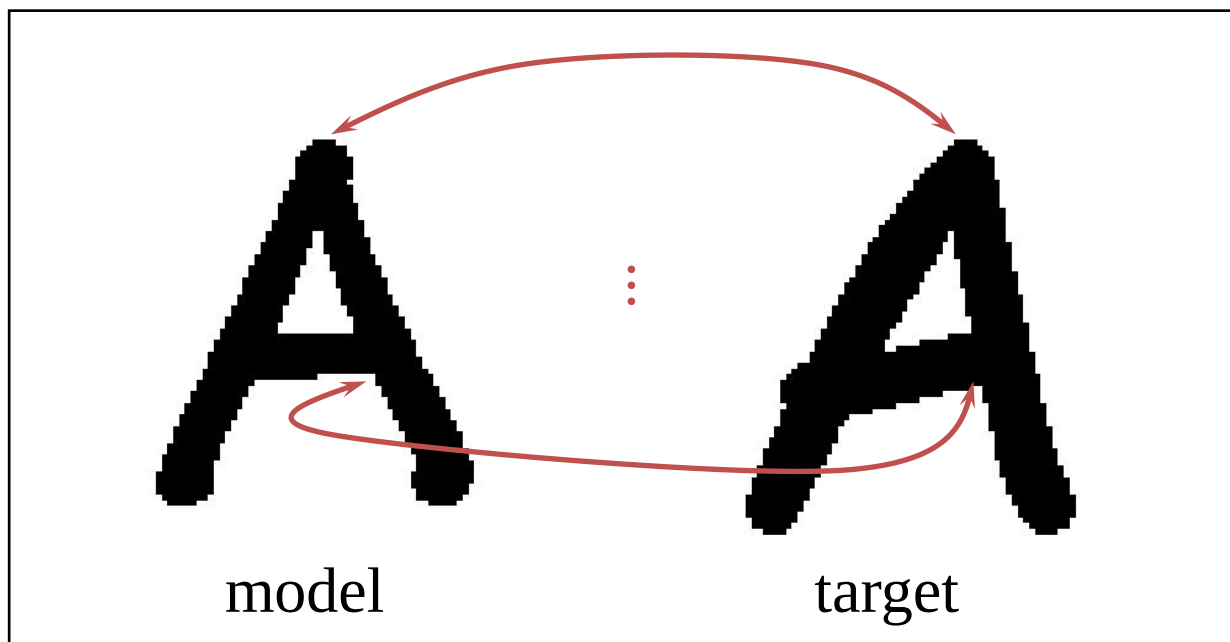
□ 基于局部特征的图像表达

- 简单的局部视觉特征
 - ✓ 局部二值模式 (LBP)
 - ✓ 梯度方向直方图 (HOG)
 - ✓ 形状上下文 (Shape Context)
- 基于关键点的局部视觉特征
 - ✓ 局部视觉特征描述: SIFT, SURF, LIOP, ORB
- 特征编码与聚合
- 乘积量化

形状上下文 (Shape Context)

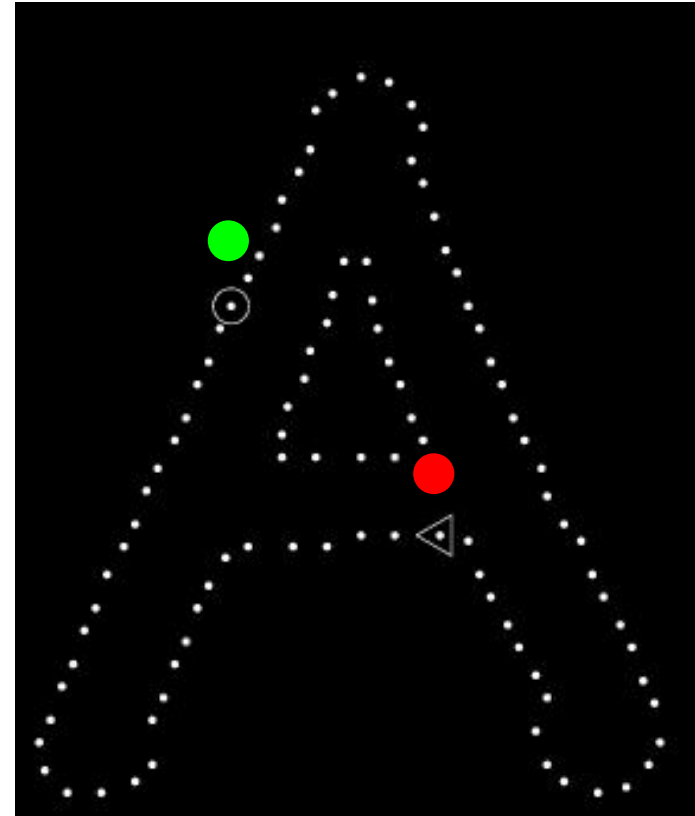
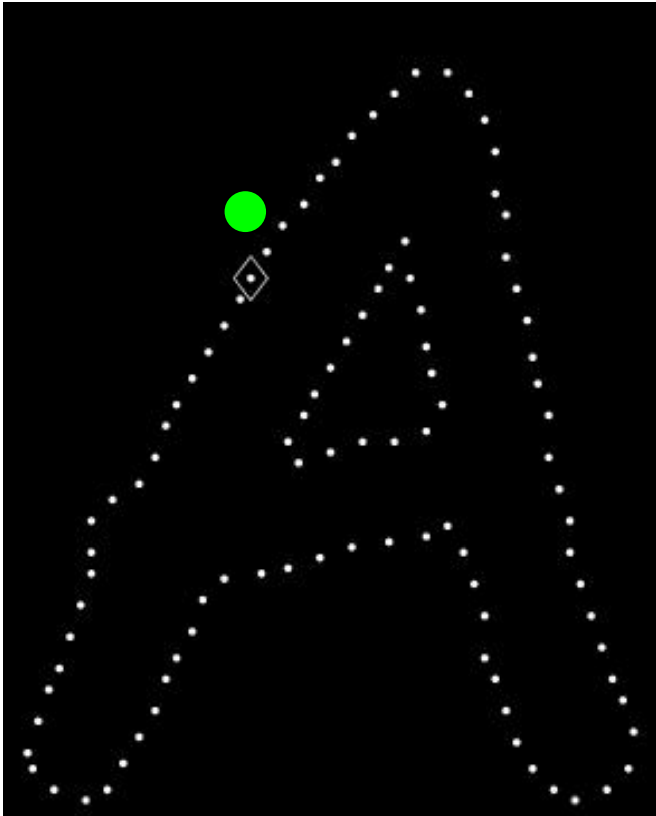


□ 动机：如何度量两个二值图像中的目标相似性



形状上下文：形状比较

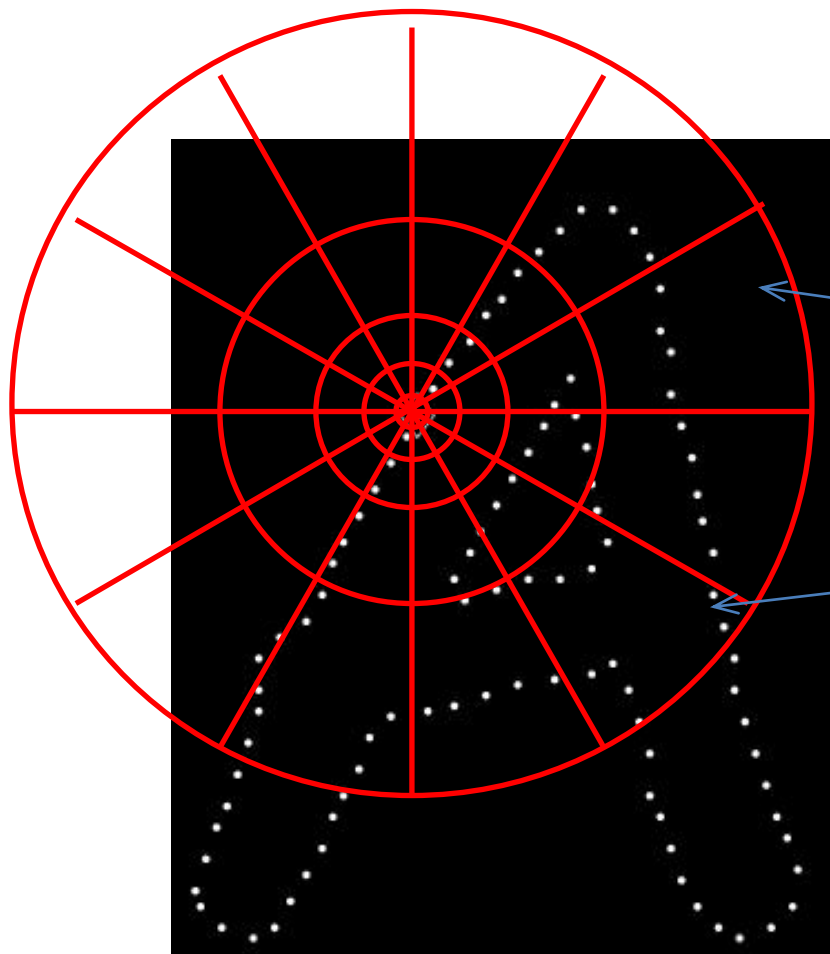
- 提取二值图像中的物体轮廓，比较轮廓点的相似性，以及匹配关系



形状上下文



□ 以某一个轮廓点为参考原点，统计其他轮廓点的分布



检查落在每个bin区间中的点的数量:

Count = 4

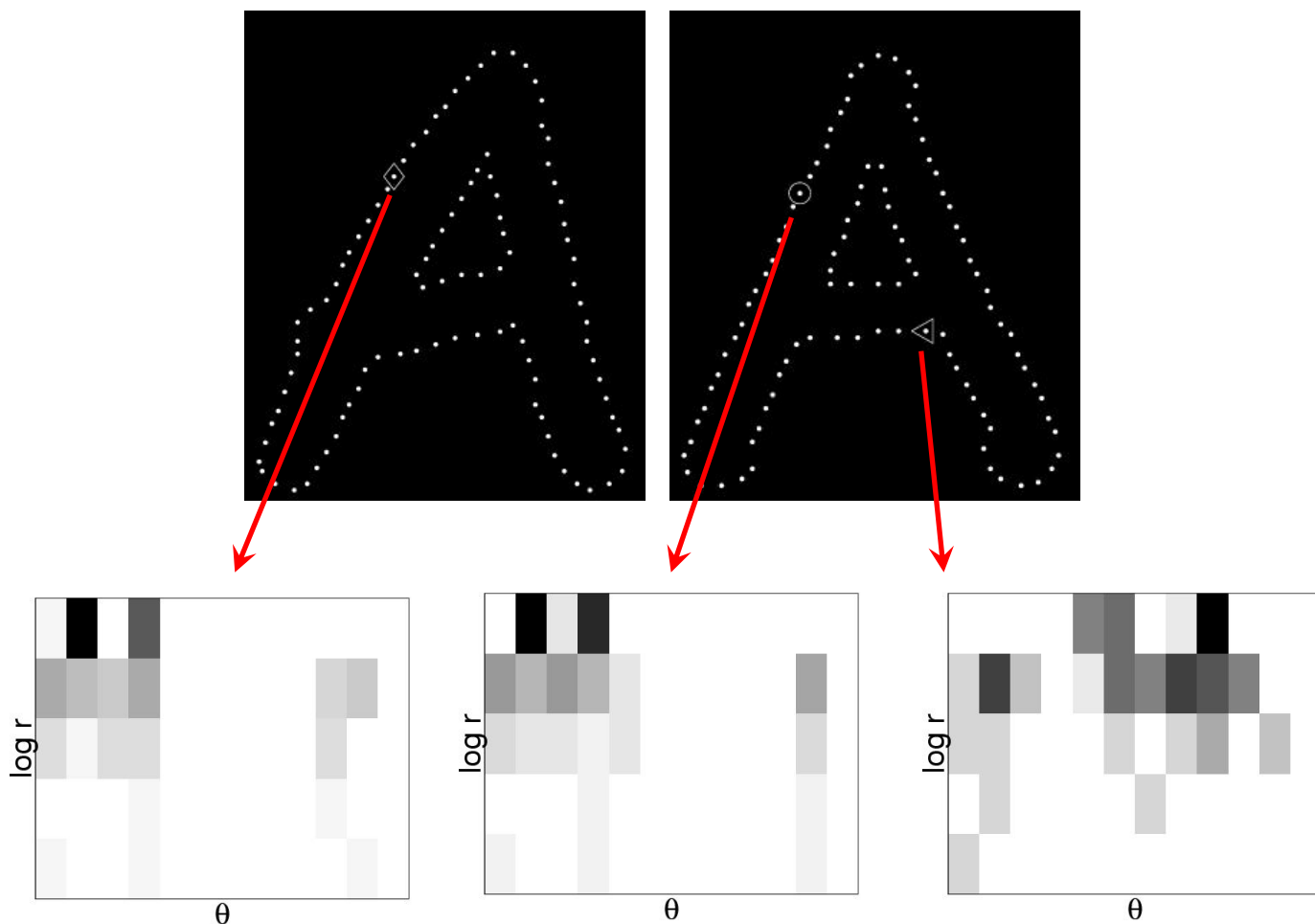
⋮

Count = 10

形状上下文



- 极坐标空间划分涉及两个变量 r 和 θ ，轮廓点分布的直方图可以用二维矩阵表示

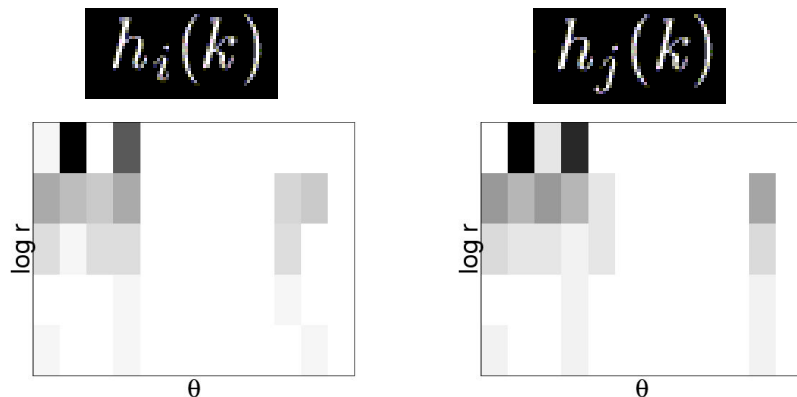




形状上下文：性质

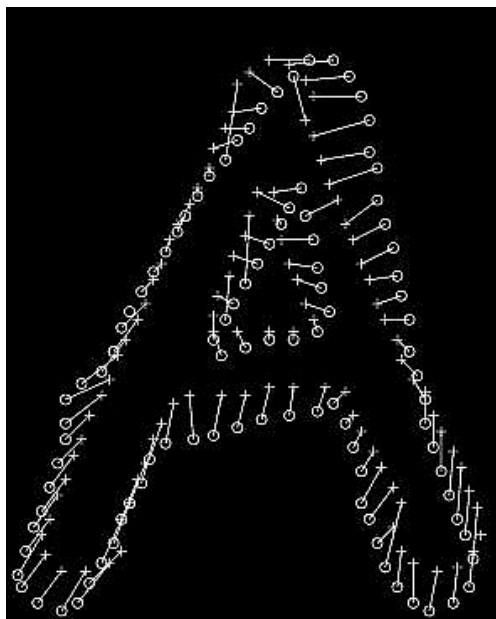
- 对平移变换具有不变性
- 通过选择合适的半径，可对缩放变换具有不变性
- 通过选择局部切向作为bin编号的参考方向，可对旋转具有不变性
- 可容忍小的仿射畸变

形状上下文：距离度量



基于Chi Squared distance, 计算两个轮廓的距离：

$$C_{ij} = \frac{1}{2} \sum_{k=1}^K \frac{[h_i(k) - h_j(k)]^2}{h_i(k) + h_j(k)}$$

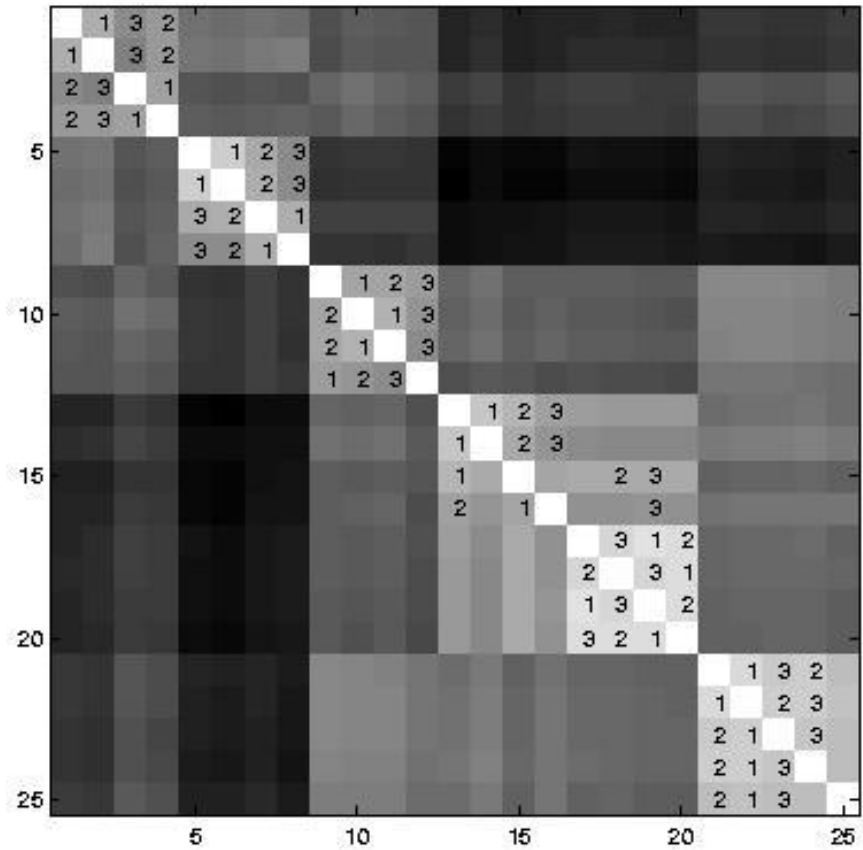
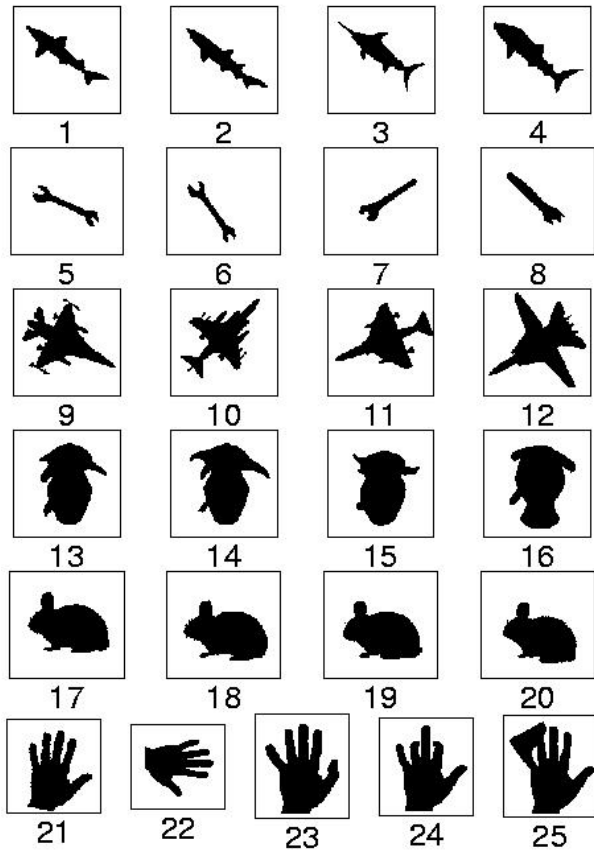


□ 利用匹配代价 $C_{i,j}$, 通过求解线性分配问题, 得到两组轮廓点的对应关系:

$$H(\pi) = \sum_i C(p_i, q_{\pi(i)})$$

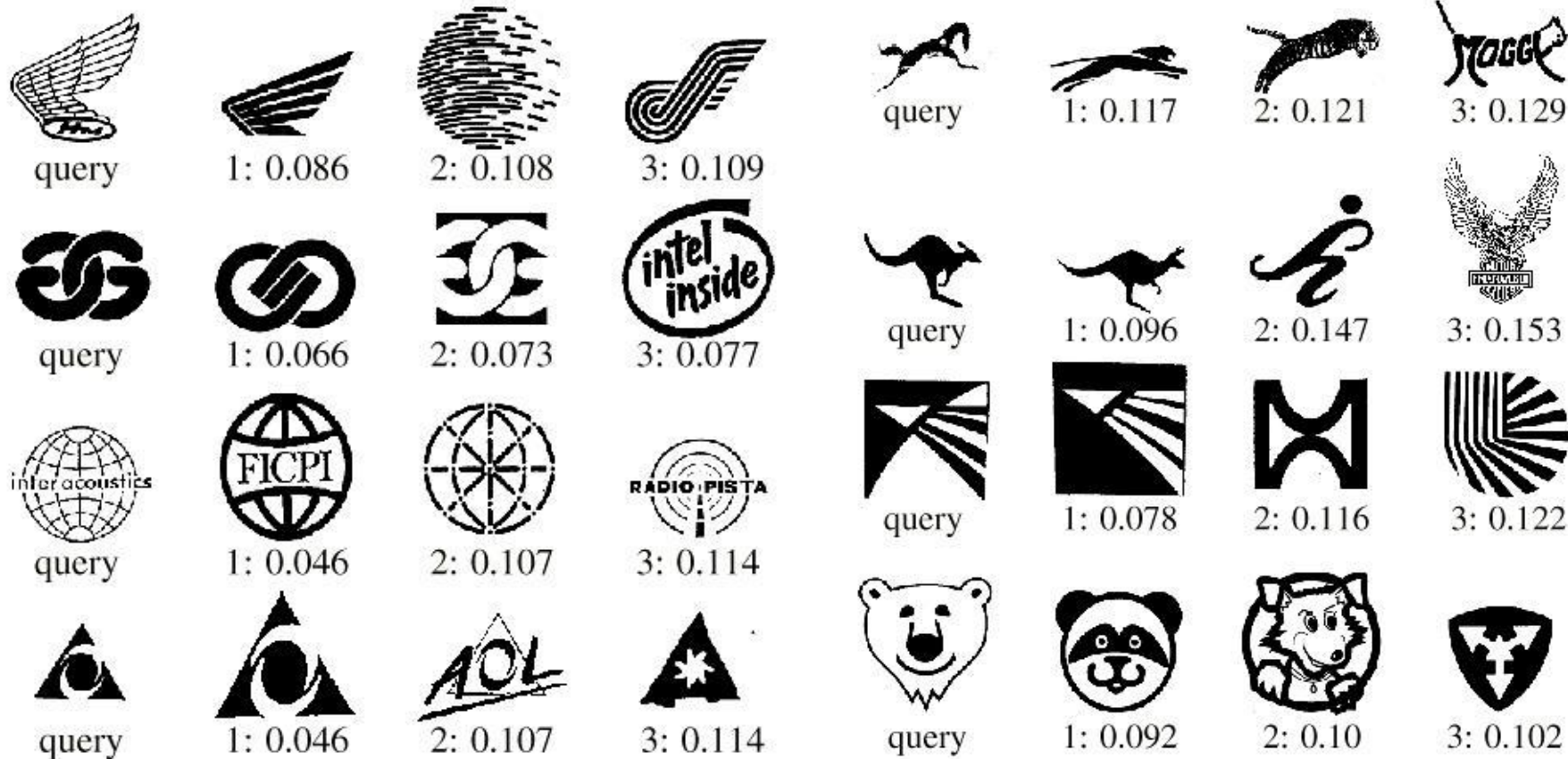
二分图匹配算法：匈牙利算法

形状上下文的应用-I



- S. Belongie, J. Malik, and J. Puzicha, “Shape matching and object recognition using shape contexts”, IEEE TPAMI 2002.

形状上下文的应用-II: 商标检索



- S. Belongie, J. Malik, and J. Puzicha, “Shape matching and object recognition using shape contexts”, IEEE TPAMI 2002.



图像表达

□ 基于全局特征的图像表达

- Histogram
- Color Name
- GIST

□ 基于局部特征的图像表达

- 简单的局部视觉特征
 - ✓ 局部二值模式 (LBP)
 - ✓ 梯度方向直方图 (HOG)
 - ✓ 形状上下文 (Shape Context)
- 基于关键点的局部视觉特征
 - ✓ 局部视觉特征描述: SIFT, SURF, LIOP, ORB
- 特征编码与聚合
- 乘积量化

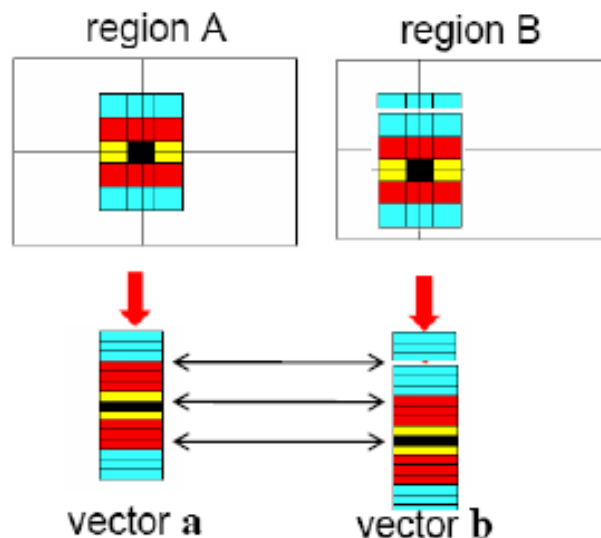
SIFT局部区域特征描述

□ 目标

- 将图像局部区域变换为固定维度的特征向量
- 实现不变性：旋转不变，亮度不变

□ 将原始的像素灰度作为局部区域的描述？

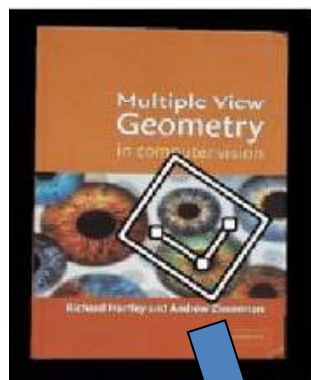
- 最简单的方式：将特征点邻域的像素灰度值逐个排列，得到一个特征向量
- 但是，这种方式对微小平移和旋转非常敏感！！！！



SIFT局部区域特征描述

□ 如何实现几何变换的不变性？

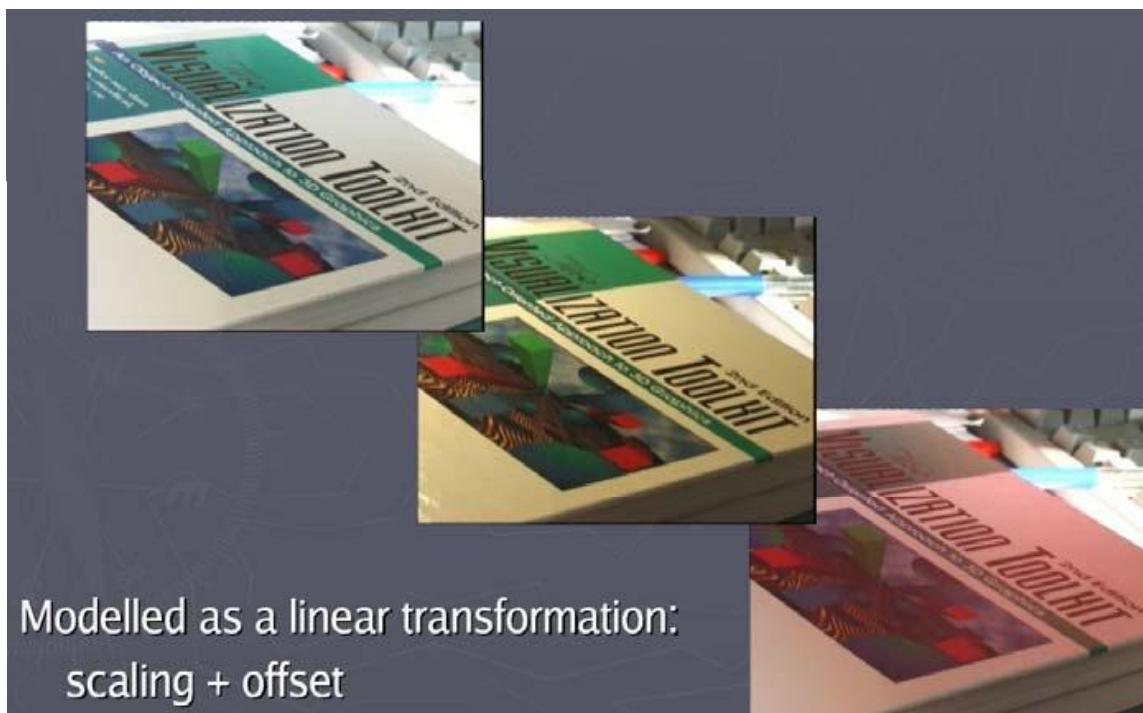
- 旋转对齐：主方向估计
- 尺寸对齐：缩放到统一尺寸



SIFT局部区域特征描述

□ 如何实现亮度变化不变性？

- 一般将RGB图像转为灰度图像
- 假设局部区域的灰度变换为线性：梯度方向直方图（SIFT）
- 假设局部区域的像素相对灰度大小不变：LBP，及其变种



SIFT局部区域特征描述

- SIFT描述子由英属哥伦比亚大学Prof. Lowe于1999年提出，发表在ICCV'99，随后扩展期刊版本发表于IJCV'03
 - 两篇论文目前总被引超过8万次
 - SIFT: Scale Invariant Feature Transform



David Lowe
Professor
University of British
Columbia

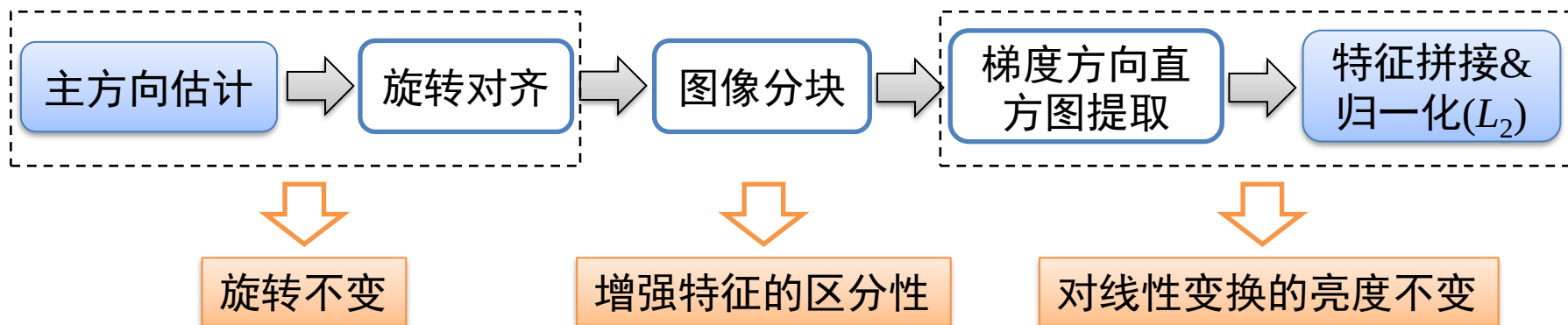
Distinctive image features from scale-invariant keypoints

[DG Lowe](#) - International journal of computer vision, 2004 - Springer

This paper presents a method for extracting distinctive invariant features from images that can be used to perform reliable matching between different views of an object or scene. The features are invariant to image scale and rotation, and are shown to provide robust matching across a substantial range of affine distortion, change in 3D viewpoint, addition of noise, and change in illumination. The features are highly distinctive, in the sense that a single feature can be correctly matched with high probability against a large database of features from ...

☆ 99 被引用次数: 62199 相关文章 所有 156 个版本

□ SIFT描述子的基本流程



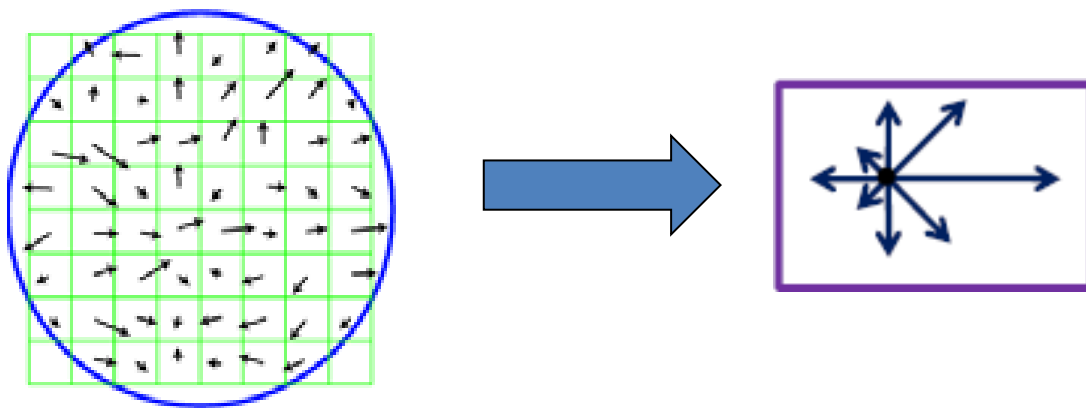
局部区域描述：SIFT描述子

□ 主方向估计

- 目标：检测图像块的主方向(dominant orientation), 用于旋转对齐

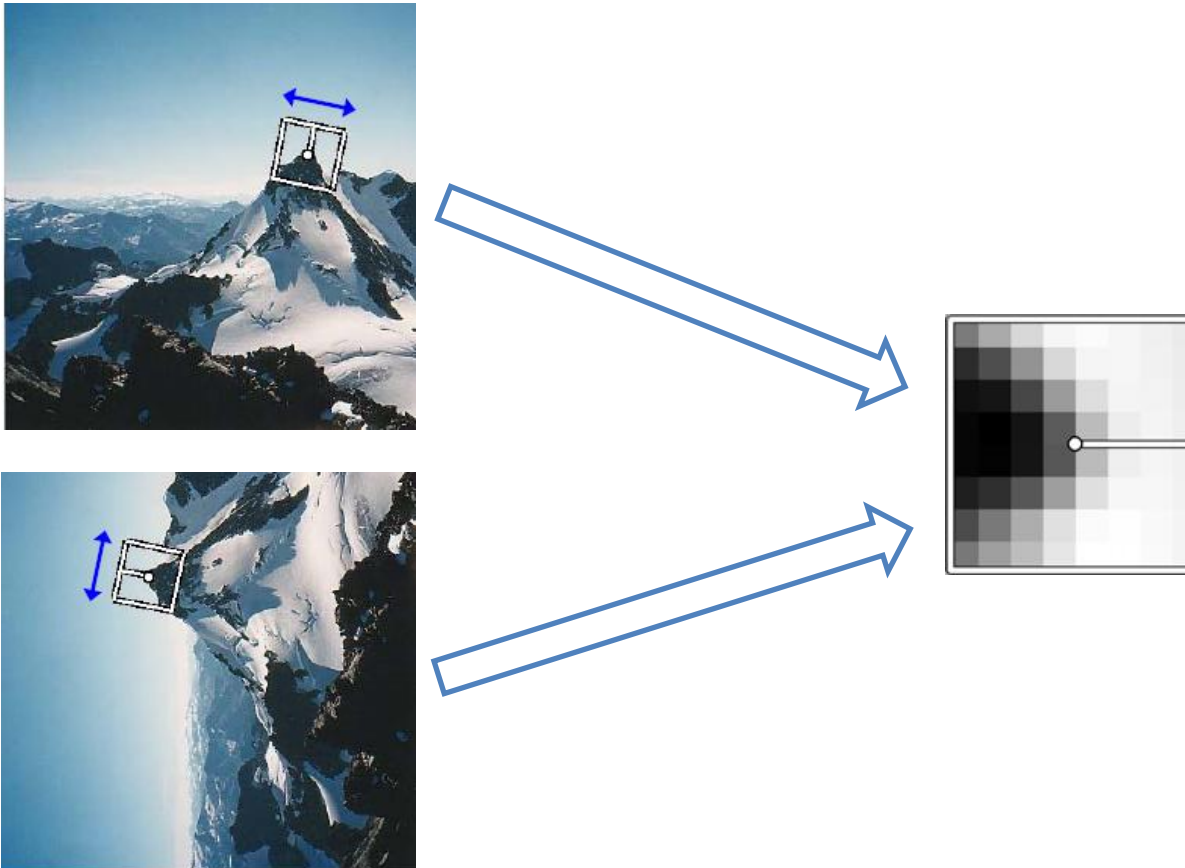
□ 具体过程

- 对图像块计算梯度，构建含36个方向的梯度方向直方图（HOG）
 - ✓ 根据梯度幅值和与特征点的距离，进行高斯加权
- 梯度方向直方图中，峰值对应的bin（方向）为主方向
 - ✓ 其他bin上的值如果大于最大值的80%，也视为主方向
 - ✓ 一个图像块可能可以得到多个主方向



局部区域描述：SIFT描述子

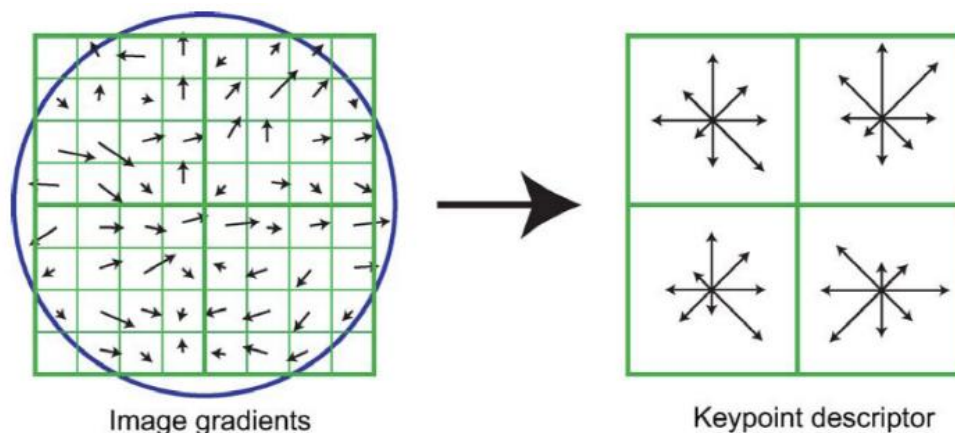
- 基于主方向，对图像块旋转对齐
 - 将图像块修正到一个canonical orientation.



局部区域描述：SIFT描述子

□ SIFT描述子生成

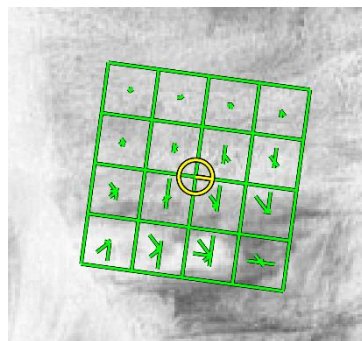
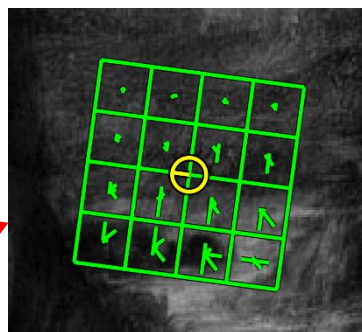
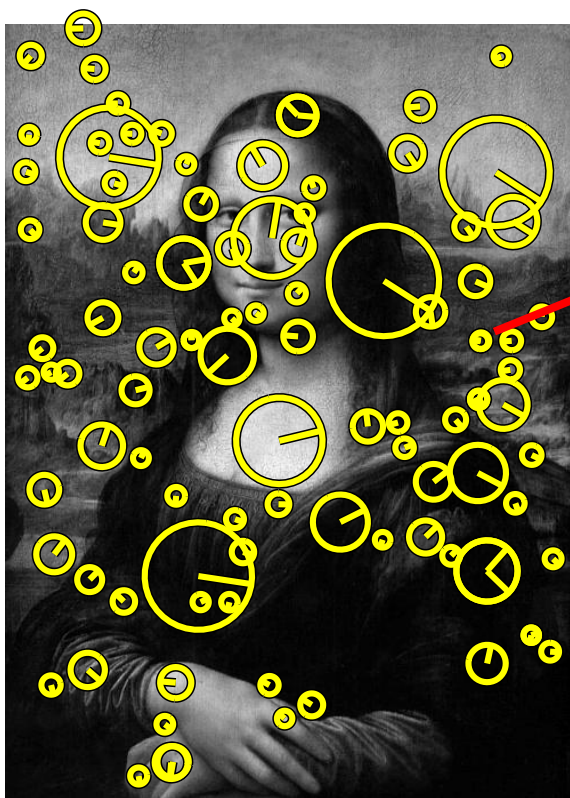
- 将图像区域缩放到固定大小，均分为 4×4 的小块
- 逐块计算梯度方向直方图（8-D）
- 将各块直方图拼接，得到128维的梯度方向直方图
- 最后，对直方图进行 L_2 归一化处理
 - ✓ 实际实现时，会将归一化的特征向量的模设为512，并裁减最大值为255，这样每一维用一个字节即可存储。



示例： 2×2 区域分块，并计算梯度方向直方图。**实际SIFT采用的是 4×4 分块。**

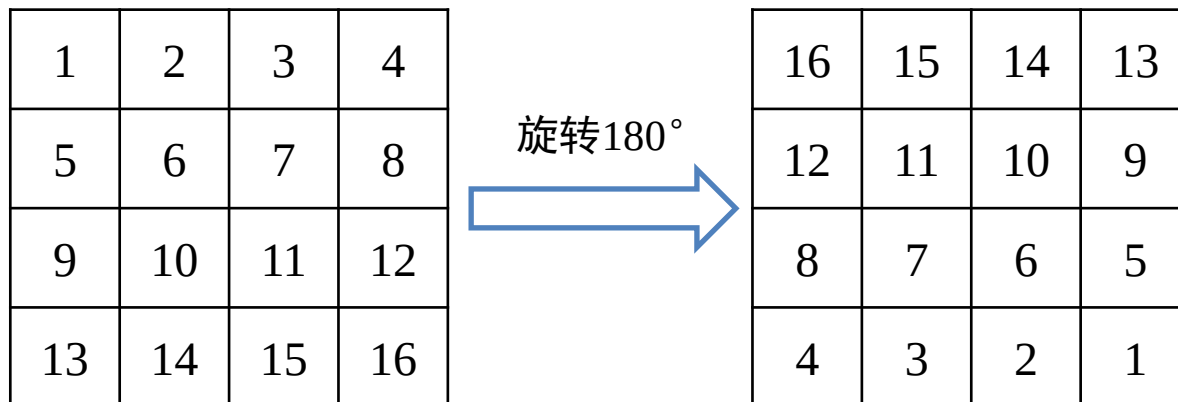
图像反色对SIFT特征提取的影响

- 如果一副灰度图像发生反色变换，即 $f(I) = 255 - I$ ，图像中的SIFT特征（采样DoG检测关键点）会发生什么样的变化？
 - SIFT特征的数量，每个SIFT特征的关键点位置、主方向、特征尺度、128D描述子？



图像反色对SIFT特征提取的影响

- 如果一副灰度图像发生反色变换，即 $f(I) = 255 - I$ ，图像中的SIFT特征会发生什么样的变化？
 - SIFT特征的数量：不变
 - 每个SIFT特征的关键点位置、特征尺度：不变
 - 每个SIFT特征的主方向：相差180度
 - 每个SIFT特征的128D描述子
 - ✓ 16个格子中每个格子对应8维梯度方向直方图不变，但是16个格子的排列顺序颠倒了





图像表达

□ 基于全局特征的图像表达

- Histogram
- Color Name
- GIST

□ 基于局部特征的图像表达

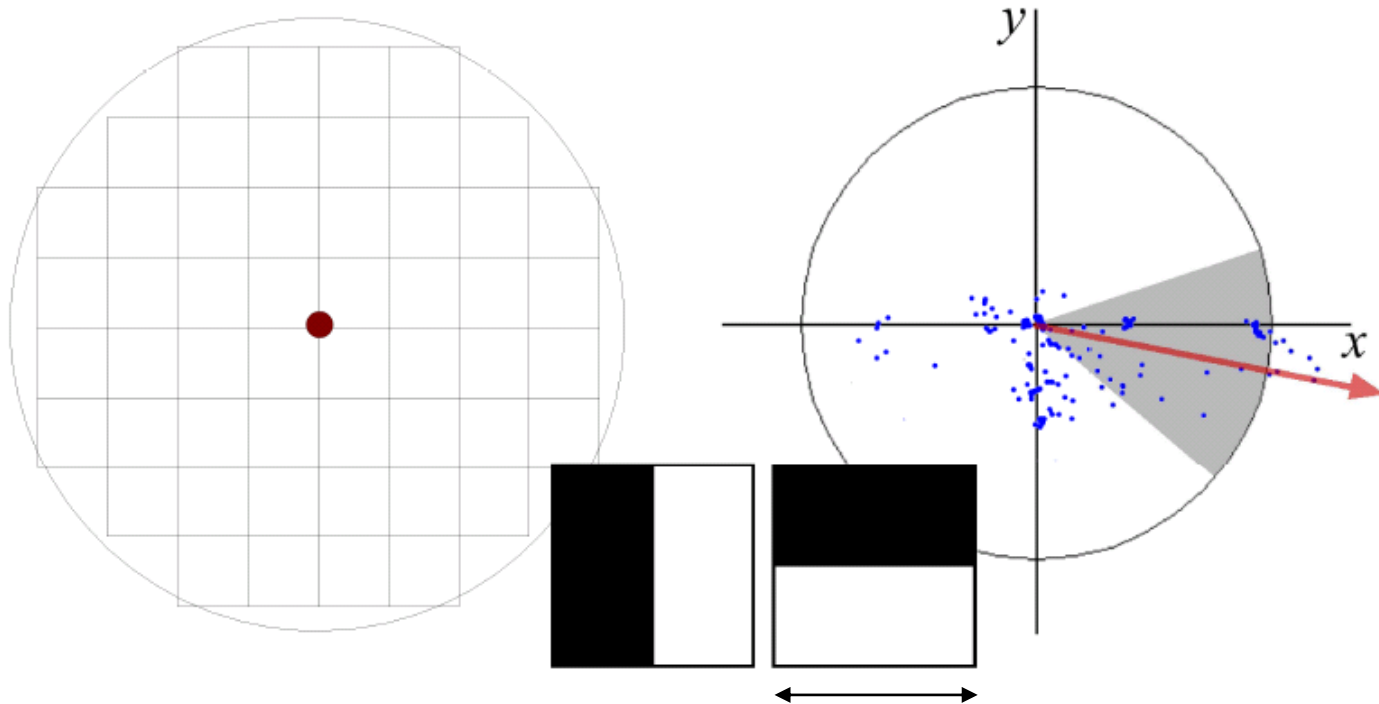
- 简单的局部视觉特征
 - ✓ 局部二值模式 (LBP)
 - ✓ 梯度方向直方图 (HOG)
 - ✓ 形状上下文 (Shape Context)
- 基于关键点的局部视觉特征
 - ✓ 局部视觉特征描述: SIFT, SURF, LIOP, ORB
- 特征编码与聚合
- 乘积量化

SURF特征描述

□ SURF (Speeded Up Robust Features)

■ 主方向估计

- ✓ Circular neighborhood of radius $6s$ around the interest point (s = the scale at which the point was detected)



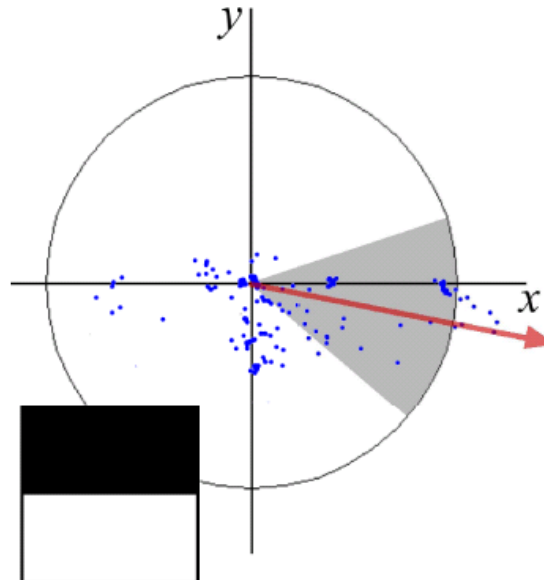
Side length = $4s$, Cost 6 operation
to compute the response

SURF特征描述

□ SURF (Speeded Up Robust Features)

■ 主方向估计

- ✓ The Haar wavelet responses are represented as vectors
- ✓ Sum all responses within a sliding orientation window covering an angle of 60 degree
- ✓ The two summed response yield a new vector
- ✓ The longest vector is the dominant orientation
- ✓ Second longest is ignored

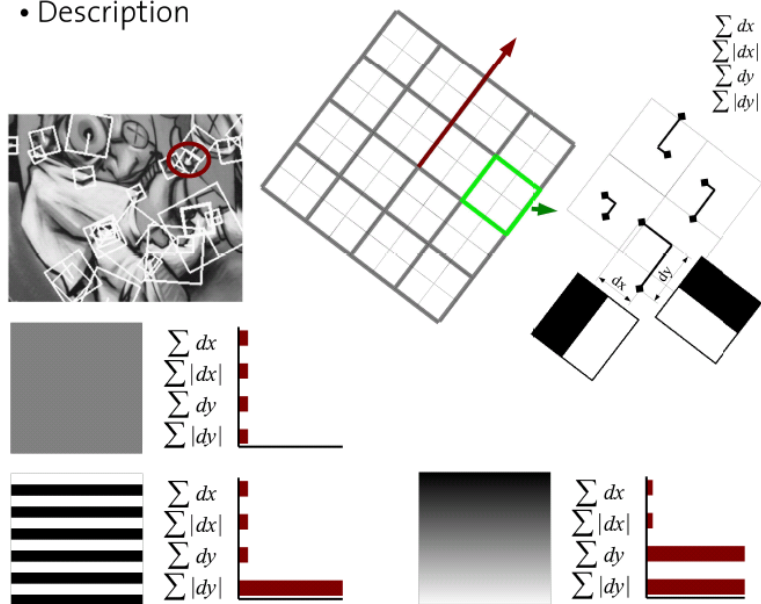


SURF特征描述

□ SURF (Speeded Up Robust Features) 描述子

- 将局部区域均匀划分为 4×4 的方形子区域，每个子区域含 5×5 像素
- 计算每个像素的梯度 d_x 和 d_y ，并用高斯核对梯度值加权
- 计算每个子区域的 d_x 的和(sum)，以及 d_x 的和
 - ✓ 拼接16个子区域的梯度和，得到32维向量
- 为避免梯度正负相消，在求和前，分别计算各像素的 d_x 和 d_y 绝对值
 - ✓ 拼接16个子区域的梯度/绝对梯度的和，共得到64维向量
- 对得到的特征向量进行归一化到单位向量

• Description





图像表达

□ 基于全局特征的图像表达

- Histogram
- Color Name
- GIST

□ 基于局部特征的图像表达

- 简单的局部视觉特征
 - ✓ 局部二值模式 (LBP)
 - ✓ 梯度方向直方图 (HOG)
 - ✓ 形状上下文 (Shape Context)
- 基于关键点的局部视觉特征
 - ✓ 局部视觉特征描述: SIFT, SURF, LIOP, ORB
- 特征编码与聚合
- 乘积量化

局部区域描述：LIOP描述子

□ SIFT特征存在的问题

- 主方向估计存在误差，其影响会在基于主方向对图像块旋转对齐中积累放大
- 定量比较：逆时针旋转图像 θ 度，确定旋转前、后的图像中的对应位置的匹配的SIFT特征，计算其主方向差与 θ 的一致性

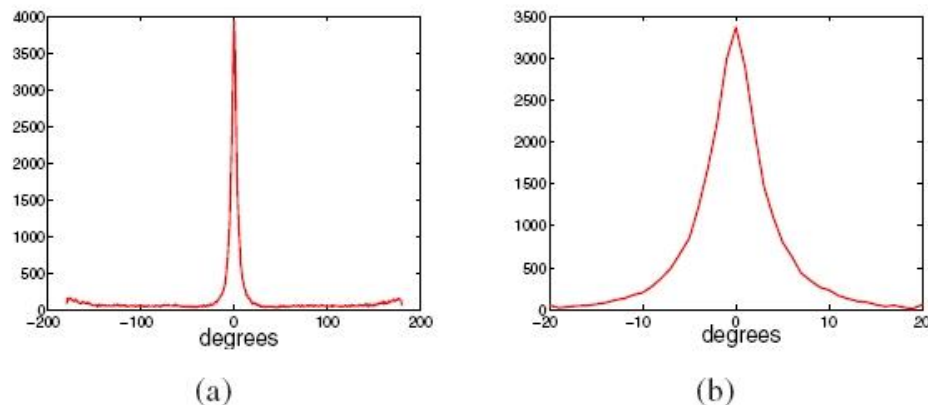


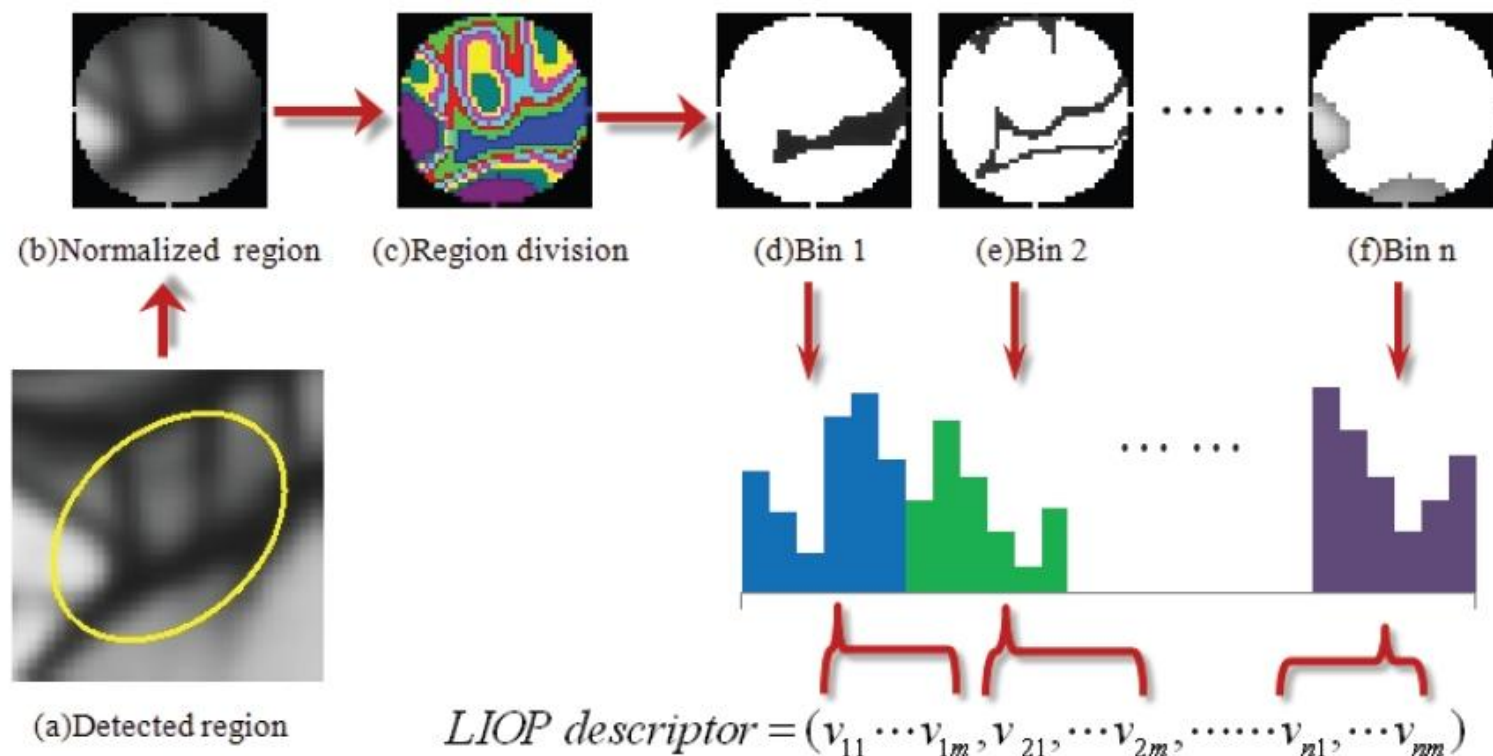
Figure. Orientation assignment errors. (a) Between corresponding points, only 63.77% of errors are in the range of $[-20, 20]$. (b) Between corresponding points that are also matched by SIFT descriptors.

- 可否绕过主方向估计，无需对齐，即可提取特征，同时保持旋转不变性？

局部区域描述：LIOP描述子

□ LIOP特征：自适应分块

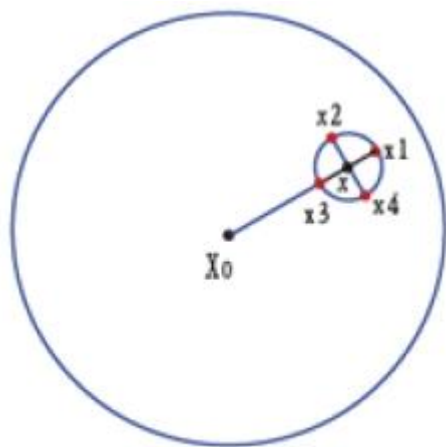
- 基本假设：图像区域中像素灰度的相对大小关系保持不变
- 利用像素的灰度值排序，将图像划分为n个不规则的区域
- 每个区域提取一个m维的特征向量，最后将所有区域的特征进行拼+立



局部区域描述：LIOP描述子

□ LIOP特征：旋转不变性特征

- 基于极坐标空间，提取各像素梯度
- 类似于LBP，基于相对大小关系，将梯度值变为一种索引编号
- 对每一个子区域，计算像素的索引编号的直方图



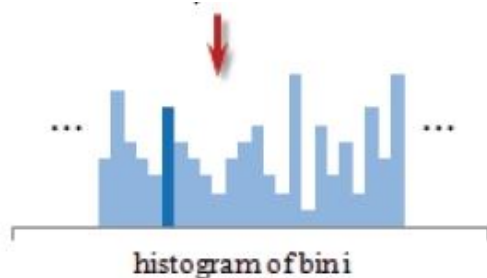
(a) Original patch

For a point x:

$$P(x) = (I(x_1), I(x_2), I(x_3), I(x_4)) \\ = (86, 217, 152, 101)$$

$$\gamma(P(x)) = (1, 4, 3, 2)$$

$$LIOP(x) = \phi(\gamma(P(x))) \\ = (0, 0, 0, 0, 0, 1, 0, \dots, 0)$$



π	$Ind(\pi)$
1,2,3,4	1
1,2,4,3	2
1,3,2,4	3
1,3,4,2	4
1,4,2,3	5
1,4,3,2	6
2,1,3,4	7
2,1,4,3	8
...	...
4,3,1,2	23
4,3,2,1	24



图像表达

□ 基于全局特征的图像表达

- Histogram
- Color Name
- GIST

□ 基于局部特征的图像表达

- 简单的局部视觉特征
 - ✓ 局部二值模式 (LBP)
 - ✓ 梯度方向直方图 (HOG)
 - ✓ 形状上下文 (Shape Context)
- 基于关键点的局部视觉特征
 - ✓ 局部视觉特征描述: SIFT, SURF, LIOP, ORB
- 特征编码与聚合
- 乘积量化

ORB局部二值特征

□ ORB: Oriented FAST and Rotated BRIEF

- 主方向：将像素视为密度，质心相对几何中心的偏移
 - ✓ 将像素的灰度值视为相对局部区域中心的偏移量(offset)
 - ✓ 计算局部区域的灰度矩：

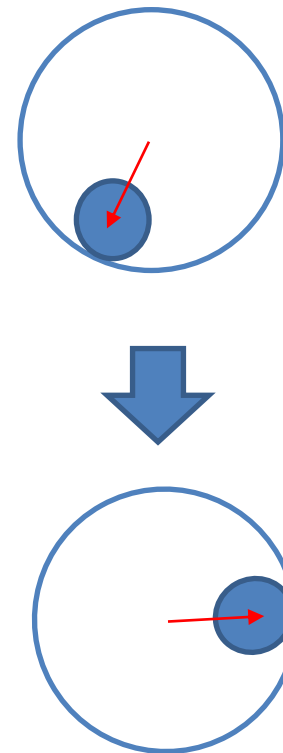
$$m_{pq} = \sum_{x,y} x^p y^q I(x,y)$$

- ✓ 定义局部区域的灰度质心：

$$C = \left(\frac{m_{10}}{m_{00}}, \frac{m_{01}}{m_{00}} \right)$$

- ✓ 构建从局部区域几何中心点到灰度质心点的向量：

$$\theta = \text{atan2}(m_{01}, m_{10})$$



ORB局部二值特征

□ 主方向计算方法

■ 特征点局部梯度

$$\mathbf{g}(\mathbf{p}_i, \mathbf{p}_j) = (\mathbf{p}_j - \mathbf{p}_i) \cdot \frac{I(\mathbf{p}_j, \sigma_j) - I(\mathbf{p}_i, \sigma_i)}{\|\mathbf{p}_j - \mathbf{p}_i\|^2}.$$

■ 定义短距离点对子集、长距离点对子集：

$$\mathcal{A} = \{(\mathbf{p}_i, \mathbf{p}_j) \in \mathbb{R}^2 \times \mathbb{R}^2 \mid i < N \wedge j < i \wedge i, j \in \mathbb{N}\}$$

$$\mathcal{S} = \{(\mathbf{p}_i, \mathbf{p}_j) \in \mathcal{A} \mid \|\mathbf{p}_j - \mathbf{p}_i\| < \delta_{max}\} \subseteq \mathcal{A}$$

$$\mathcal{L} = \{(\mathbf{p}_i, \mathbf{p}_j) \in \mathcal{A} \mid \|\mathbf{p}_j - \mathbf{p}_i\| > \delta_{min}\} \subseteq \mathcal{A}.$$

■ 局部梯度均值： $\mathbf{g} = \begin{pmatrix} g_x \\ g_y \end{pmatrix} = \frac{1}{L} \cdot \sum_{(\mathbf{p}_i, \mathbf{p}_j) \in \mathcal{L}} \mathbf{g}(\mathbf{p}_i, \mathbf{p}_j).$

■ 主方向： $\alpha = \arctan2(g_y, g_x)$

□ 二值特征生成

$$b = \begin{cases} 1, & I(\mathbf{p}_j^\alpha, \sigma_j) > I(\mathbf{p}_i^\alpha, \sigma_i) \\ 0, & \text{otherwise} \end{cases} \quad \forall (\mathbf{p}_i^\alpha, \mathbf{p}_j^\alpha) \in \mathcal{S}$$



图像表达

□ 基于全局特征的图像表达

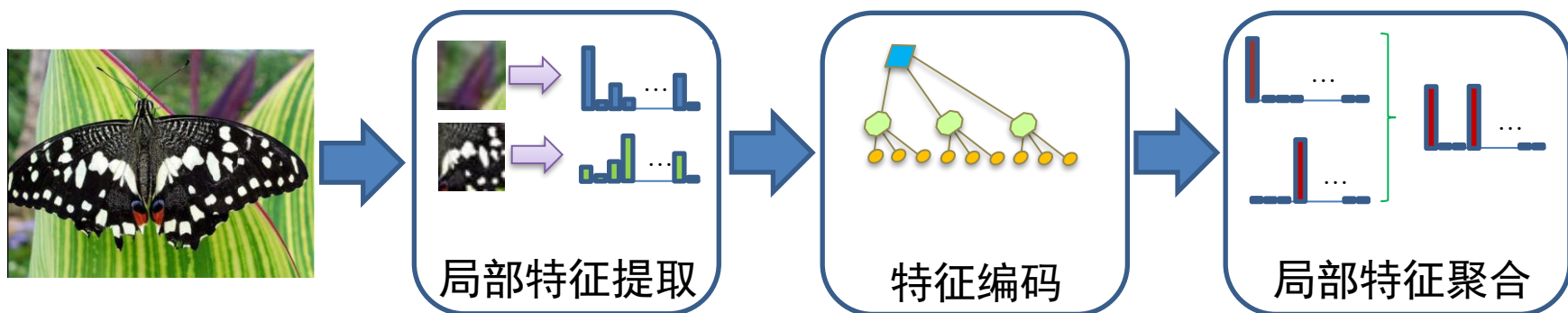
- Histogram
- Color Name
- GIST

□ 基于局部特征的图像表达

- 简单的局部视觉特征
 - ✓ 局部二值模式 (LBP)
 - ✓ 梯度方向直方图 (HOG)
 - ✓ 形状上下文 (Shape Context)
- 基于关键点的局部视觉特征
 - ✓ 局部视觉特征描述: SIFT, SURF, LIOP, ORB
- 特征编码与聚合
- 乘积量化

基于局部特征的图像表达基本框架

- 在图像分析中，许多问题可归结为图像间的比较
- **局部视觉特征**对视觉内容有良好的区分性和表达力
 - 区分性：可辨别类间差异
 - 表达力：可包容类内差异
 - 相似视觉内容的视觉特征向量距离小，不同视觉内容的视觉特征向量距离大
- 特征聚合
 - 局部视觉特征维度高，且不同图像中包含的视觉特征**数量**不同
 - 有必要对一副图像中的局部特征进行量化（编码）和聚合，生成一个**固定长度**的矢量表达



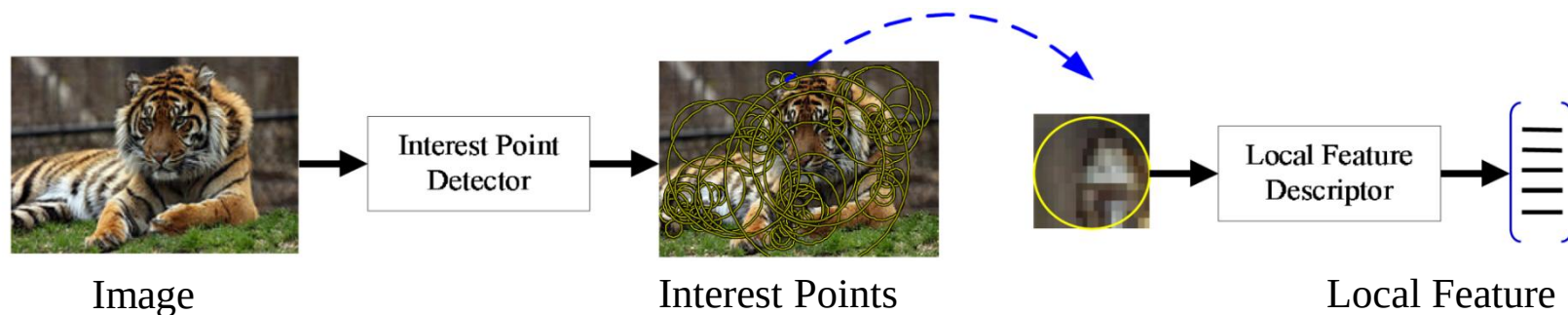
局部视觉特征

□ 视觉特征的不变性 (invariance)

- 以不变应万变：图像在**像素空间**发生了**亮度变换**或**几何变换**，但是在**特征空间**没有变化，或变化很小
- 常见的不变性
 - ✓ 亮度不变性，平移不变性，旋转不变性，尺度（缩放）不变性

□ 局部视觉特征提取

- 局部关键点检测
 - ✓ 确定关键点的**位置**和对应局部区域的**大小**
- 局部区域描述
 - ✓ 将图像局部区域**像素信号**变换为固定维度的**特征向量**





Fisher Vector Representation

□ 基本思想

- 将图像中的每个局部视觉特征当作一个服从某种分布的随机变量
- 将一组特征表示为一个向量
 - ✓ 对数似然函数的梯度

□ 生成模型

$$\nabla_{\lambda} \log p(X | \lambda)$$

$X = \{x_t\}, (t = 1, \dots, T)$,
 $p(\cdot)$: the PDF function,
 λ : a set of parameters

□ 假设特征之间相互独立

$$L(X|\lambda) = \log p(X|\lambda) = \log \prod_{t=1}^T p(x_t|\lambda) = \sum_{t=1}^T \log p(x_t|\lambda)$$

Fisher Vector Representation

■ 用混合高斯模型GMM对分布进行近似

$$p(x_t | \lambda) = \sum_{i=1}^N w_i p_i(x_t | \lambda)$$

$$p_i(x | \lambda) = \frac{\exp\{-\frac{1}{2}(x - \mu_i)^T \Sigma_i^{-1}(x - \mu_i)\}}{(2\pi)^{D/2} |\Sigma_i|^{1/2}}$$

$$\sum_{i=1}^N w_i = 1$$

$$\lambda = \{w_i, \mu_i, \Sigma_i, i = 1, \dots, N\}$$

■ 计算对数似然函数关于各参数的梯度值：

$$\nabla_{\alpha_i} \log p(X | \lambda) = \sum_{t=1}^T \left[\frac{\gamma_t(i)}{w_i} - \frac{\gamma_t(1)}{w_1} \right]$$

$$\nabla_{\mu_i^d} \log p(X | \lambda) = \sum_{t=1}^T \gamma_t(i) \left[\frac{x_t^d - \mu_t^d}{(\sigma_i^d)^2} \right]$$

$$\nabla_{\sigma_i^d} \log p(X | \lambda) = \sum_{t=1}^T \gamma_t(i) \left[\frac{(x_t^d - \mu_i^d)^2}{(\sigma_i^d)^2} - \frac{1}{\sigma_i^d} \right]$$

$$w_k = \frac{\exp(\alpha_k)}{\sum_{j=1}^K \exp(\alpha_j)}$$

$$\gamma_t(i) = p(i | x_t, \lambda) = \frac{w_i p_i(x_t | \lambda)}{\sum_{j=1}^N w_j p_j(x_t | \lambda)}$$

$$w_i = \frac{\exp(\alpha_k)}{\sum_{j=1}^N \exp(\alpha_j)}$$



FV与BoW和VLAD的关系

□ 词袋模型 (BOW: bag-of-words model)

- The gradient with respect to the weight of GMM:

$$\nabla_{\alpha_i} \log p(X | \lambda) = \sum_{t=1}^T \left[\frac{\gamma_t(i)}{w_i} - \frac{\gamma_t(1)}{w_1} \right]$$

- Soft version of BoW:

- **Sparseness** $b_i \propto \sum_{t=1}^T \gamma_t(i)$ for the i -th visual word
✓ Suitable for the inverted index

□ VLAD (vector of locally aggregated descriptor)

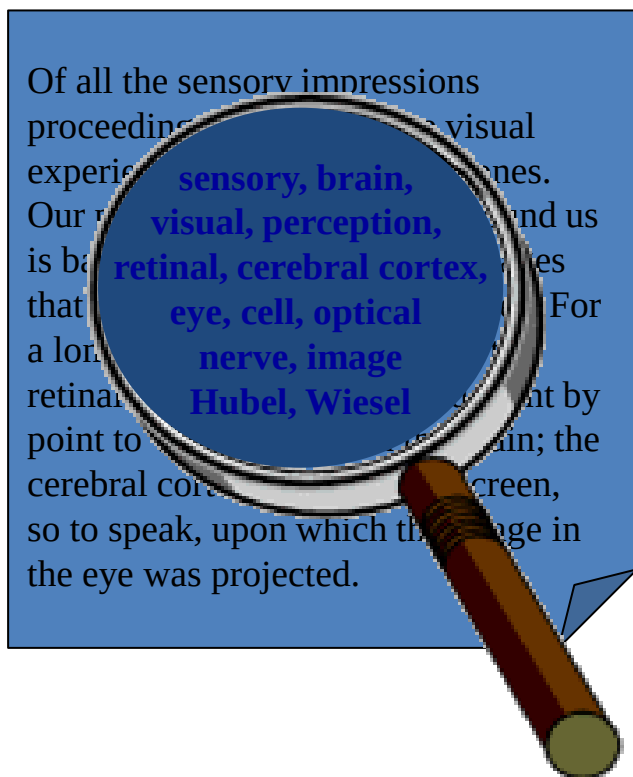
- The gradient with respect to the mean vector of GMM

- Disadvantage $\nabla_{\mu_i^d} \log p(X | \lambda) = \sum_{t=1}^T \gamma_t(i) \left[\frac{x_t^d - \mu_t^d}{(\sigma_i^d)^2} \right]$
✓ **Non-sparse**: unsuitable to apply the inverted index

词袋模型 (Bag-of-Words Model)

- 在信息检索中，基于**文本单词字典**表达文本文档
 - 紧凑性
 - 描述力

Bag-of-Word model



Retrieve



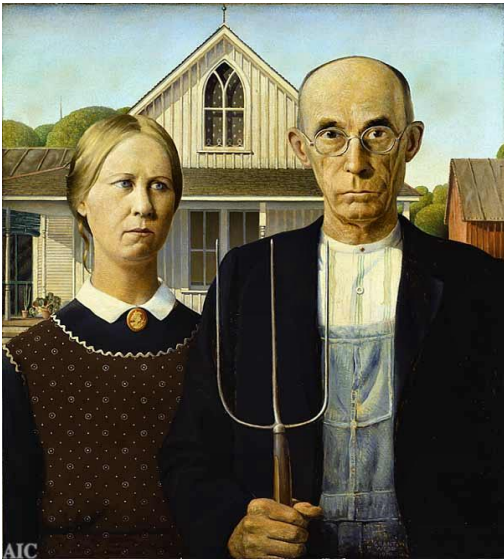
视觉词袋模型(Bag of Visual Words)

□ 图像也表达为Bag-of-Visual Words?

Image



Bag of 'visual words'



?



视觉词袋模型(Bag of Visual Words)



□ 三个基本问题

■ Q1: 如何定义视觉单词和视觉码本？

- ✓ A1: 对大量的局部视觉特征进行聚类（如k-means），聚类中心视为视觉单词，所有的聚类中心构成了视觉码本。

■ Q2: 如何将每个局部视觉特征与视觉单词对应？

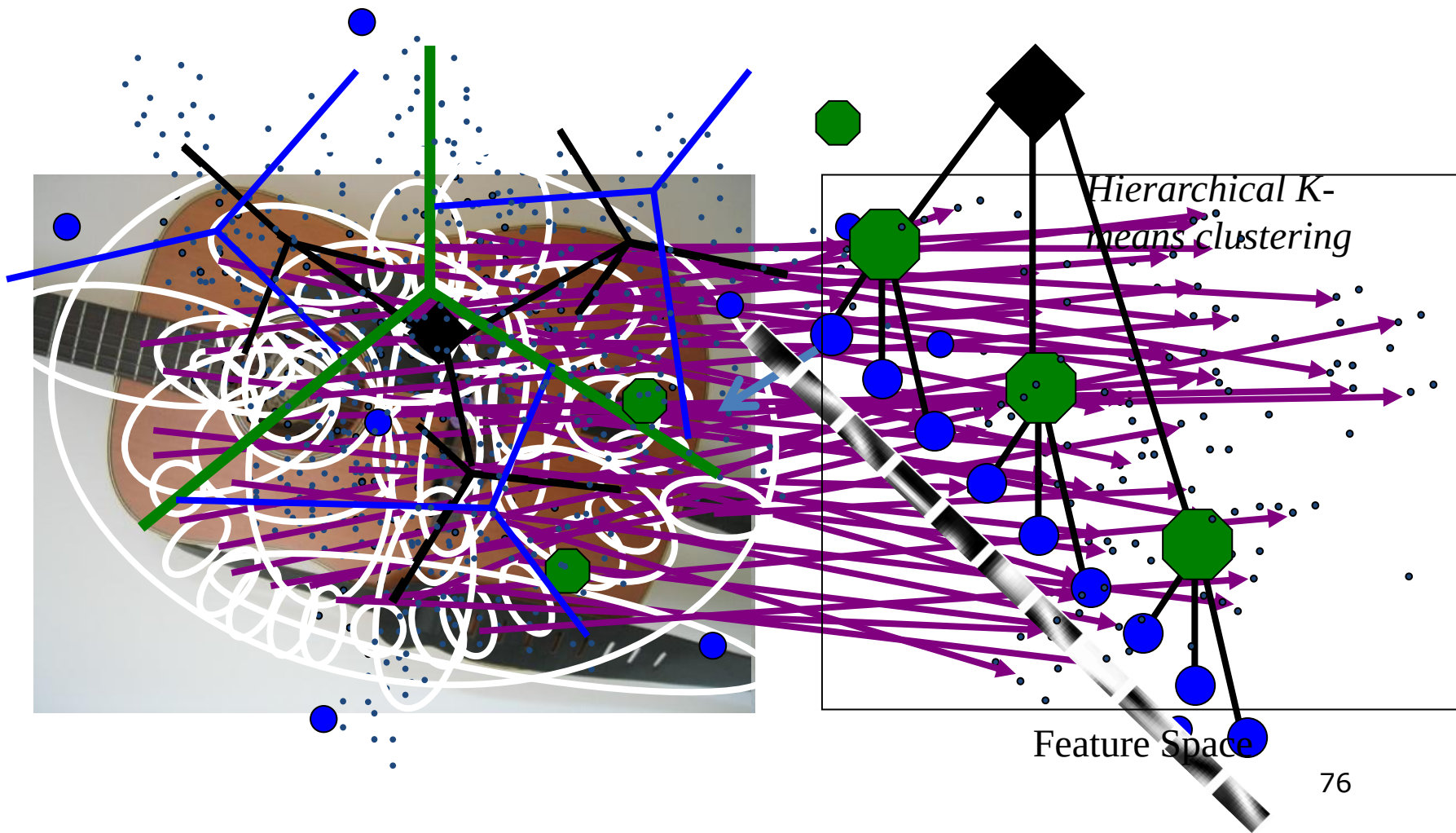
- ✓ A2: 基于视觉码本，通过矢量量化，将局部视觉特征量化到最近的视觉单词。
- ✓ 有损压缩 (lossy compression)，表达紧凑。

■ Q3: 如何基于对一幅图像的局部视觉特征集合进行紧凑表达？

- ✓ A2: 基于视觉单词在图像中出现的频率，构造视觉单词直方图

基于层级聚类的视觉码本生成

□ 关键点检测和局部特征提取



视觉词袋模型：特征向量量化与匹配

□ 视觉码本的物理意义

- 通过视觉特征向量聚类，将特征空间划分为若干子空间，
- 每个聚类中心向量可视为一个视觉单词，对应一个上述子空间
- 所有的视觉单词的集合，构成视觉码本

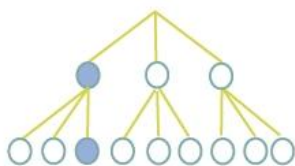
□ 通过矢量量化，实现特征匹配

- 给定一个视觉特征向量，矢量量化从视觉码本中返回与其最近的视觉单词的编号： $q(\mathbf{x}) = \arg \min_k \|\mathbf{v}_k - \mathbf{x}\|$
- 通过矢量量化，可知视觉特征向量落在哪个子空间

□ 高效的矢量量化方法

- Hierarchical k-means [Nister 06]

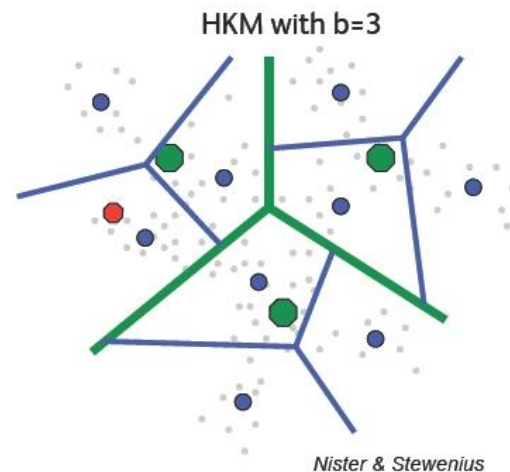
► K-means tree of height h



► Branching factor b : $k = b^h$

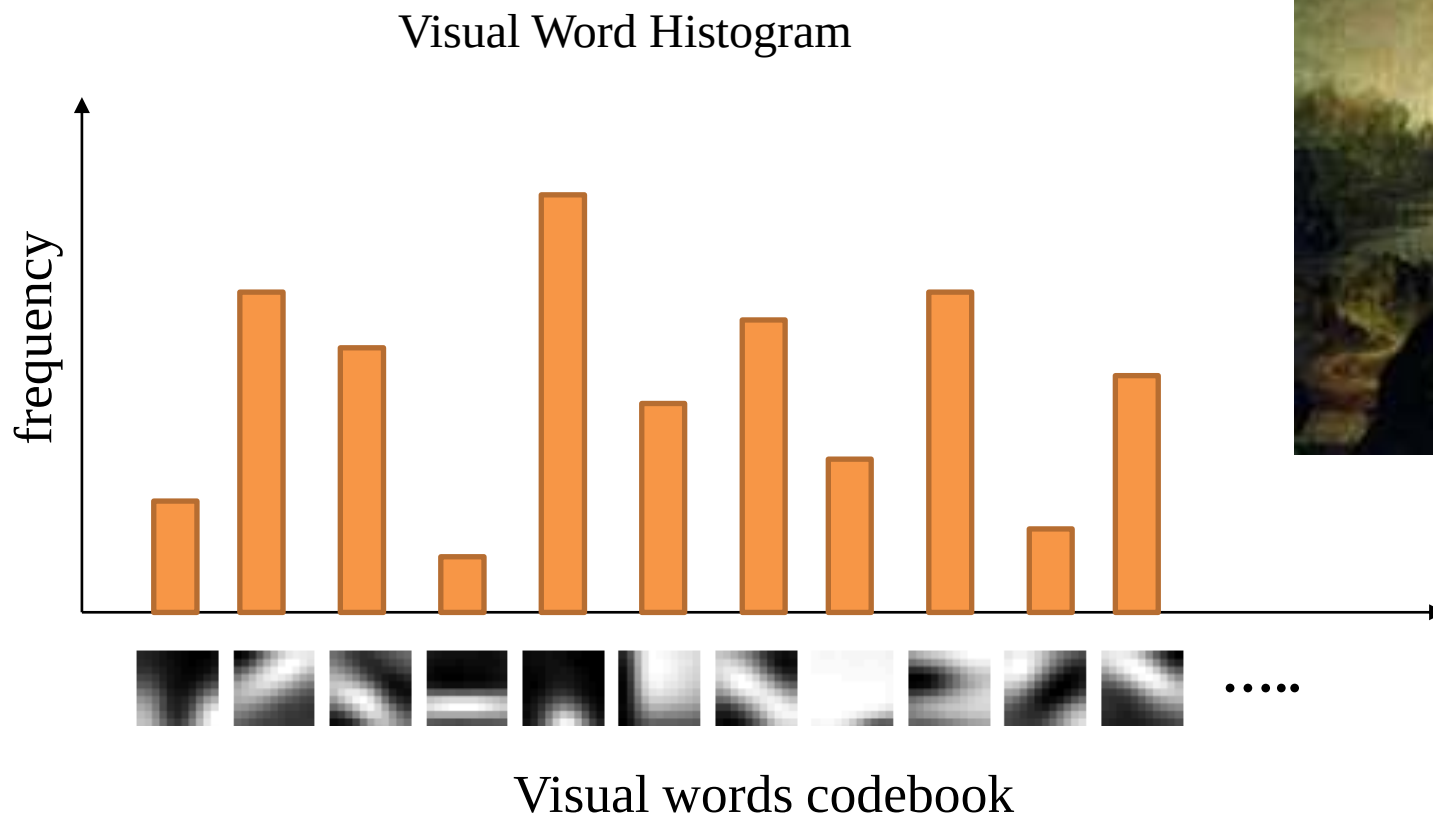
► Assignment Complexity:

$$\mathcal{O}(d h b) = \mathcal{O}(d h k^{\frac{1}{h}})$$



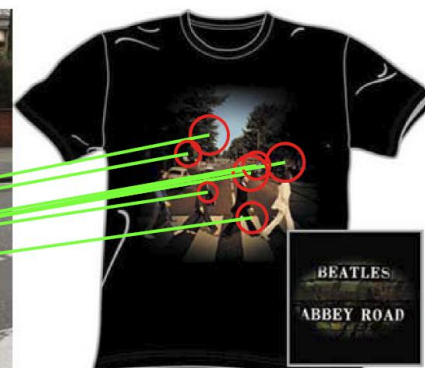
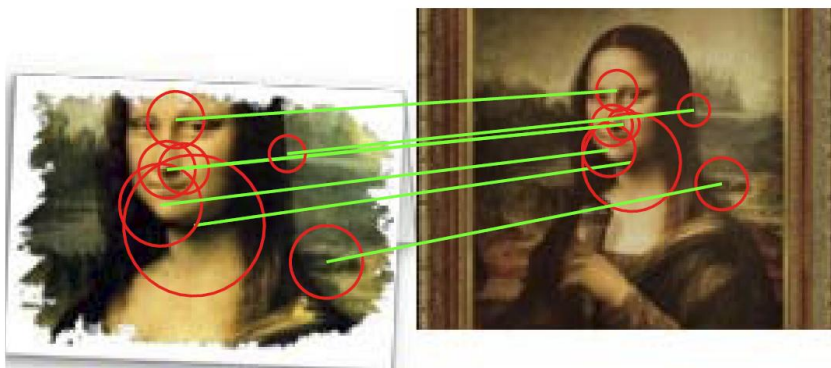
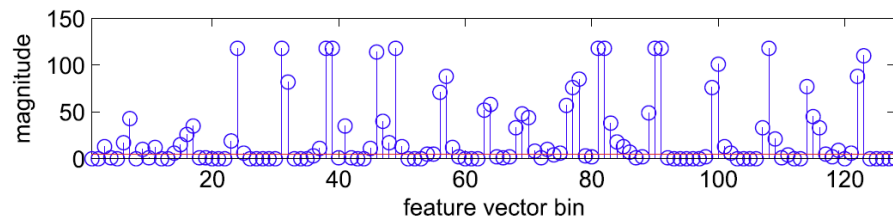
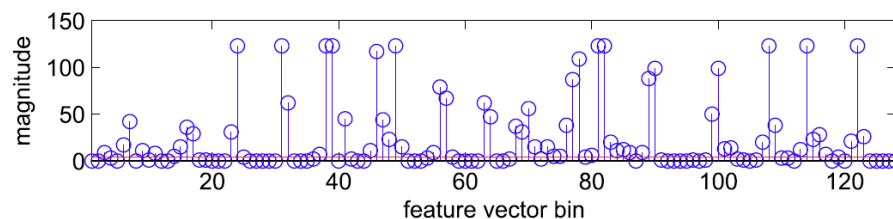
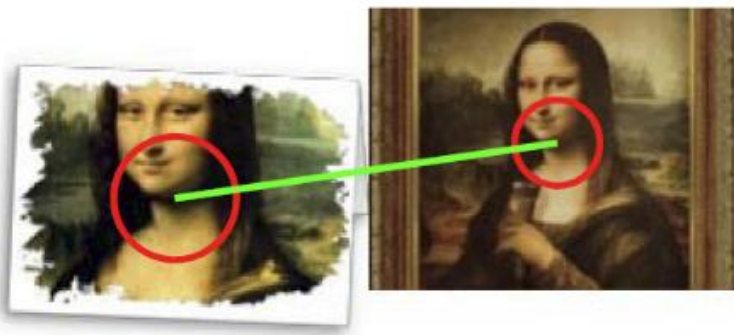
视觉词袋模型

□ 将一副图像表达为视觉单词的分布直方图



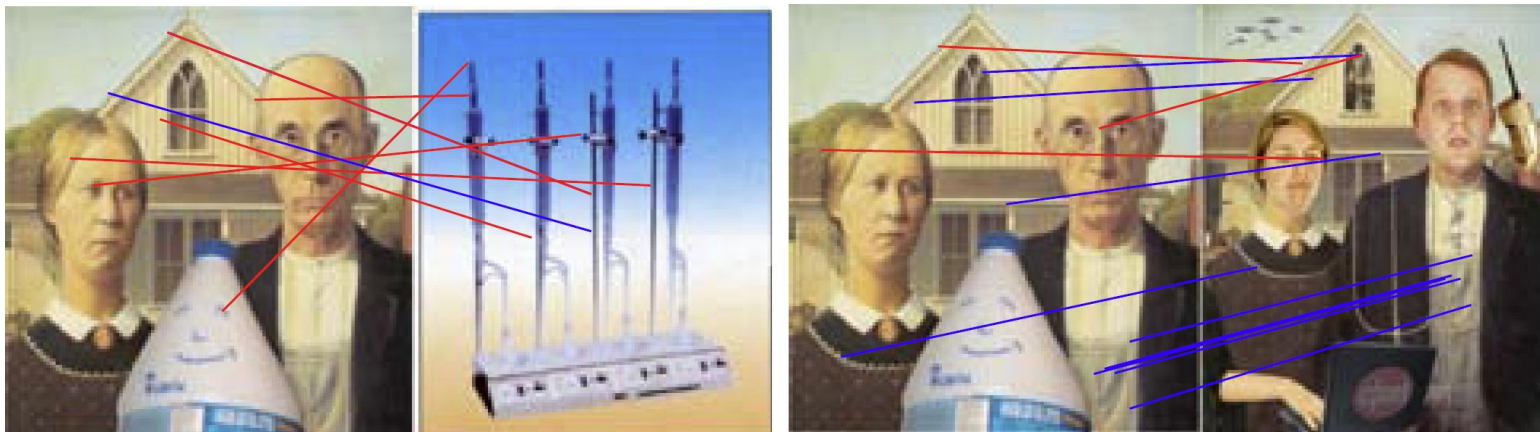
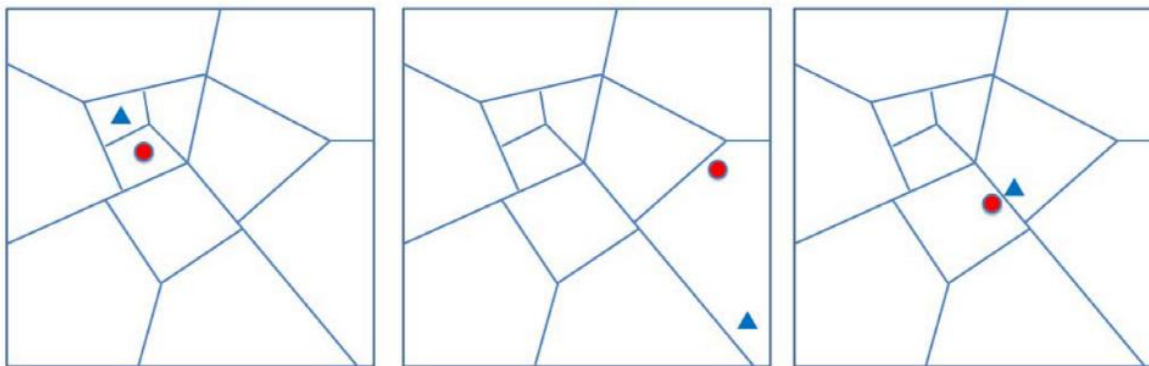
基于词袋模型的图像匹配-I

- 来自两幅图像中的SIFT特征被量化到同一个视觉单词，则构成两图之间的一对局部匹配



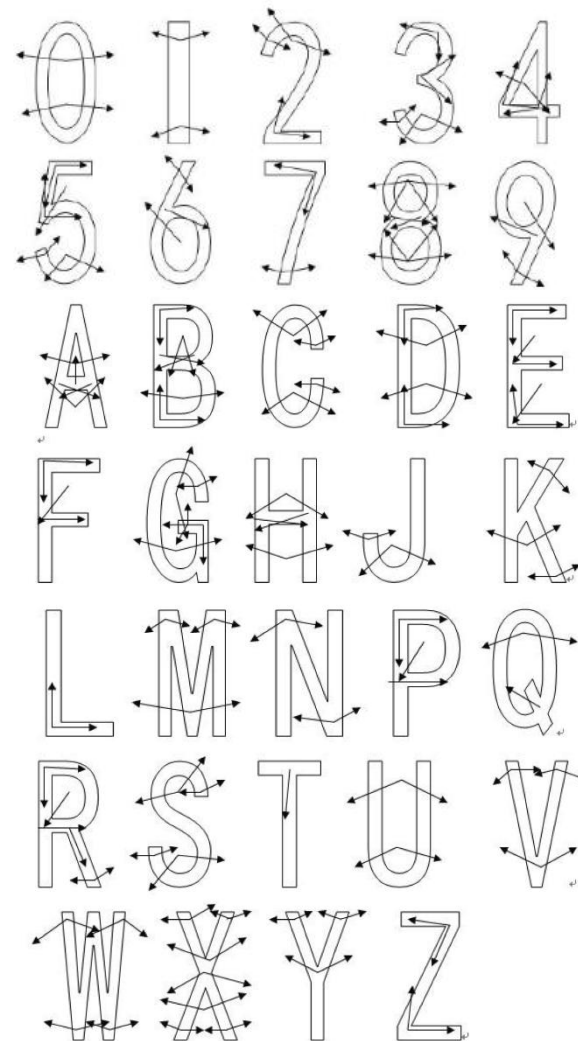
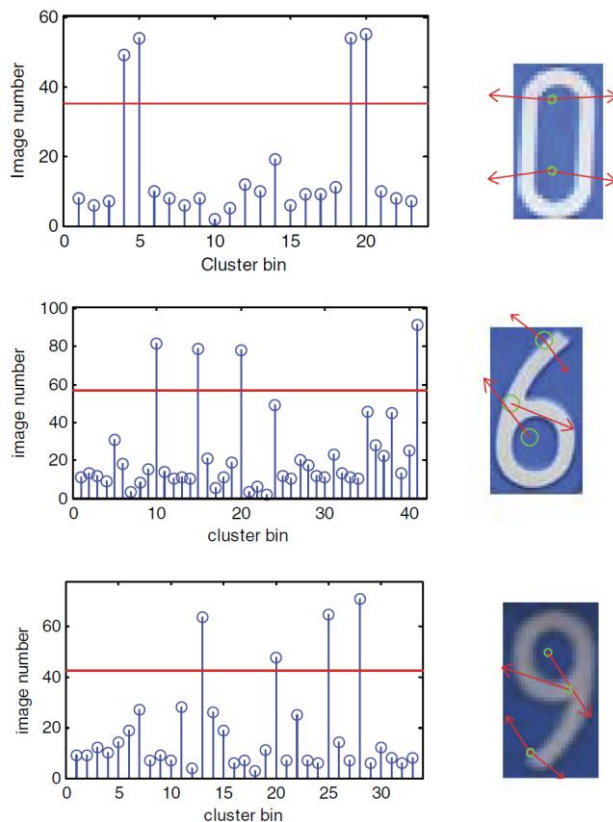
基于词袋模型的图像匹配-II

- 来自两幅图像中的SIFT特征被量化到同一个视觉单词，则构成两图之间的一对局部匹配
 - 硬量化可能会产生错误匹配



应用：车牌字符的视觉单词-I

- 对车牌中的各个字符收集图像样本，对其中的SIFT特征进行AP聚类，保留较大的类簇，作为其视觉单词：



- Wengang Zhou, Houqiang Li, Yijuan Lu, and Qi Tian, “Principal Visual Word Discovery for Automatic License Plate Detection,” *IEEE Transactions on Image Processing (TIP)*, vol. 21, no. 6, pp. 4269-4279, 2012.

应用：车牌字符的视觉单词-I

□ 基于车牌主视觉单词的车牌检测



(a)



(b)



(c)



(d)

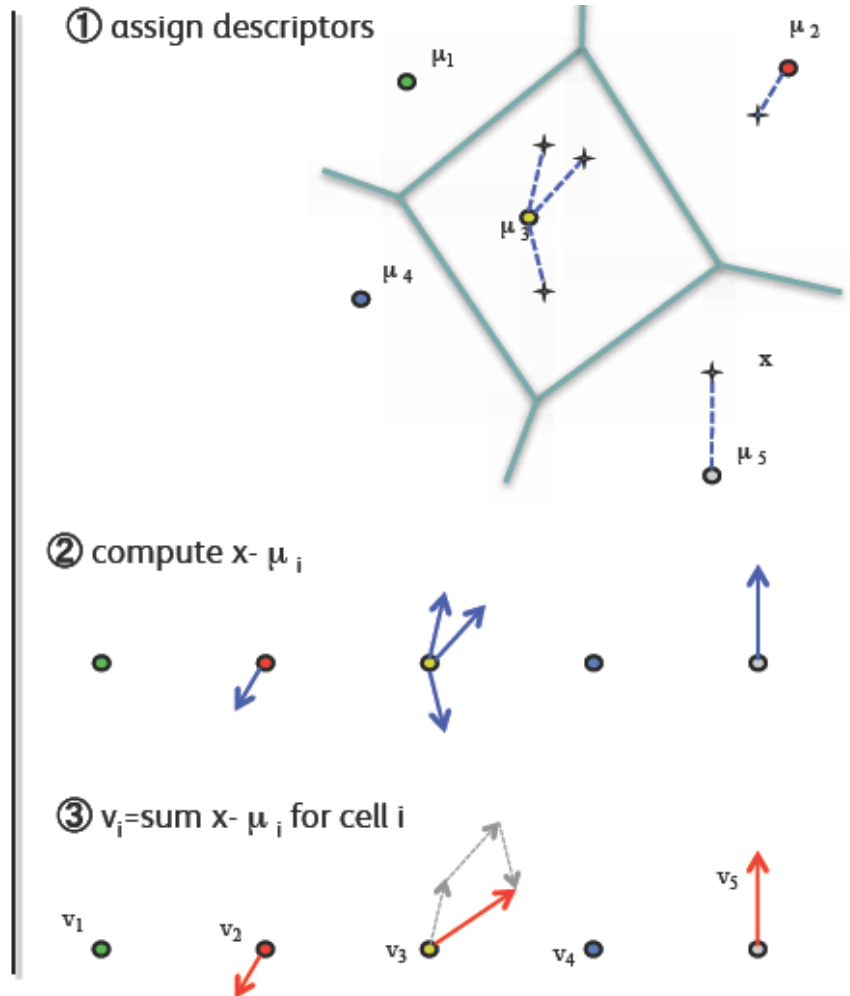
- Wengang Zhou, Houqiang Li, Yijuan Lu, and Qi Tian, “Principal Visual Word Discovery for Automatic License Plate Detection,” *IEEE Transactions on Image Processing (TIP)*, vol. 21, no. 6, pp. 4269-4279, 2012.

局部聚合描述子向量

(VLAD: Vector of Locally Aggregated Descriptor)

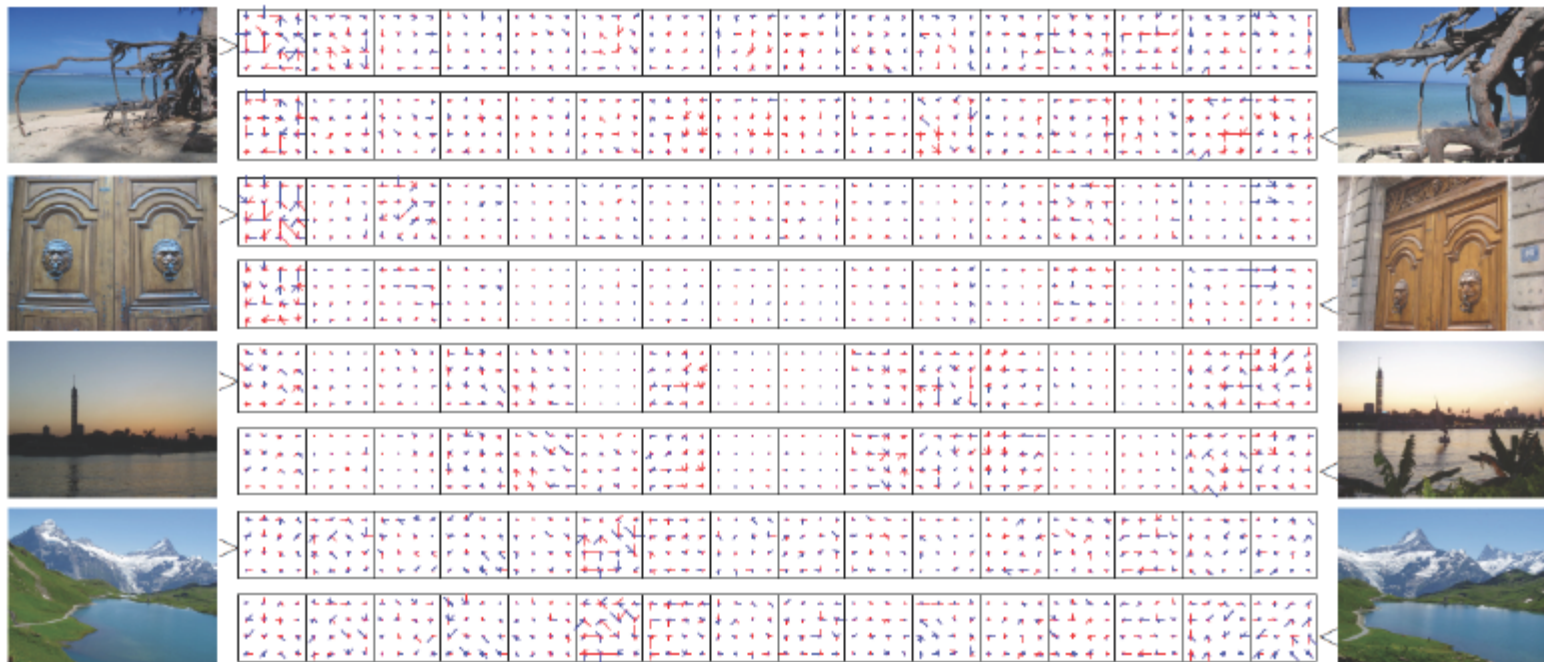
Given a codebook $\{\mu_i, i = 1 \dots N\}$,
e.g. learned with K-means, and a set of
local descriptors $X = \{x_t, t = 1 \dots T\}$:

- ① assign: $\text{NN}(x_t) = \arg \min_{\mu_i} \|x_t - \mu_i\|$
- ②③ compute: $v_i = \sum_{x_t: \text{NN}(x_t) = \mu_i} x_t - \mu_i$
- concatenate v_i 's + ℓ_2 normalize



局部聚合描述子向量：实例

A graphical representation of $v_i = \sum_{x_t: \text{NN}(x_t) = \mu_i} x_t - \mu_i$





图像表达

□ 基于全局特征的图像表达

- Histogram
- Color Name
- GIST

□ 基于局部特征的图像表达

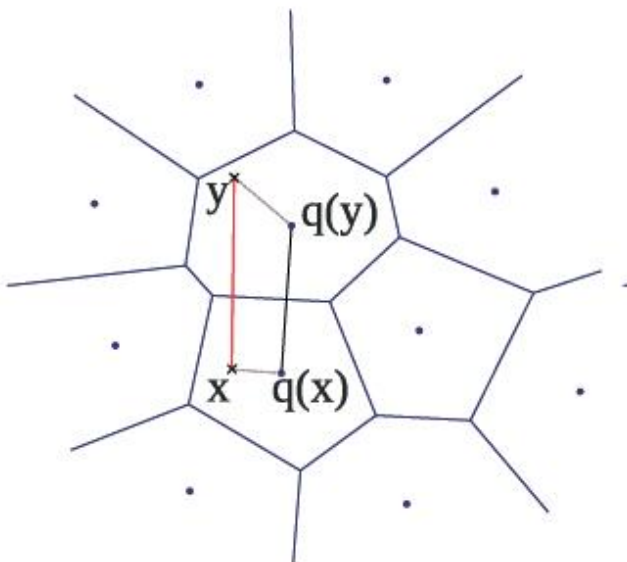
- 简单的局部视觉特征
 - ✓ 局部二值模式 (LBP)
 - ✓ 梯度方向直方图 (HOG)
 - ✓ 形状上下文 (Shape Context)
- 基于关键点的局部视觉特征
 - ✓ 局部视觉特征描述: SIFT, SURF, LIOP, ORB
- 特征编码与聚合
- 乘积量化

乘积量化 (Product Quantization)



□ 动机

- 给定码本，对高维特征矢量进行**矢量量化**，用**码字**来替换原始的特征矢量，可减小存储开销
- 由于码字矢量间距离可以事先计算，高维特征矢量间的距离可以用其对应的码字之间的距离来近似
- 显然：当码本越大时，每个码字对应的子空间越小，量化误差也越小，上述距离近似的精度越高



乘积量化 (Product Quantization)



□ 基本思想：对特征空间进行极度精细的划分，减小量化误差

■ 将特征向量划分为 m 段子向量

$$\underbrace{x_1, \dots, x_{D^*}, \dots, x_{D-D^*+1}, \dots, x_D}_{u_1(x)} \rightarrow q_1(u_1(x)), \dots, q_m(u_m(x)),$$

■ 对每一段子向量 $u_i(\mathbf{x})$ ，用一个量化器 $q(\cdot)$ 进行量化

✓ 在 D^* 维空间，通过聚类生成一组聚类中心向量 $\{\mathbf{v}_1, \dots, \mathbf{v}_K\}$

✓ 相当于：每一段子向量对应的 D^* 维空间被划分为 K 个子空间

✓ 等价于：整个 D 维特征空间被划分为 K^m 个子空间

✓ 定义量化器： $q(u_i(\mathbf{x})) = \arg \min_k \|\mathbf{v}_k - u_i(\mathbf{x})\|, (k = 1, \dots, K)$

➤ 各段子向量维度相同，可独立学习量化器，或共享同一个量化器

□ 乘积量化的优势：

■ 量化计算复杂度低，可轻松实现对特征空间的极度精细的划分

乘积量化 (Product Quantization)



□ 向量距离近似计算

$$\|\mathbf{x} - \mathbf{y}\|_2^2 = \sum_{i=1}^D (x_i - y_i)^2 = \sum_{k=1}^m \|u_k(\mathbf{x}) - u_k(\mathbf{y})\|_2^2 \approx \sum_{k=1}^m \|\mathbf{c}(a_k) - \mathbf{c}(b_k)\|_2^2 = \sum_{k=1}^m d_{a_k b_k}$$

where $a_k = q_k(u_k(\mathbf{x}))$, $b_k = q_k(u_k(\mathbf{y}))$

↙
 $u_k(\mathbf{x})$ 被量化到第 a_k 个聚类中心向量



□ 预先计算好的距离表

- 聚类中心两两之间距离
- 在线计算向量距离时，仅需查表
- 计算复杂度从 $O(D)$ 降为 $O(m)$

d_{11}	d_{12}	...	d_{1K}
$\vdots$			
d_{K1}			d_{KK}

乘积量化 (Product Quantization)



□ 误差分析

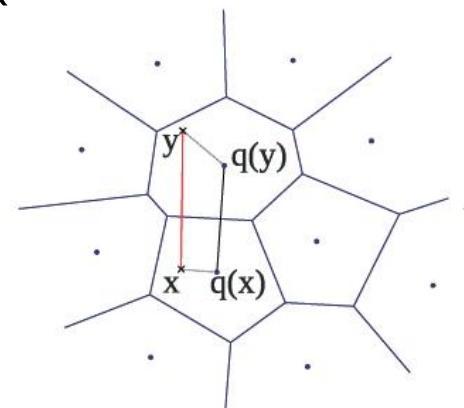
- 均方距离误差(MSDE)的上界是量化器的均方差(MSE)

$$d(x, q(y)) - d(y, q(y)) \leq d(x, y) \leq d(x, q(y)) + d(y, q(y)),$$

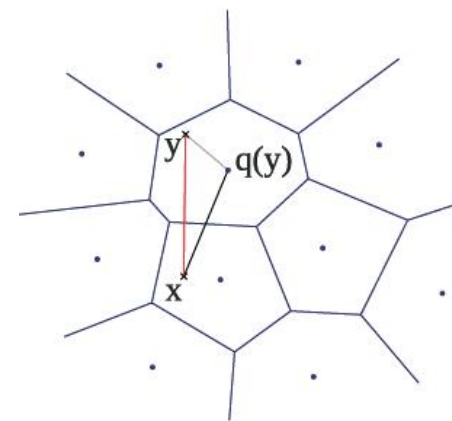
$$(d(x, y) - d(x, q(y)))^2 \leq d(y, q(y))^2.$$

$$\tilde{d}(x, y) = d(x, q(y)) = \sqrt{\sum_j d(u_j(x), q_j(u_j(y)))^2}$$

$$\begin{aligned} \text{MSDE}(q) &\triangleq \iint (d(x, y) - \tilde{d}(x, y))^2 p(x) dx p(y) dy. \\ &\leq \int p(x) \left(\int d(y, q(y))^2 p(y) dy \right) dx \\ &\leq \text{MSE}(q), \end{aligned}$$



symmetric case



asymmetric case