

第九章：图像特征表达

中国科学技术大学
电子工程与信息科学系



图像表达

- 基于全局特征的图像表达
- 基于局部特征的图像表达
 - 简单的局部视觉特征
 - 基于关键点检测的局部视觉特征
 - ✓ 图像表达基本框架
 - ✓ 局部视觉特征描述
 - ✓ 特征编码与聚合

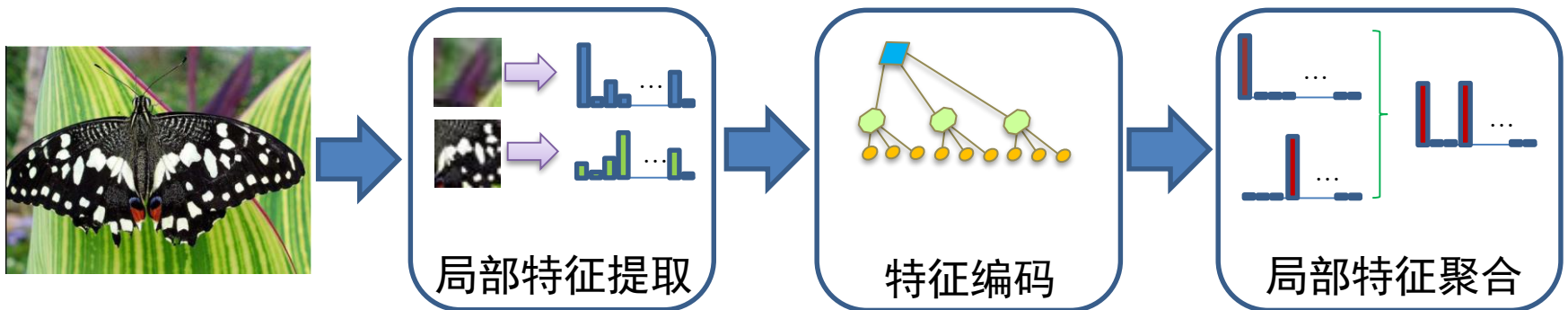


基于局部特征的图像表达

- 基于局部特征的图像表达基本框架
- 局部视觉特征描述
 - 局部关键点检测
 - ✓ 角点检测: Harris
 - ✓ 块检测: DoG, MSER
 - 局部区域描述
 - ✓ SIFT
 - ✓ LIOP
- 特征编码与聚合
 - Fisher Vector
 - 词袋模型(BOW)
 - VLAD
 - 乘积量化(PQ)

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



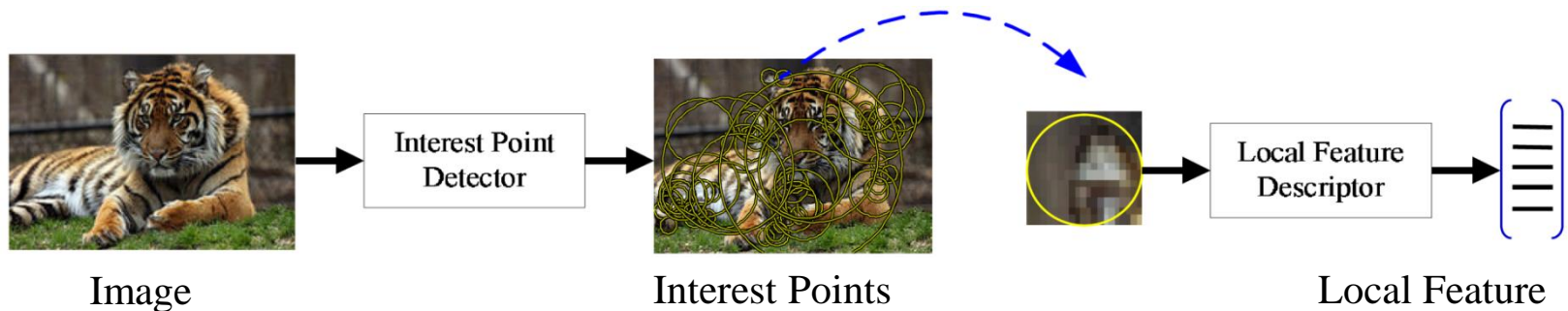
局部视觉特征

□ 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**





局部特征性质

- 可重复性 (Repeatability)
 - 图像经过变换后，对应的局部特征仍然存在

- 显著性 (Saliency)
 - 每个局部特征都对应一个有区别性的描述

- 紧凑性和高效性 (Compactness and efficiency)
 - 特征数目远少于图像像素数

- 局部性 (Locality)
 - 每个特征对应着一个相对较小的图像区域;
 - 对于混乱背景和遮挡具有鲁棒性

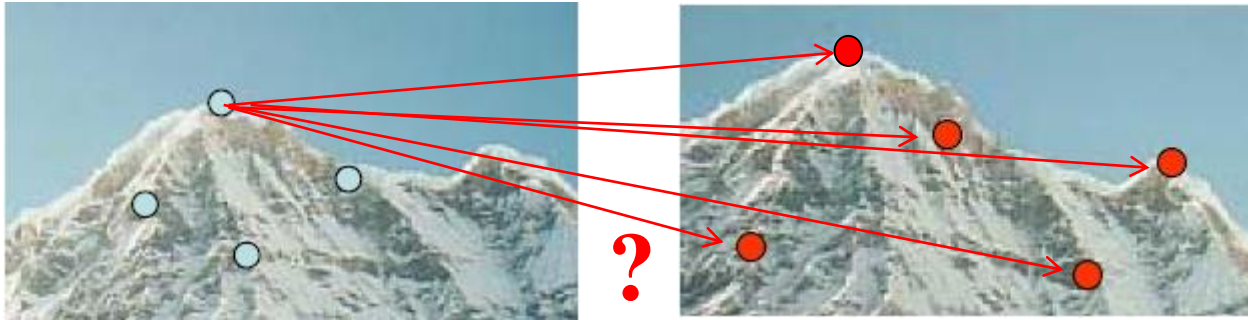
目标: 关键点检测的可重复性

- 对每副图像，局部关键点检测独立进行
 - 类似“管中窥豹”
- 基本要求
 - 在两副相关图像中能够独立地检测到若干**相同**的关键点
 - 如果关键点检测没有重复性保证，则不可能完成后续的匹配

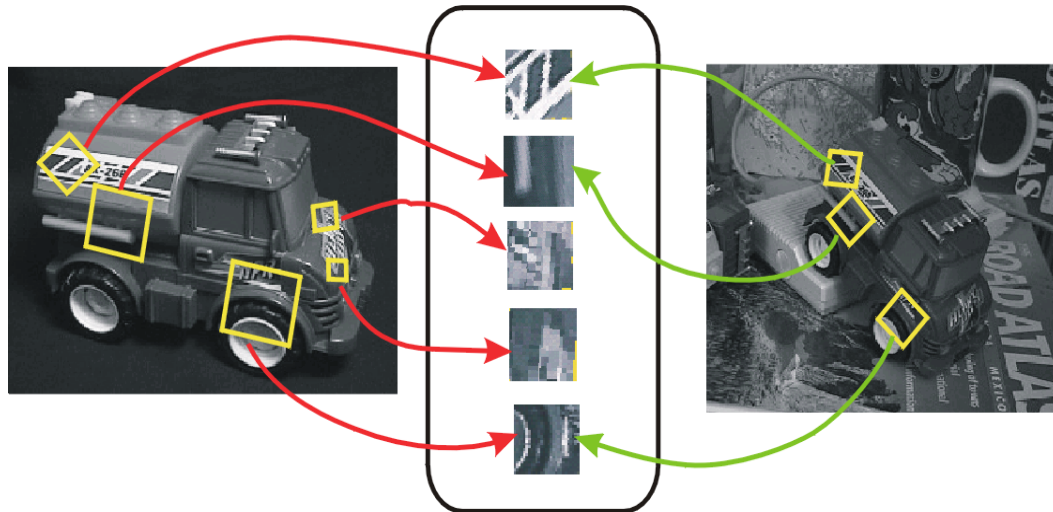


目标: 描述子的可区分性

- 可靠地确定兴趣点之间的对应关系

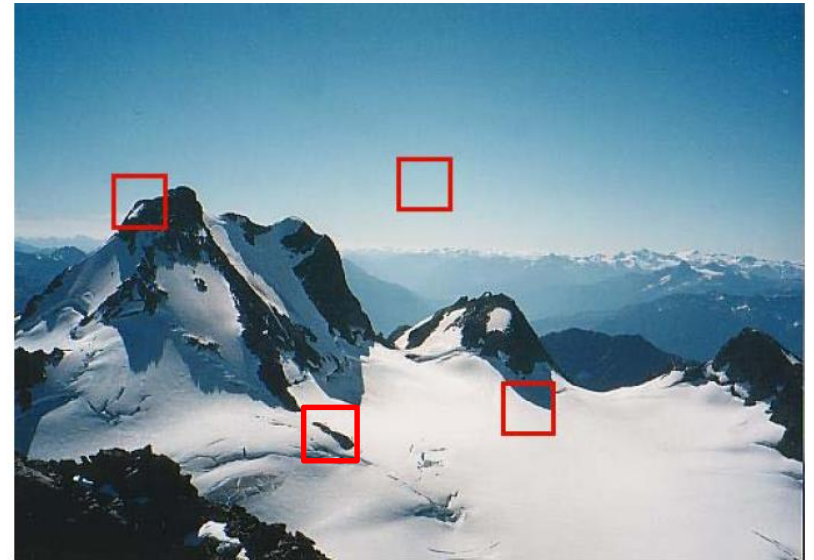
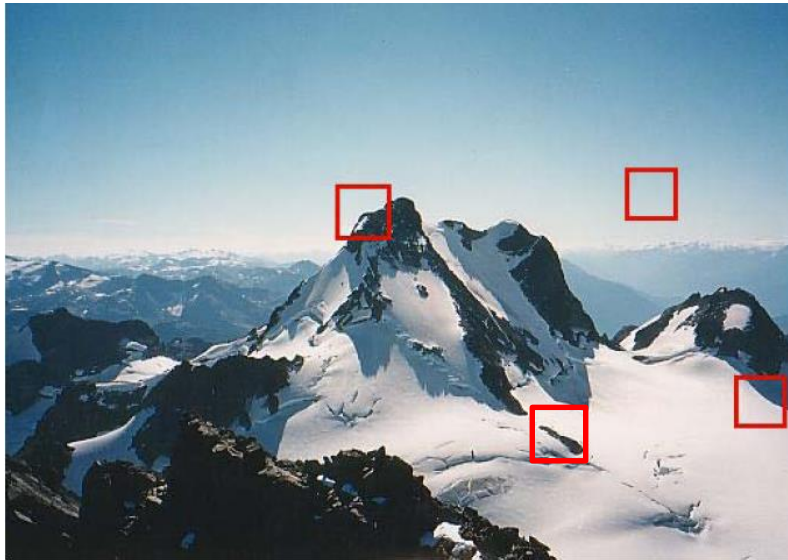


- 对于两个视角，描述子必须具有对几何和光照变化的不变性



局部视觉特征

- 在图像中寻找哪些与众不同的点/区域？
 - 目标：在图像之间建立明确的、无二义性的匹配
 - 如何定义这种“与众不同”？



平滑区域
(×)

边缘
(×)

角点
(✓)

块(blob)
(✓)





基于局部特征的图像表达

- 图像表达基本框架
- 局部视觉特征描述
 - 局部关键点检测
 - ✓ 角点检测：Harris
 - ✓ 块检测：DoG, MSER
 - 局部区域描述
 - ✓ SIFT
- 特征编码与聚合
 - Fisher Vector
 - Bow
 - VLAD
 - PQ



局部视觉特征描述

□ 局部关键点检测

■ 角点检测

✓ Harris 角点检测子

■ 块检测

✓ 高斯差分(DoG) 检测子

✓ MSER 检测子

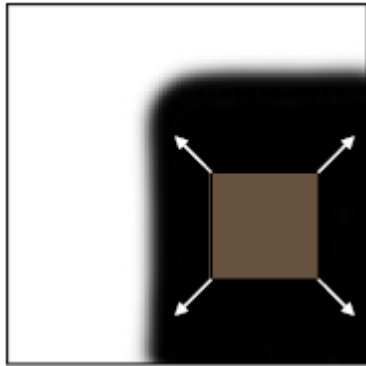
□ 局部区域描述

■ SIFT

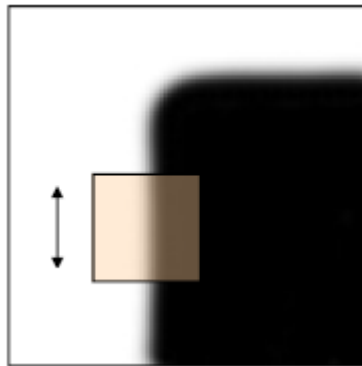
■ LIOP

关键点检测：角点

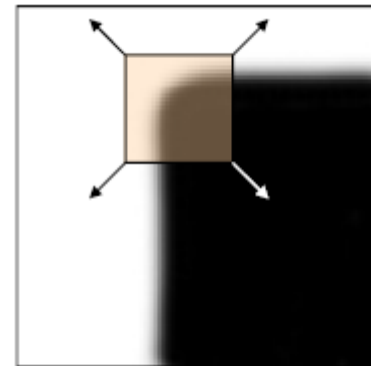
- 通过查看一个小窗口，即可简单的识别角点
 - “管中窥豹”
- 在角点上，向任何一个方向移动窗口，都会产生灰度的较大变化



“平坦”区域：
任意方向灰度
均无变化



“边缘”：沿边
缘方向，灰度无
变化



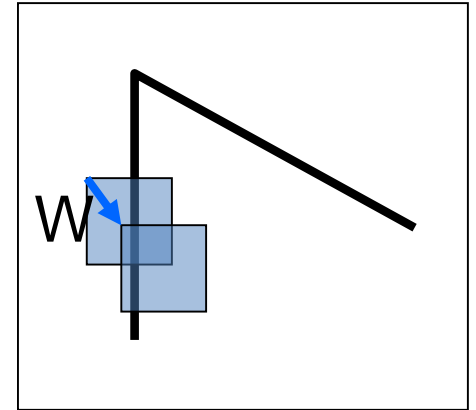
“角点”：所有
方向上，灰度均
有较大变化

Harris 角点检测

□ 考虑一个滑窗 W ，平移量为 (u, v)

- 滑动前后，滑窗 W 中的像素灰度如何变化？
- 比较滑窗内对应像素灰度，计算平方差之和

$$E(u, v) = \sum_{(x, y) \in W} [I(x + u, y + v) - I(x, y)]^2$$



- 对 I 关于 x 、 y 做泰勒展开：

$$I(x + u, y + v) = I(x, y) + \frac{\partial I}{\partial x}u + \frac{\partial I}{\partial y}v + \text{higher order terms}$$

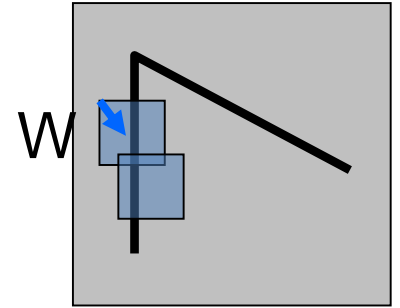
- 如果滑窗平移量 (u, v) 足够小，则可去掉高阶项：

$$\begin{aligned} I(x + u, y + v) &\approx I(x, y) + \frac{\partial I}{\partial x}u + \frac{\partial I}{\partial y}v \\ &\approx I(x, y) + [I_x \ I_y] \begin{bmatrix} u \\ v \end{bmatrix} \end{aligned}$$

Harris 角点检测

$$E(u, v) = \sum_{(x,y) \in W} [I(x+u, y+v) - I(x, y)]^2$$

$$I(x+u, y+v) \approx I(x, y) + [I_x \ I_y] \begin{bmatrix} u \\ v \end{bmatrix}$$



$$\begin{aligned} E(u, v) &= \sum_{(x,y) \in W} [I(x+u, y+v) - I(x, y)]^2 \\ &\approx \sum_{(x,y) \in W} [I(x, y) + [I_x \ I_y] \begin{bmatrix} u \\ v \end{bmatrix} - I(x, y)]^2 \\ &\approx \sum_{(x,y) \in W} \left[[I_x \ I_y] \begin{bmatrix} u \\ v \end{bmatrix} \right]^2 \end{aligned}$$

Harris 角点检测

□ 上式可重写为：

$$E(u, v) = [u \quad v] \sum_{(x,y) \in W} \begin{bmatrix} I_x^2 & I_x I_y \\ I_y I_x & I_y^2 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = [u \quad v] \mathbf{M} \begin{bmatrix} u \\ v \end{bmatrix}$$



□ 对于上方的例子

- 可以将蓝色滑窗的中心移到移到绿色单位圆上的任意位置
- 朝那个方向 $[u \quad v]$ 移动，会得到**最大**和**最小**的 E 值呢？
 - ✓ 可通过对矩阵 \mathbf{M} 的特征向量，得到这两个方向



回顾：特征向量、特征值

- 对于矩阵 \mathbf{A} ，其特征向量为 \mathbf{x} ，当其满足：

$$\mathbf{Ax} = \lambda\mathbf{x}$$

标量 λ 为特征向量 \mathbf{x} 所对应的特征值

- 特征值可通解如下方程得到： $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$

- 在上页PPT上，我们的 $\mathbf{A} = \mathbf{M}$ 是一个 2×2 的矩阵，因此可得

$$\det \begin{pmatrix} m_{11} - \lambda & m_{12} \\ m_{21} & m_{22} - \lambda \end{pmatrix} = 0$$

其解为：

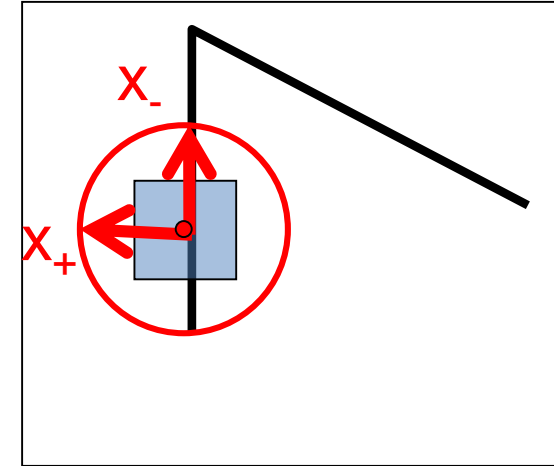
$$\lambda_{\pm} = \frac{1}{2} \left[(m_{11} + m_{22}) \pm \sqrt{4m_{12}m_{21} + (m_{11} - m_{22})^2} \right]$$

- 一旦得到特征值 λ ，可以通过解如下方程得到特征向量 \mathbf{x}

$$\begin{bmatrix} m_{11} - \lambda & m_{12} \\ m_{21} & m_{22} - \lambda \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 0$$

Harris 角点检测

$$E(u, v) = [u \ v] \underbrace{\sum_{(x,y) \in W} \begin{bmatrix} I_x^2 & I_x I_y \\ I_y I_x & I_y^2 \end{bmatrix}}_M \begin{bmatrix} u \\ v \end{bmatrix}$$



□ 矩阵M 的特征值和特征向量

- Define shifts with the largest and smallest change (E value)
- x_+ = direction of largest increase in E.
- λ_+ = amount of increase in direction x_+

$$Mx_+ = \lambda_+ x_+$$

- λ_- = direction of smallest increase in E.
- x_- = amount of increase in direction x_+

$$Mx_- = \lambda_- x_-$$



Harris 角点检测

□ 滑窗中的灰度变化：特征值分析

- λ_{\min} , λ_{\max} : 对称矩阵 \mathbf{M} 的特征值
- 如果尝试所有可能的平移方向向量 \mathbf{n} , 那么产生最大灰度变化值 $E(\mathbf{n})$ 为 λ_{\max}

$$E(u, v) \cong [u, v] M \begin{bmatrix} u \\ v \end{bmatrix}$$

$$M\mathbf{x}_{\max} = \lambda_{\max}\mathbf{x}_{\max}; M\mathbf{x}_{\min} = \lambda_{\min}\mathbf{x}_{\min}; \mathbf{x}_{\max}^T\mathbf{x}_{\min}=0$$

$$\text{定义 } \mathbf{n} = [u, v]^T = a\mathbf{x}_{\max} + b\mathbf{x}_{\min}, \text{ with } a^2 + b^2 = 1$$

$$\text{于是, } E(\mathbf{n}) = \mathbf{n}^T M \mathbf{n}$$

$$= (a\mathbf{x}_{\max} + b\mathbf{x}_{\min})^T M (a\mathbf{x}_{\max} + b\mathbf{x}_{\min})$$

$$= (a\mathbf{x}_{\max} + b\mathbf{x}_{\min})^T (a\lambda_{\max}\mathbf{x}_{\max} + b\lambda_{\min}\mathbf{x}_{\min})$$

$$= a^2\lambda_{\max} + b^2\lambda_{\min}$$

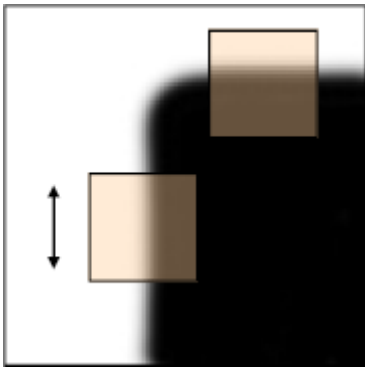
$$= a^2\lambda_{\max} + (1 - a^2)\lambda_{\min}$$

$$\text{因此, } E_{\max} = \lambda_{\max}, \text{ with } a = 1 \text{ and } \mathbf{n} = [u, v]^T = \mathbf{x}_{\max};$$

$$E_{\min} = \lambda_{\min}, \text{ with } a = 0 \text{ and } \mathbf{n} = [u, v]^T = \mathbf{x}_{\min}.$$

角点响应函数

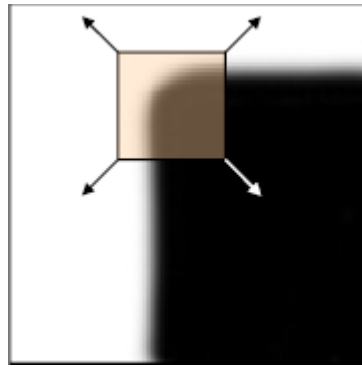
□ 定性分析矩阵对称矩阵M的特征值



边缘:

$$\lambda_1 \gg \lambda_2$$

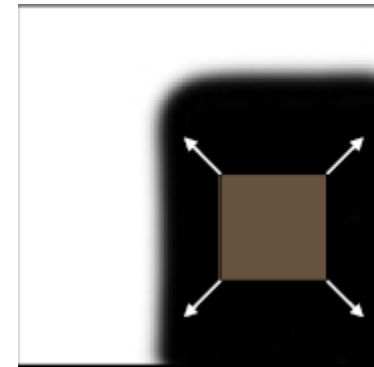
$$\lambda_2 \gg \lambda_1$$



角点:

λ_1 和 λ_2 均较大,

$$\lambda_1 \sim \lambda_2;$$

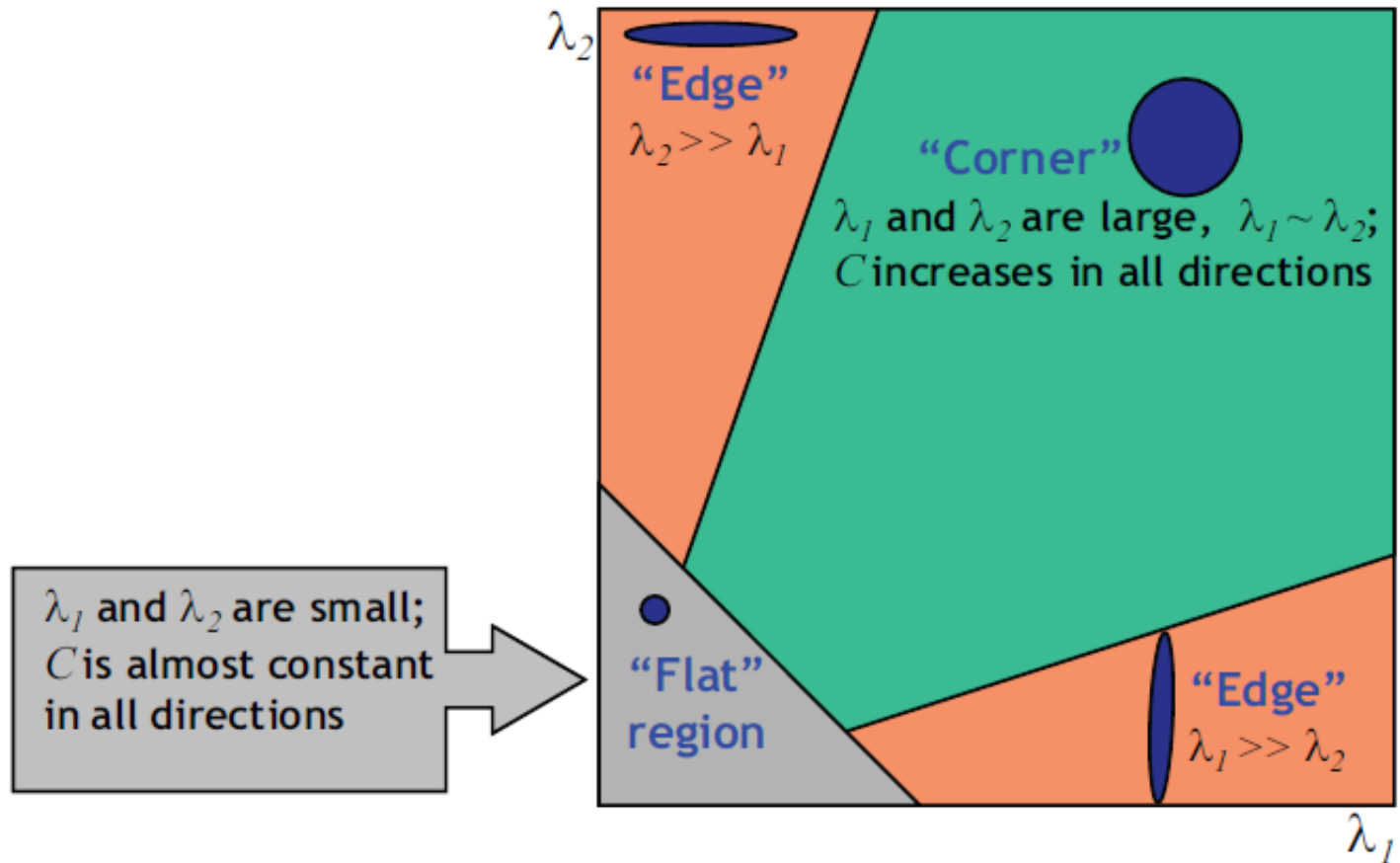


平滑区域:

λ_1 和 λ_2 均较小;

角点响应函数

$$\mathbf{M} = \begin{bmatrix} \sum_{(x,y) \in W} I_x^2 & \sum_{(x,y) \in W} I_x I_y \\ \sum_{(x,y) \in W} I_y I_x & \sum_{(x,y) \in W} I_y^2 \end{bmatrix} = \mathbf{R}^{-1} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \mathbf{R}$$





角点响应函数

- 对每个图像窗，计算 M 矩阵，基于其特征值，定义角点响应函数 R ：

$$R(x, y) = \lambda_1 \lambda_2 - \alpha (\lambda_1 + \lambda_2)^2$$

$$\det(\mathbf{C}) = \lambda_1 \lambda_2$$

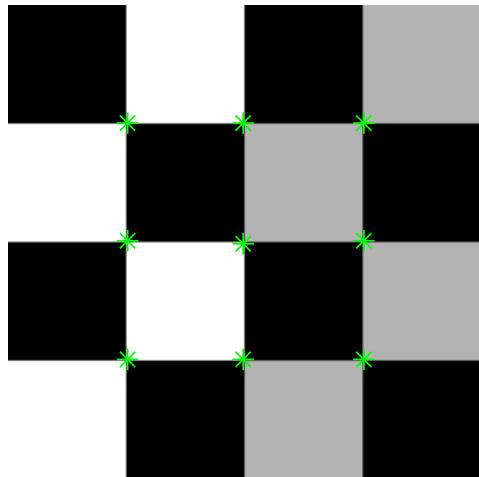
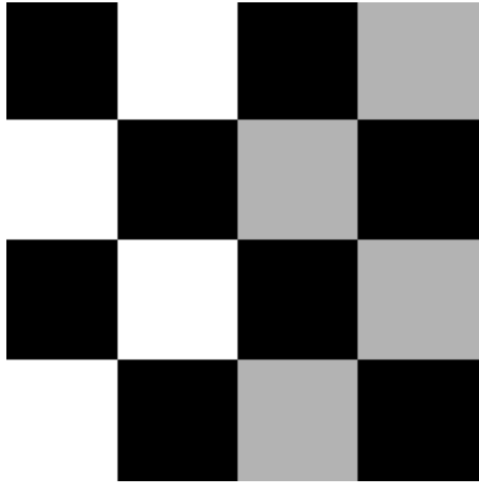
$$\text{trace}(\mathbf{C}) = \lambda_1 + \lambda_2$$

- 找到角点响应较大的值所对应的图像位置点：

$$R > \text{threshold}$$

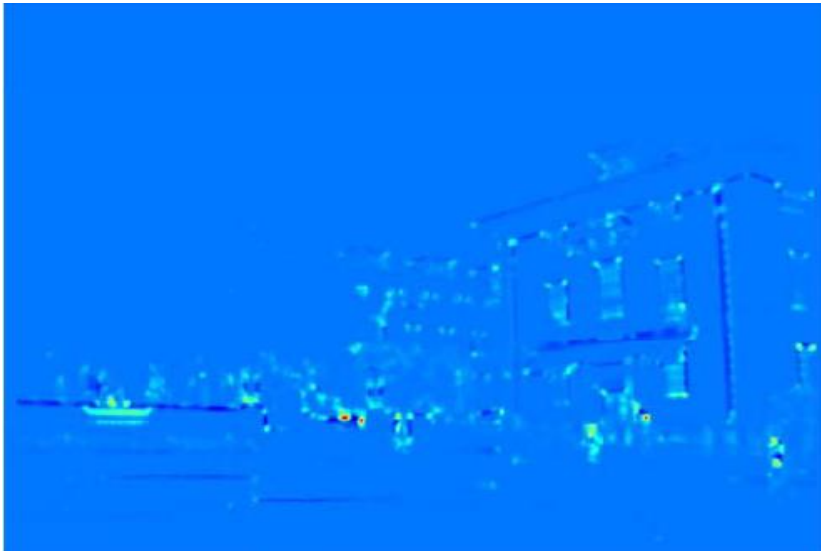
- 取 R 的局部极大值点，例如 进行非最大抑制

Harris角点检测示例



Harris角点检测示例角点响应图

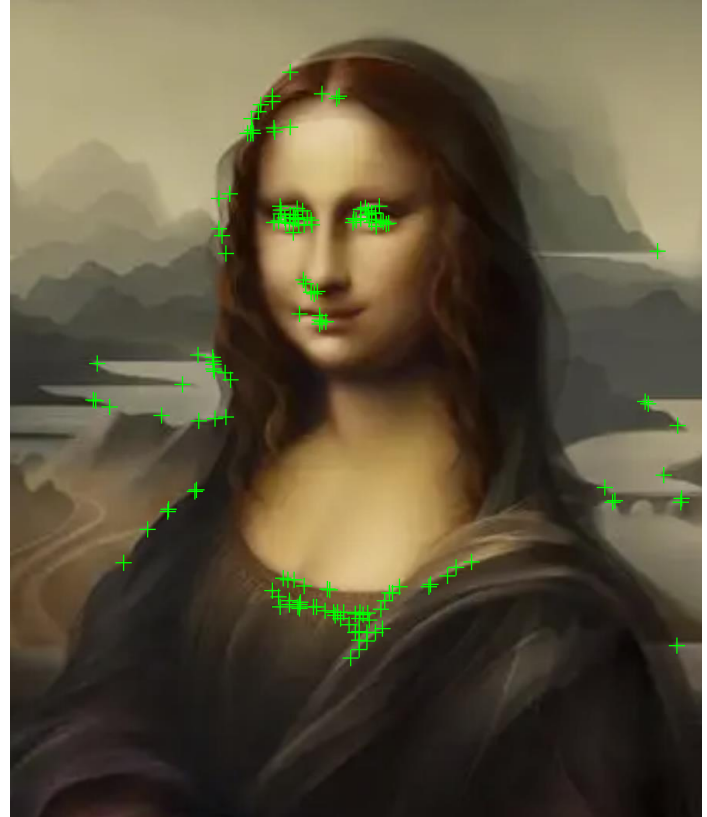
角点响应图 R



对 R 进行门限处理, 取极大值



对两张图像独立进行检测处理





Harris 角点检测子的性质

□ 旋转不变?

■ Yes

$$M = X \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} X^T$$

□ 尺度不变?

■ No

校正窗口尺度函数



Window scale = 10



Window scale = 15



Window scale = 30

Harris 检测算子检测到的特征点 取决于所选的窗的大小.



局部视觉特征描述

□ 局部关键点检测

■ 角点检测

✓ Harris 角点检测子

■ 块检测

✓ 高斯差分(DoG) 检测子

✓ MSER 检测子

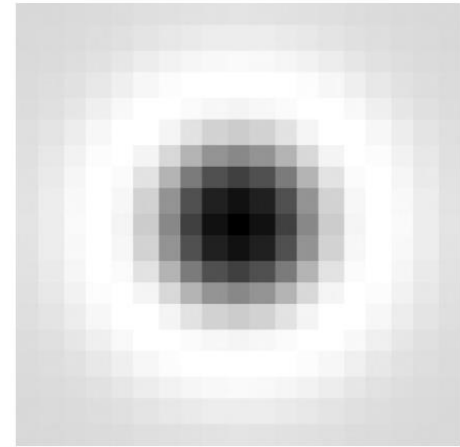
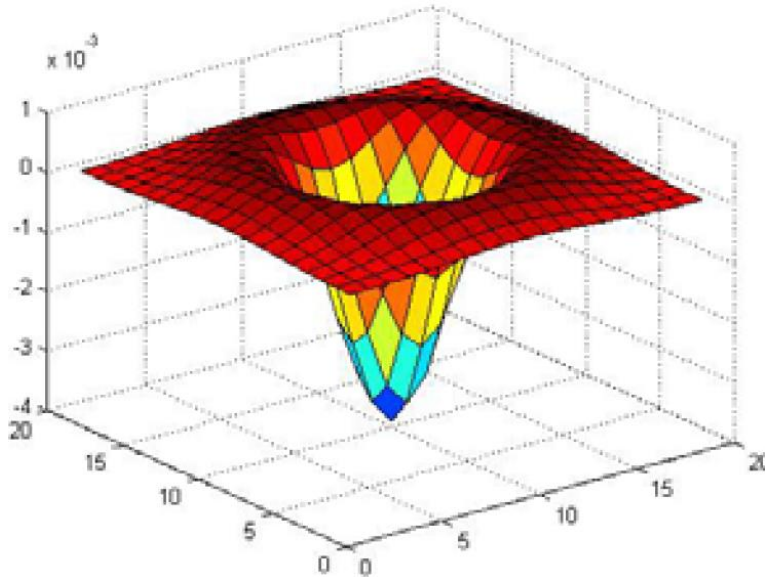
□ 局部区域描述

■ SIFT

■ LIOP

Laplacian-of-Gaussian (LoG)

- 2D的圆周对称操作子，用于块(blob)检测
 - LoG: “**blob**” detector

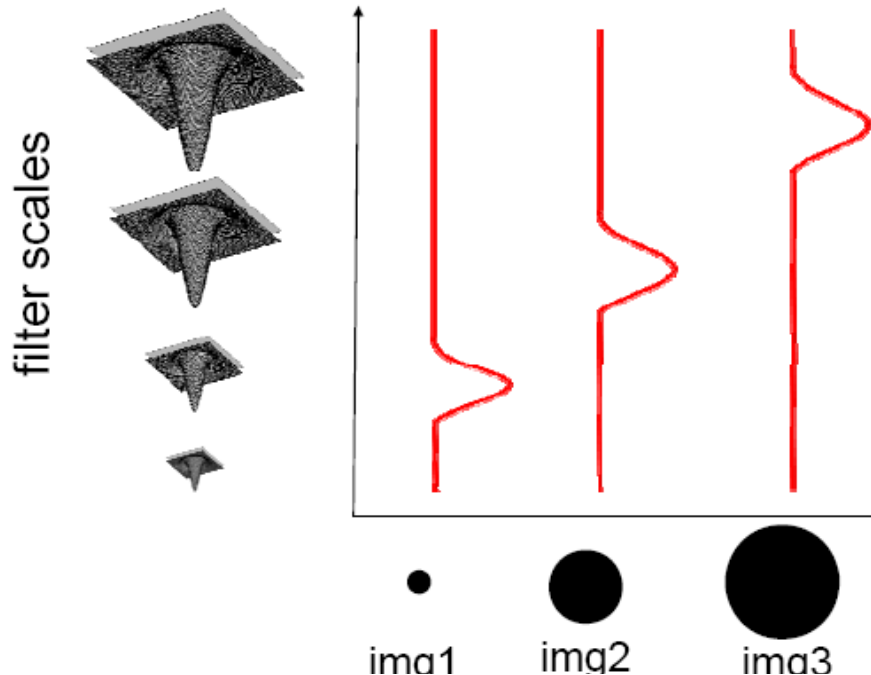


$$LoG = \nabla^2 G_\sigma(x, y) = \frac{\partial^2}{\partial x^2} G_\sigma(x, y) + \frac{\partial^2}{\partial y^2} G_\sigma(x, y)$$

$$G_\sigma(x, y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{x^2 + y^2}{2\sigma^2}\right]$$

Laplacian-of-Gaussian (LoG)

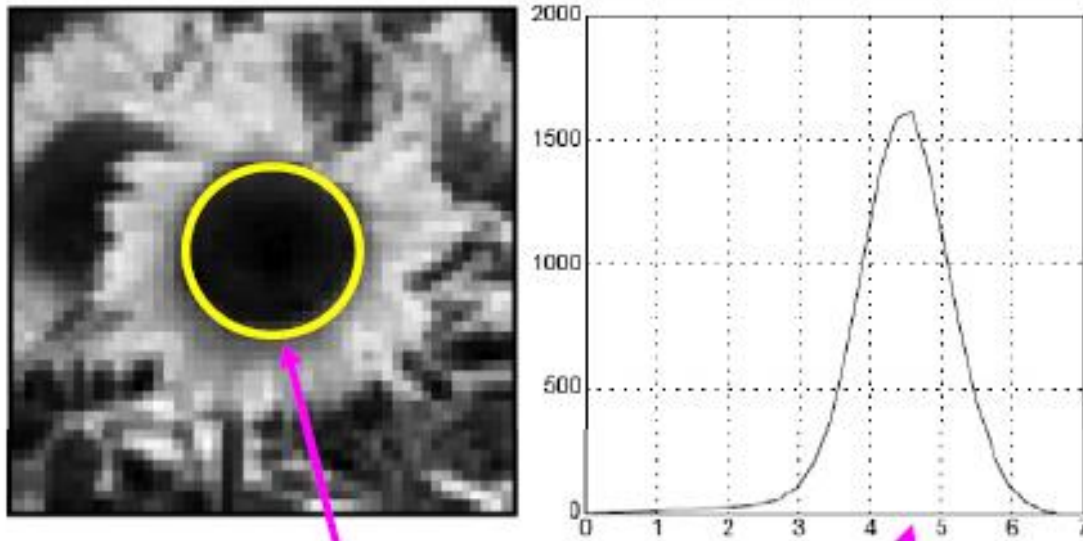
- Laplacian-of-Gaussian = “blob” detector
 - 给定图像块（patch），遍历枚举不同尺寸LoG的滤波器，将每个滤波器与图像块相卷积
 - 根据卷积响应的极大/极小值，确定适合当前图像块的LoG的滤波器，该滤波器的尺寸即定义了图像的尺度



$$\nabla^2 g = \frac{\partial^2 g}{\partial x^2} + \frac{\partial^2 g}{\partial y^2}$$

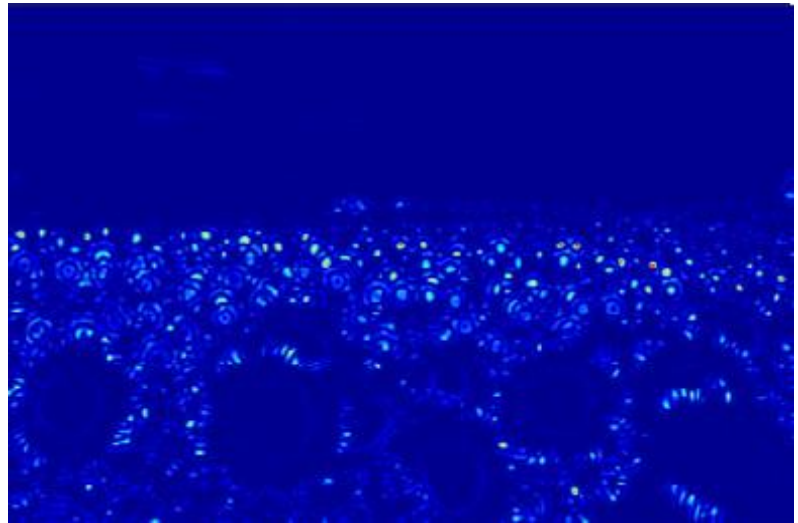
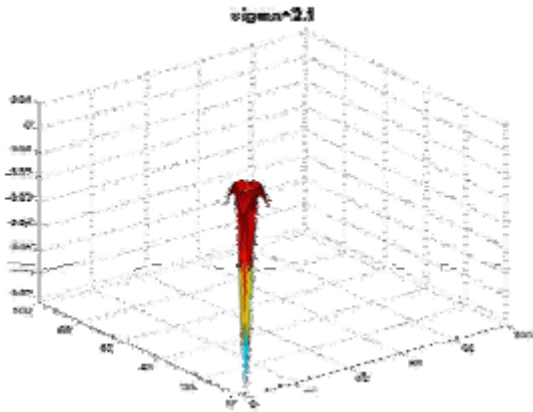
对于图像中的一个给定点:

- 特征尺度 (characteristic scale)
 - Laplacian 响应的极值所对应的尺度

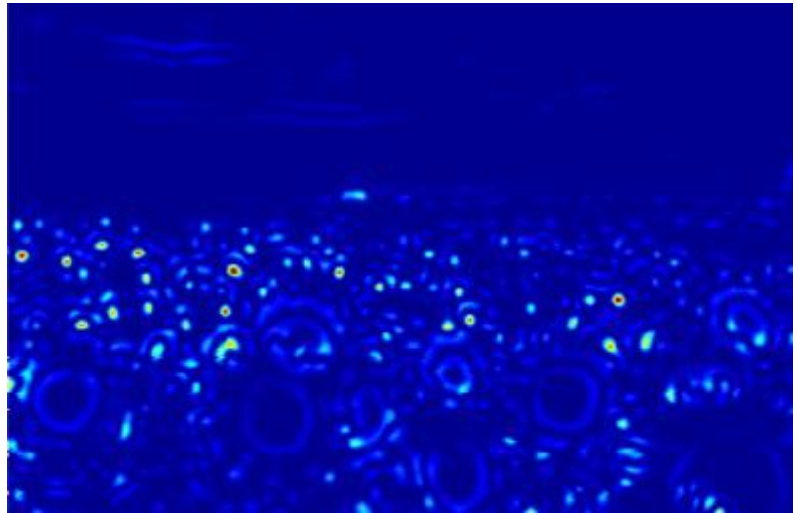
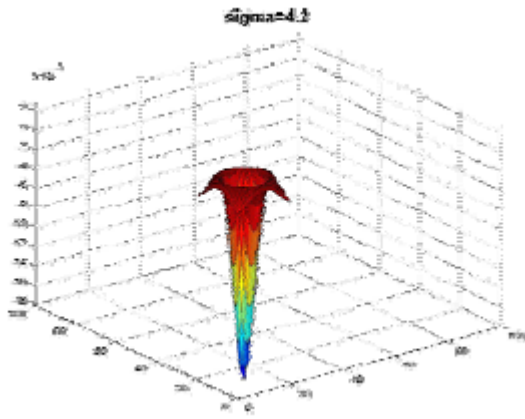


characteristic scale

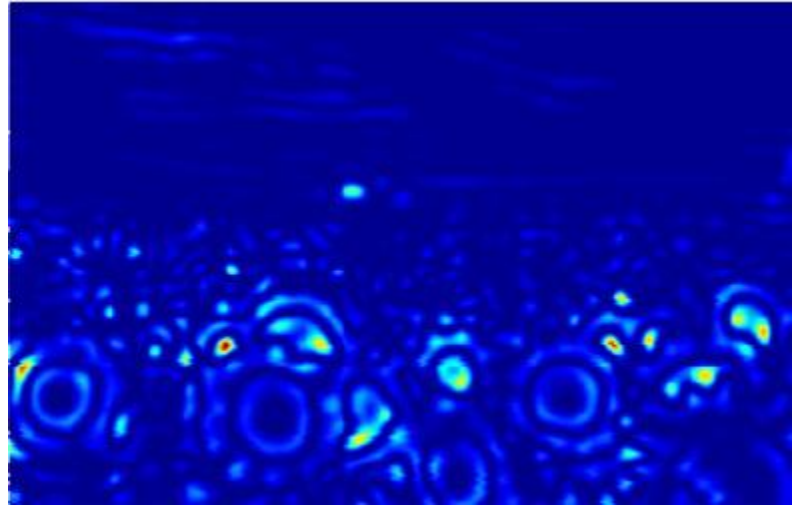
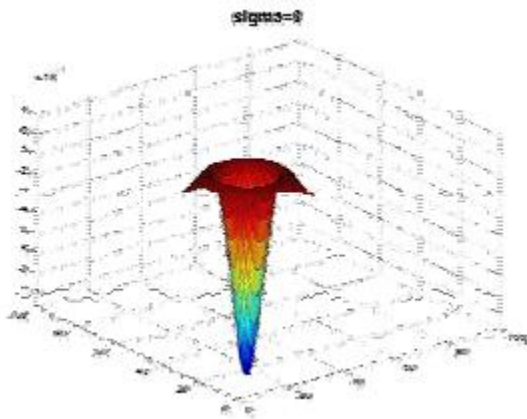
LoG响应示意图-1



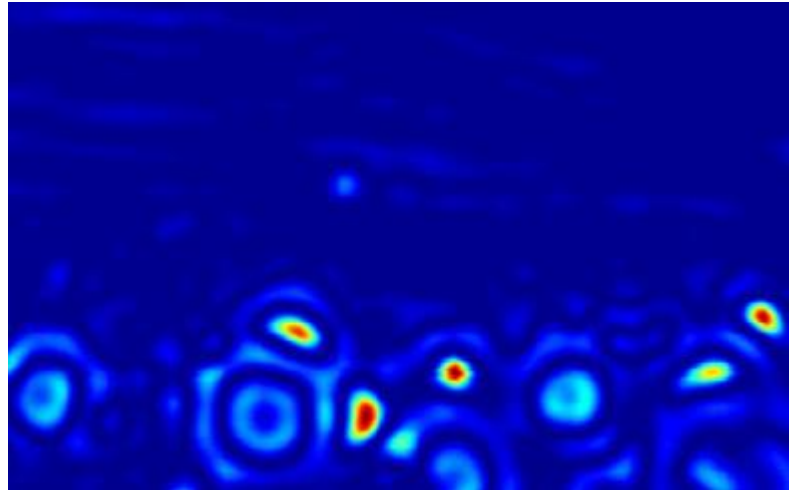
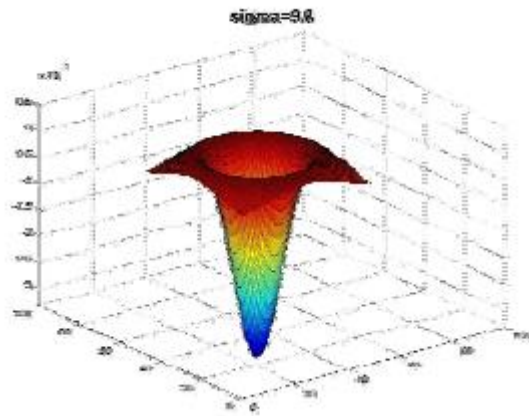
LoG响应示意图-2



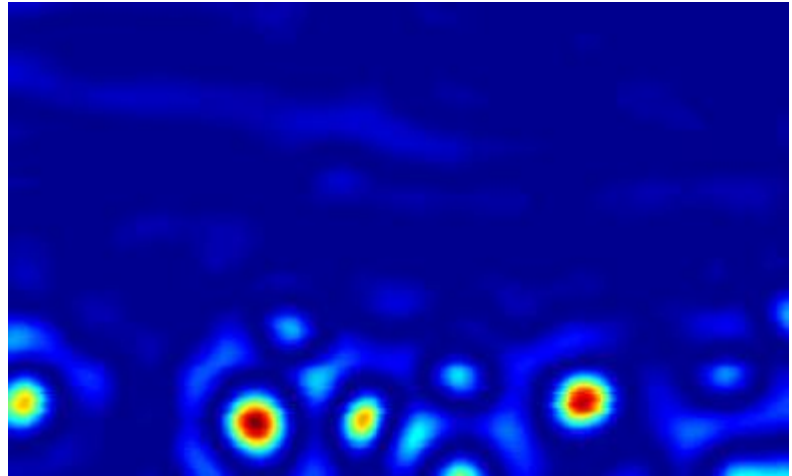
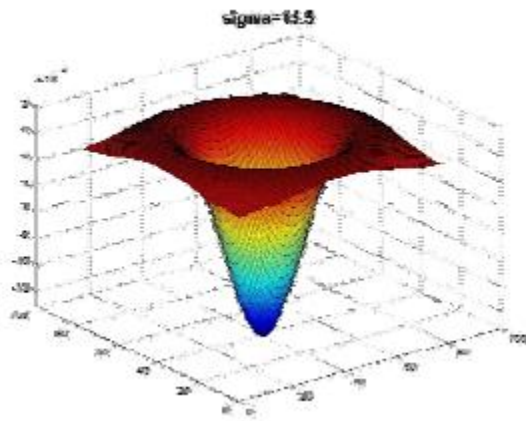
LoG响应示意图-3



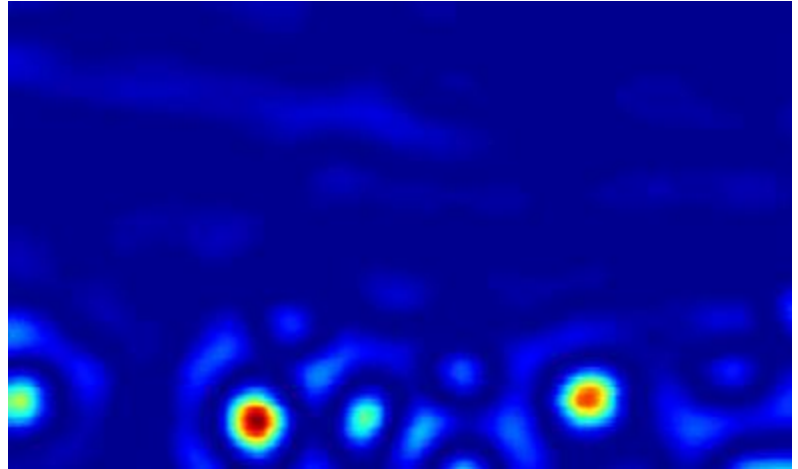
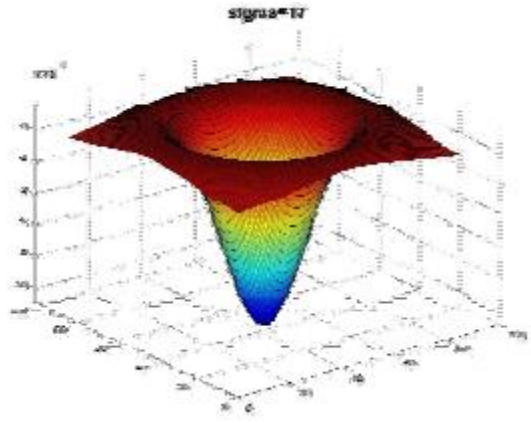
LoG响应示意图-4



LoG响应示意图-5

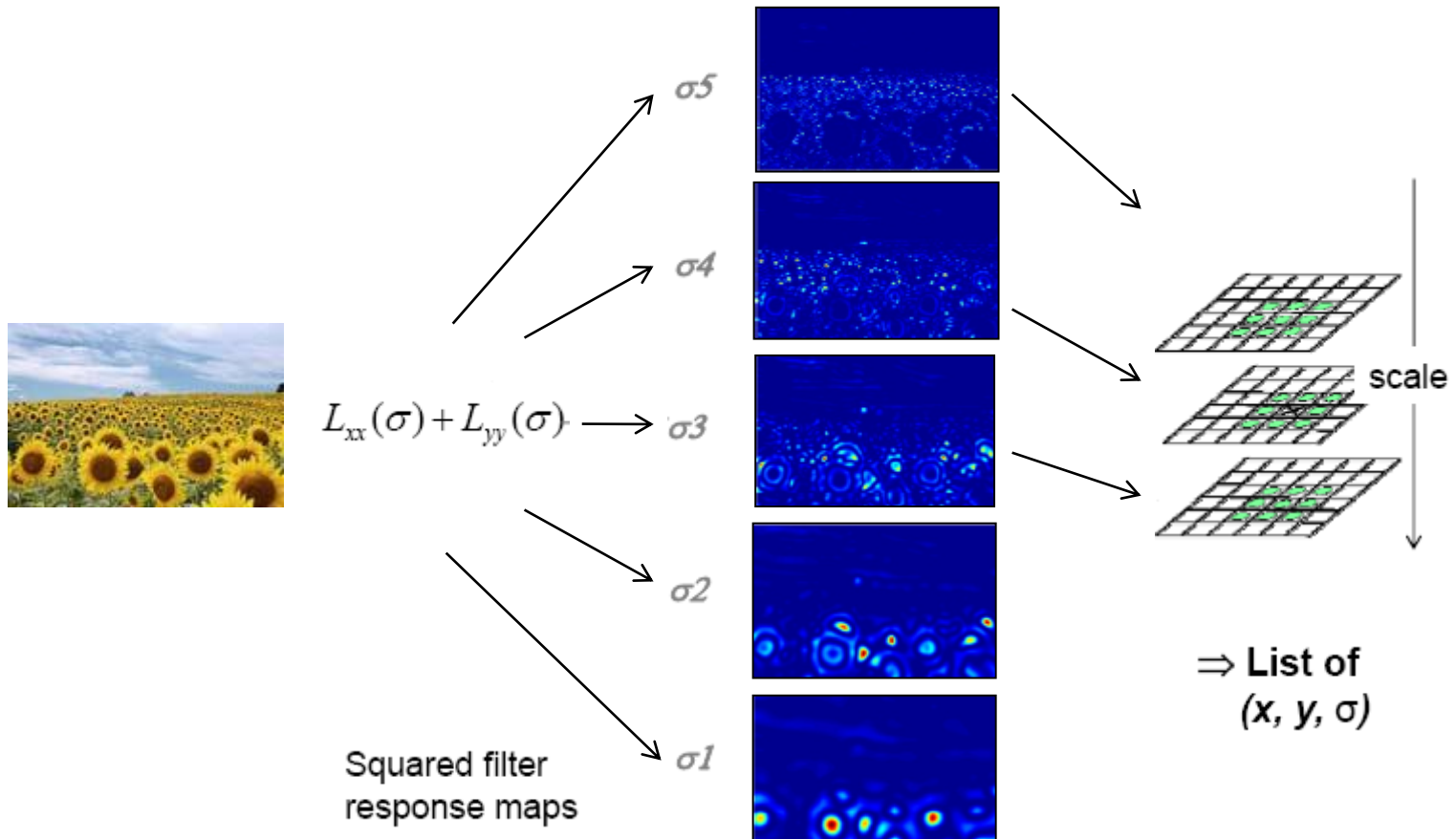


LoG响应示意图-6



尺度空间的 blob 检测

- 根据卷积响应的极大/极小值，确定适合当前像素位置的LoG的滤波器，该滤波器的尺寸即定义了当前像素的尺度



尺度空间的 blob 检测: 示例

- 检测出来的区域不一定是完整的blob或独立的区域

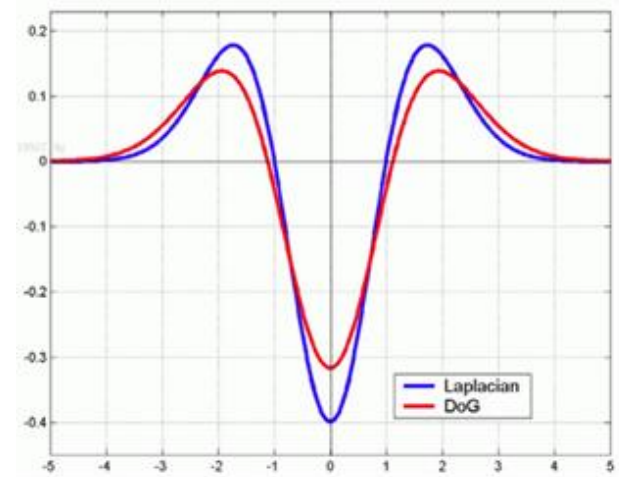


从LoG到DoG: SIFT检测子

- 采用 Difference of Gaussians (DoG) 去近似 Laplacian of Gaussians (LoG)
 - 差分阶次更低 (二阶 \rightarrow 一阶)
 - 高效的实现形式

LoG: $L = \sigma^2 (G_{xx}(x, y, \sigma) + G_{yy}(x, y, \sigma))$

DoG: $DoG = G(x, y, k\sigma) - G(x, y, \sigma)$

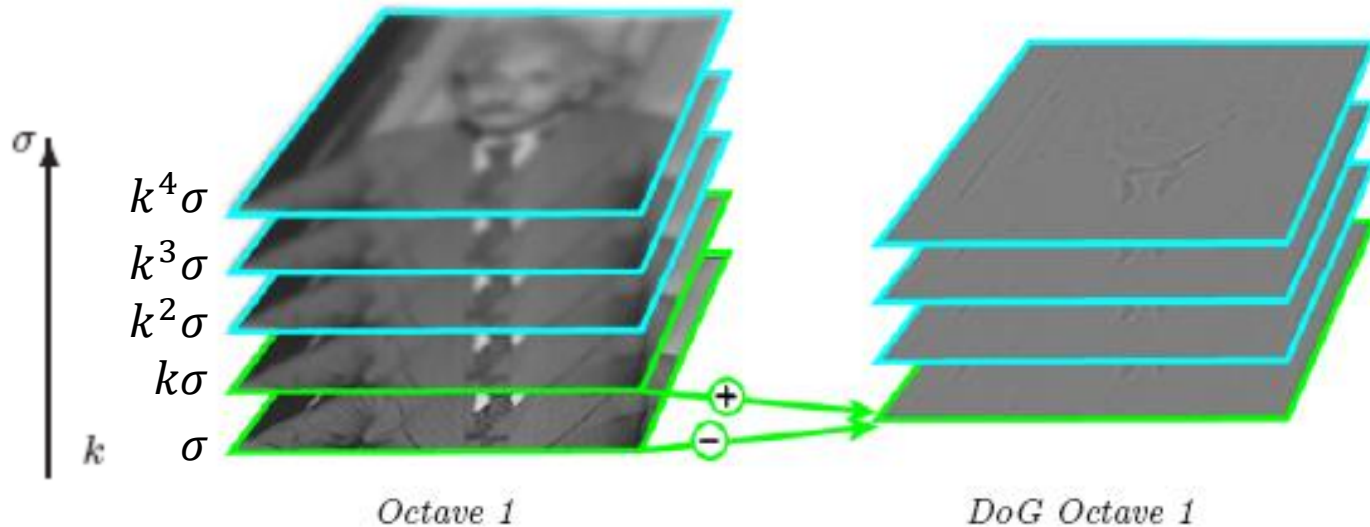


DoG 图像金字塔-1

□ 首先构建高斯金字塔

- 用尺度为 σ 的高斯核对图像做卷积，得到不同尺度的图像
- 每层octave有 $s + 3$ 副滤波结果图像， $k = 2^{\frac{1}{s}}$
- 相邻高斯图作差，得到高斯差分图，每层octave有 $s + 2$ 张图
- ✓ 节省计算开销

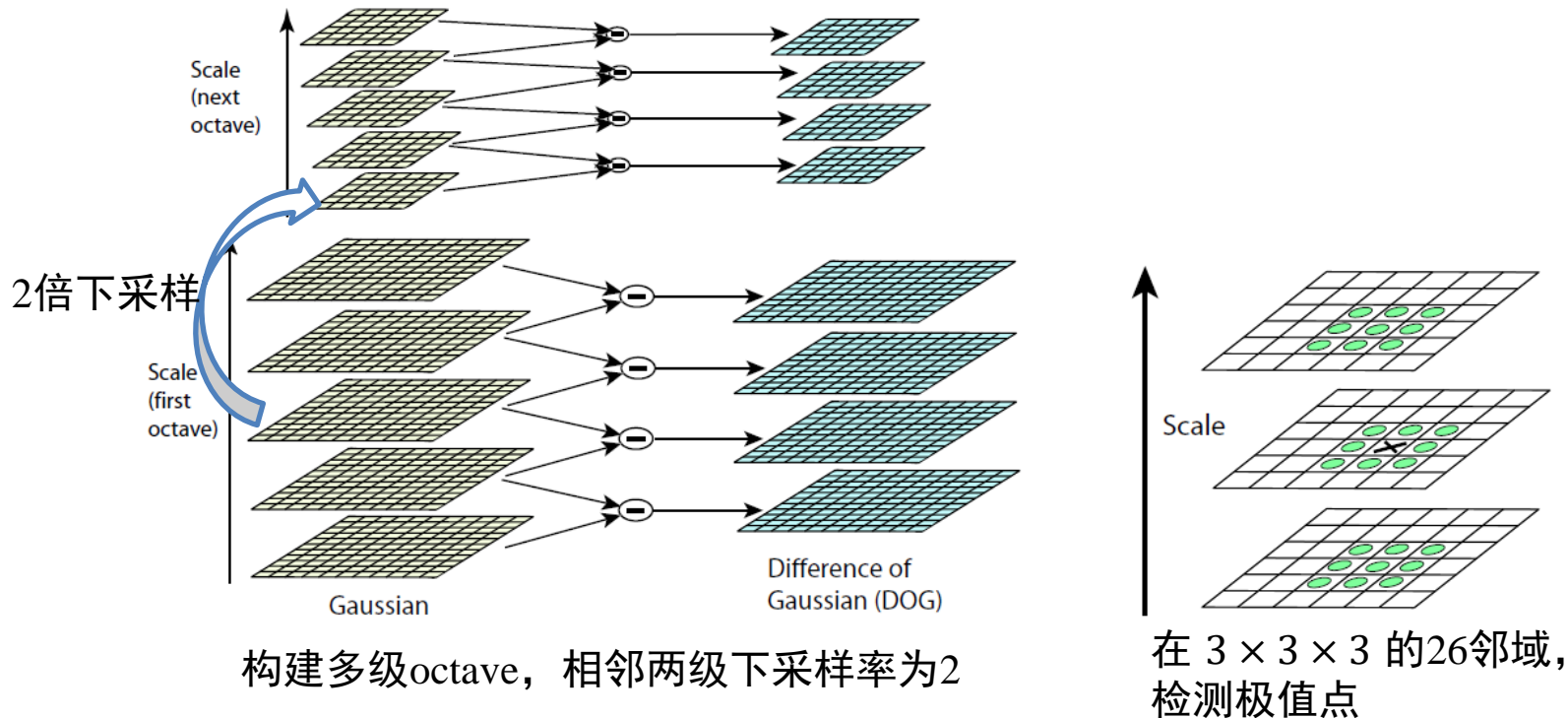
$$\begin{aligned} D(x, y, \sigma) &= (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) \\ &= L(x, y, k\sigma) - L(x, y, \sigma). \end{aligned}$$



DoG 图像金字塔-2

□ 关键点检测

- 相邻两个octave的尺度差为2
- 每组octave的 $s + 2$ 副DoG图表示 (x, y, scale) 三维尺度空间
- 在 $3 \times 3 \times 3$ 的26邻域寻找局部极大、极小值点
- 对于每个octave，局部极值点仅落在 s 副DoG图像上





DoG 图像金字塔-3

□ 小结

- 将图象中的**关键区域的尺度估计问题**转换为**DoG滤波器匹配问题**，将**最匹配的DoG滤波器的标准差**作为**关键区域的尺度估计值**
- 将图象的尺度空间划分为一系列子区间： $(\sigma, 2\sigma]$, $(2\sigma, 4\sigma]$, $(4\sigma, 8\sigma]$, \dots , $(2^m\sigma, 2^{m+1}\sigma]$, \dots 。每个尺度子区间对应一个octave
- 由于相邻两个octave对图象进行了2倍（行、列）下采样，所以每个octave中的都**等价于**在不同图象分辨率下进行**尺度范围为 $[\sigma, 2\sigma)$ 的关键区域尺度估计**
- 在第 m 个octave上的尺度估计结果，**通过乘以 2^m** ，即可得到在原始图象分辨率下的**关键区域尺度估计值**



局部视觉特征描述

□ 局部关键点检测

■ 角点检测

- ✓ Harris 角点检测子

■ 块检测

- ✓ 高斯差分(DoG) 检测子
- ✓ 最大稳定极值区域 (MSER) 检测子

□ 局部区域描述

■ SIFT

■ LIOP

最大稳定极值区域检测子: MSER

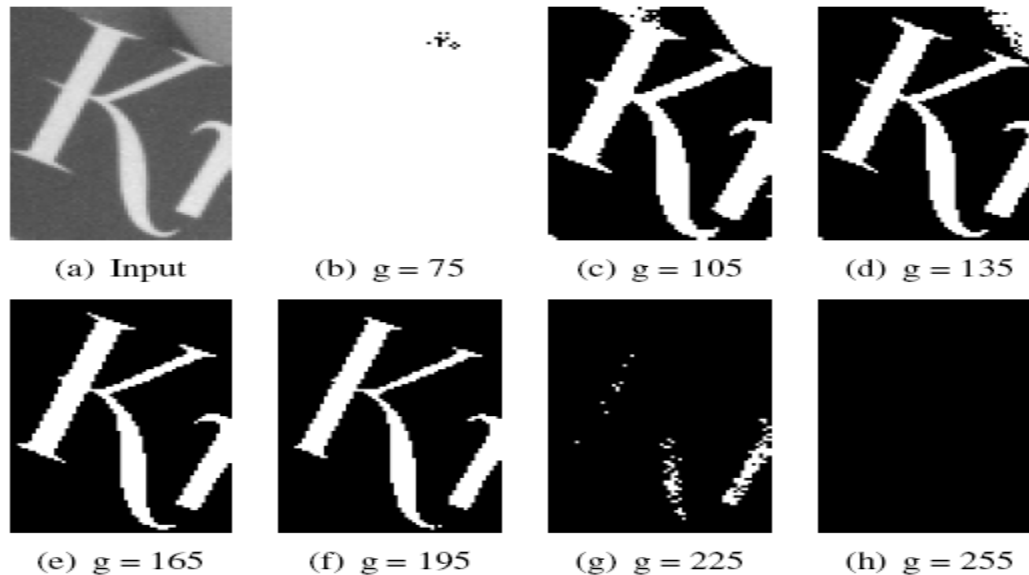
□ MSER: Maximally Stable Extremal Region

- 在一个比较大的阈值范围内, 可被独立分割出来的、面积稳定的图像区域

- 区域内部和区域边界上的像素的灰度有显著差异

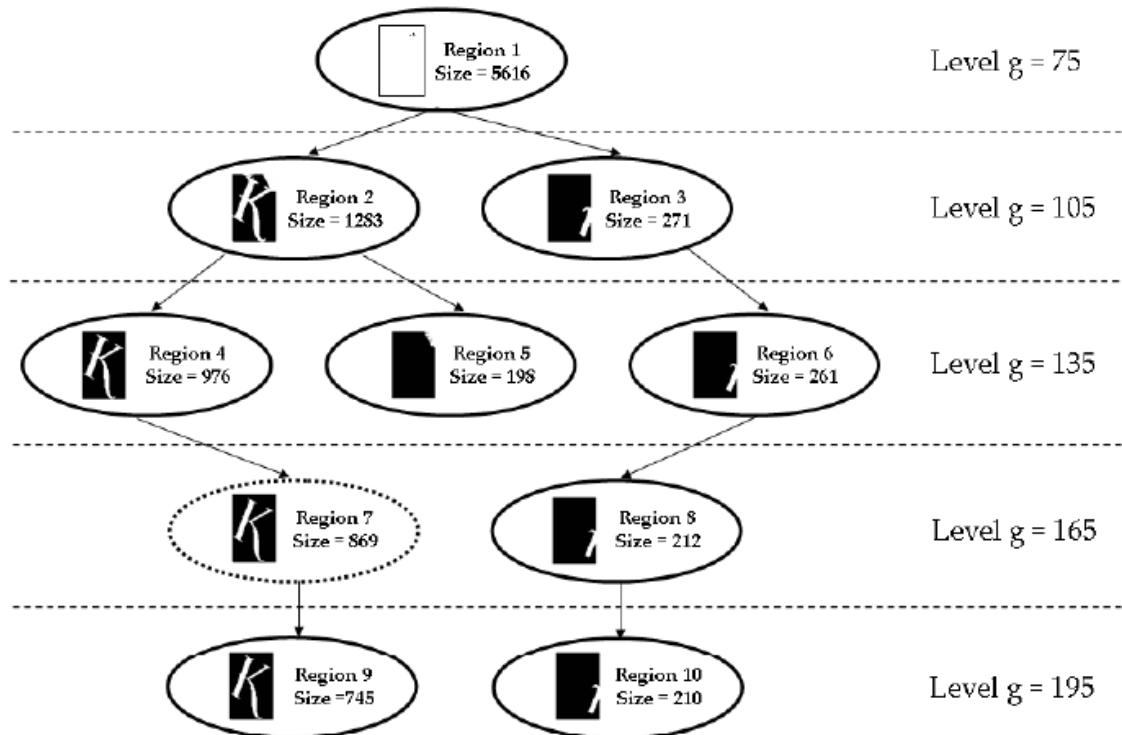
$$\forall p \in R_i, \forall q \in \text{boundary}(R_i) \rightarrow I(p) - I(q) > \tau$$

- 实现时, 可用一系列渐变的阈值对图像进行分割, 比较分割结果



最大稳定极值区域检测子: MSER

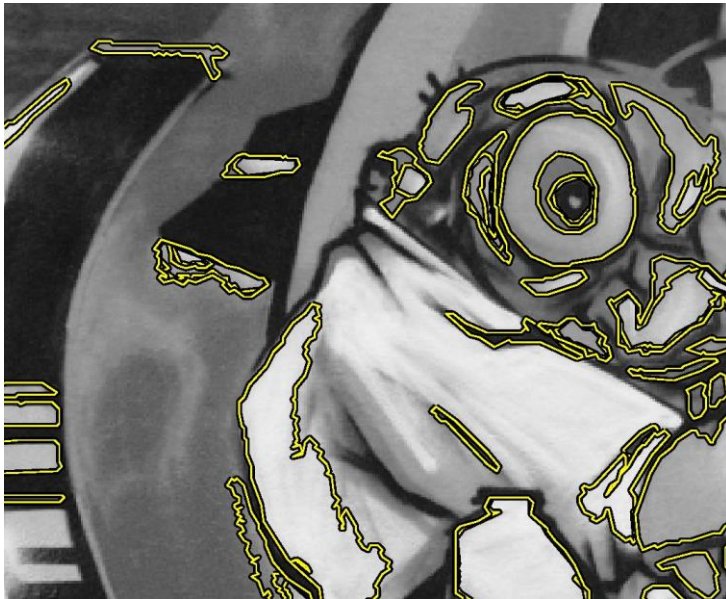
- 对比不同阈值下的分割结果，构建分割树
 - 该树结构中，不同层对应不同的分割阈值
 - 对各个叶节点，往上回溯到其第一个分叉节点（子节点数量大于2的节点），统计回溯的层级数；若层级数大于指定阈值，且区域面积变化较小，则该区域为一个MSER



最大稳定极值区域检测子: MSER

□ MSER的特点

- 对较大仿射畸变有较好的检测鲁棒性，
 - ✓ 在可重复性（repeatability）指标上，优于其他局部特征检测子
- 存在的问题：一副图象中检测得到的MSER区域数量有限
 - ✓ 例如，对于一副 640×480 的图像，MSER数量一般 < 100
 - ✓ 可以支持的图像之间的MSER匹配数量有限





局部视觉特征描述

□ 局部关键点检测

■ 角点检测

✓ Harris 角点检测子

■ 块检测

✓ 高斯差分(DoG) 检测子

✓ MSER 检测子

□ 局部区域描述

■ SIFT

■ LIOP

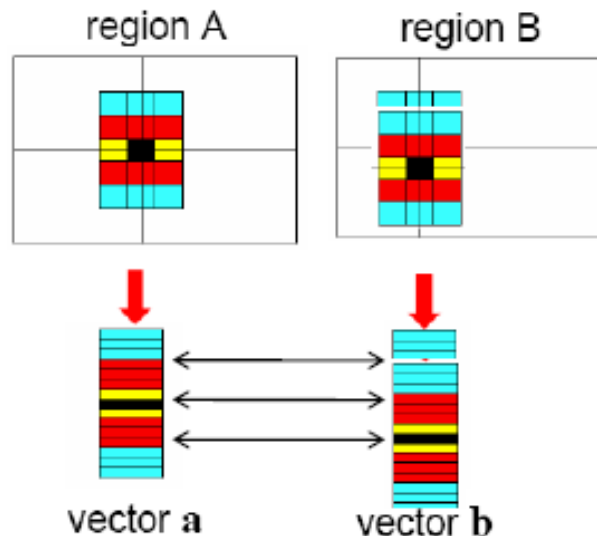
局部区域描述

□ 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

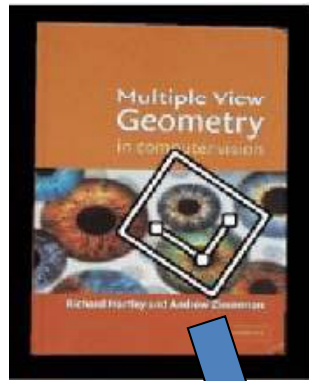
□ 将原始的像素灰度作为局部区域的描述？

- 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！！！！



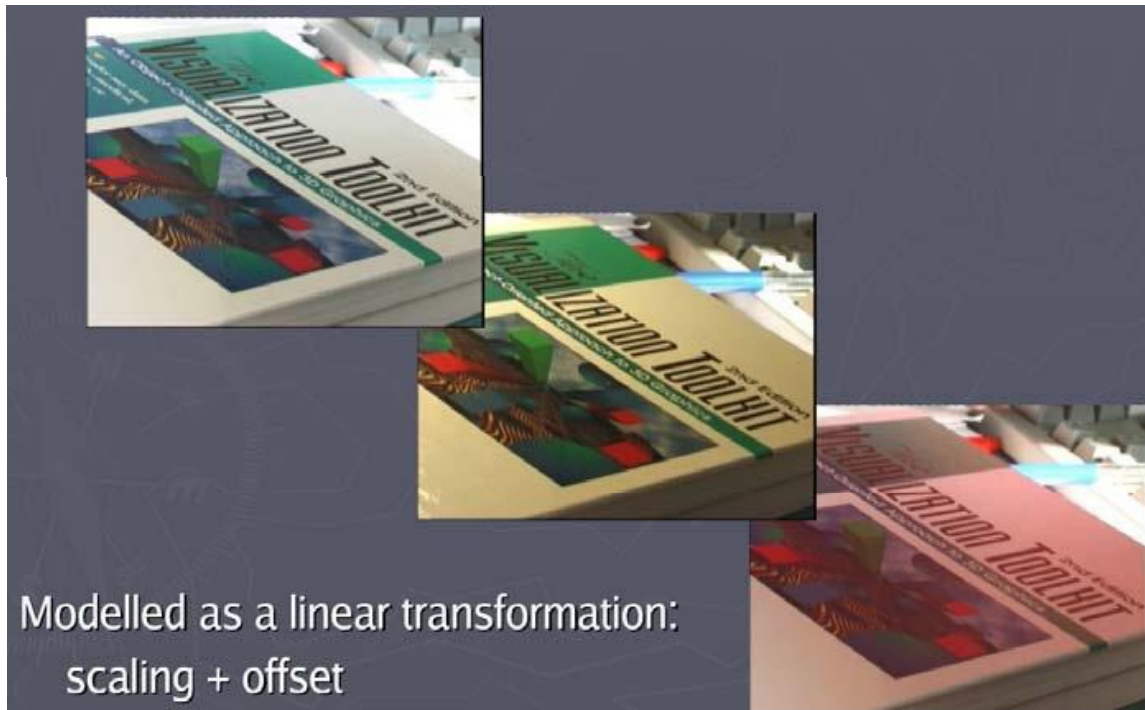
局部区域描述： 示例

- 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



局部区域描述： 示例

- 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





局部区域描述：SIFT描述子

- 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

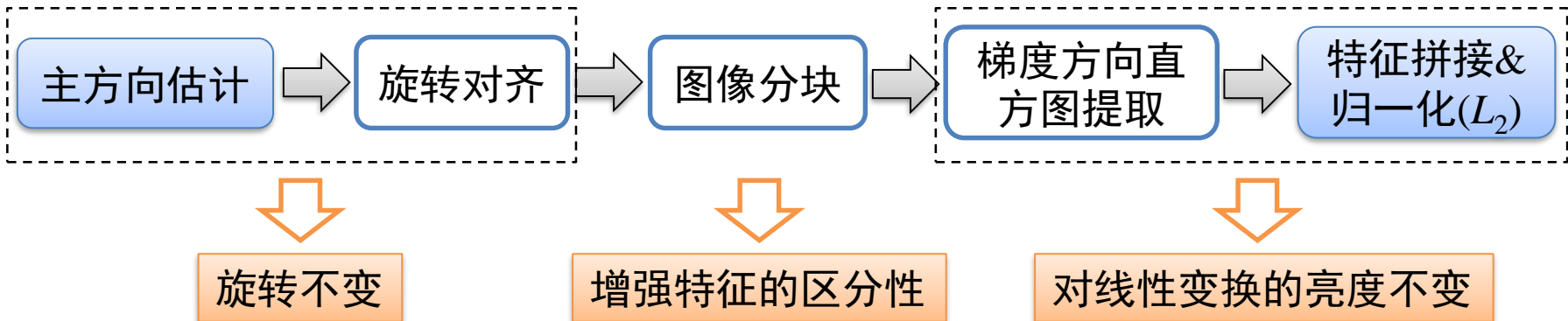
Distinctive image features from scale-invariant keypoints

[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 ...

☆ 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



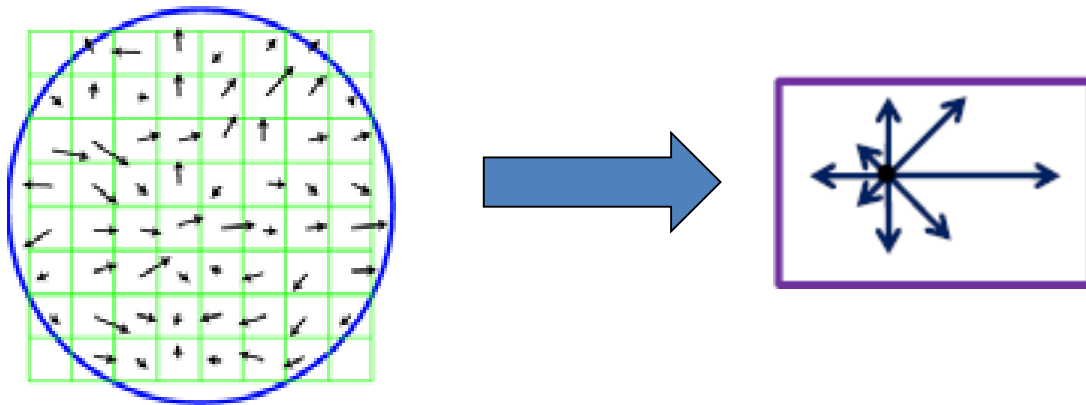
局部区域描述：SIFT描述子

□ 主方向估计

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

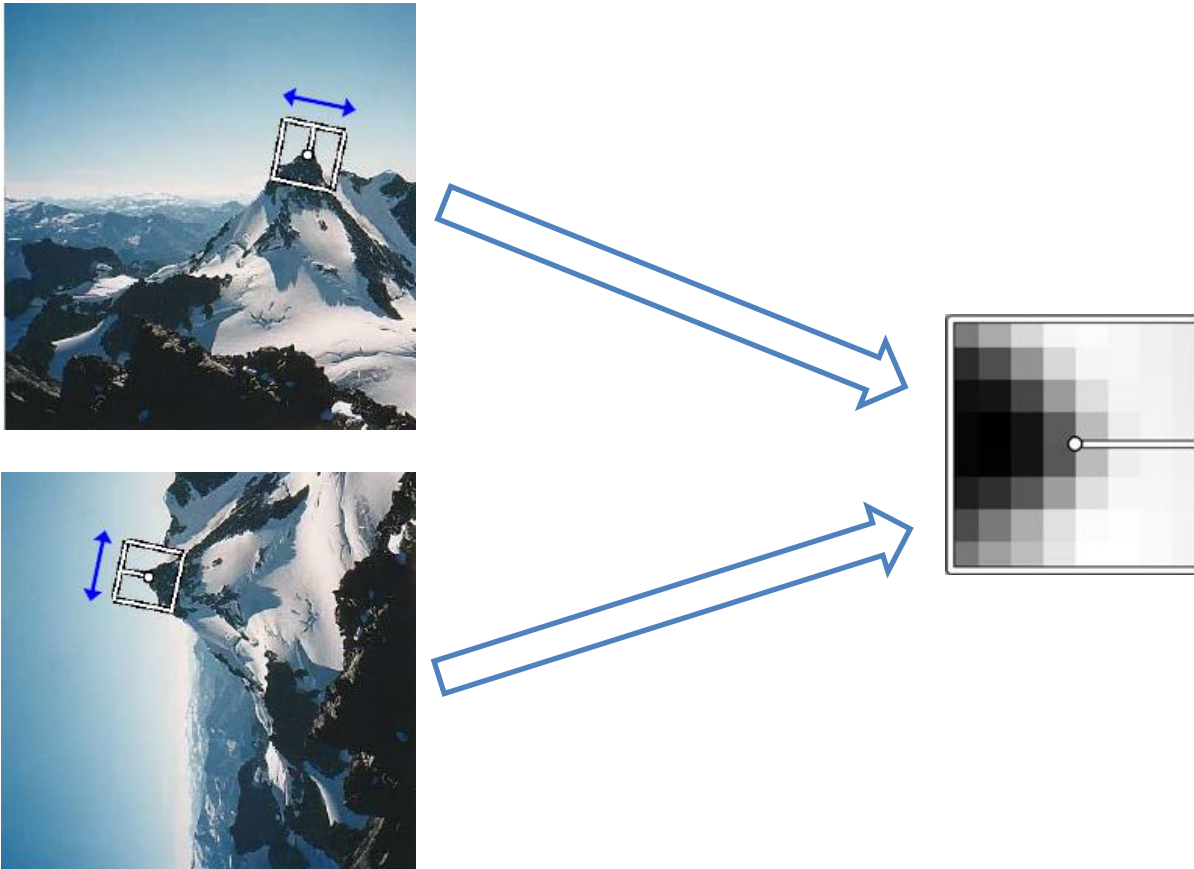
□ 具体过程

- 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

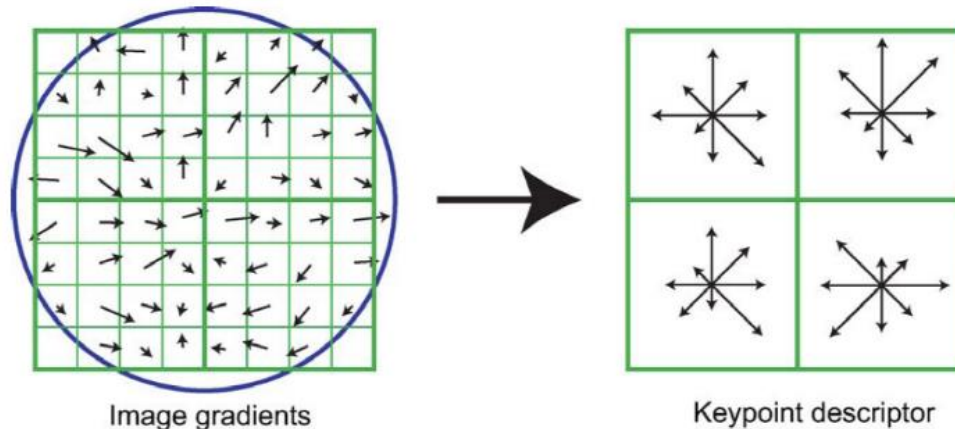
- 基于主方向，对图像块旋转对齐
 - 将图像块修正到一个canonical orientation.



局部区域描述：SIFT描述子

□ SIFT描述子生成

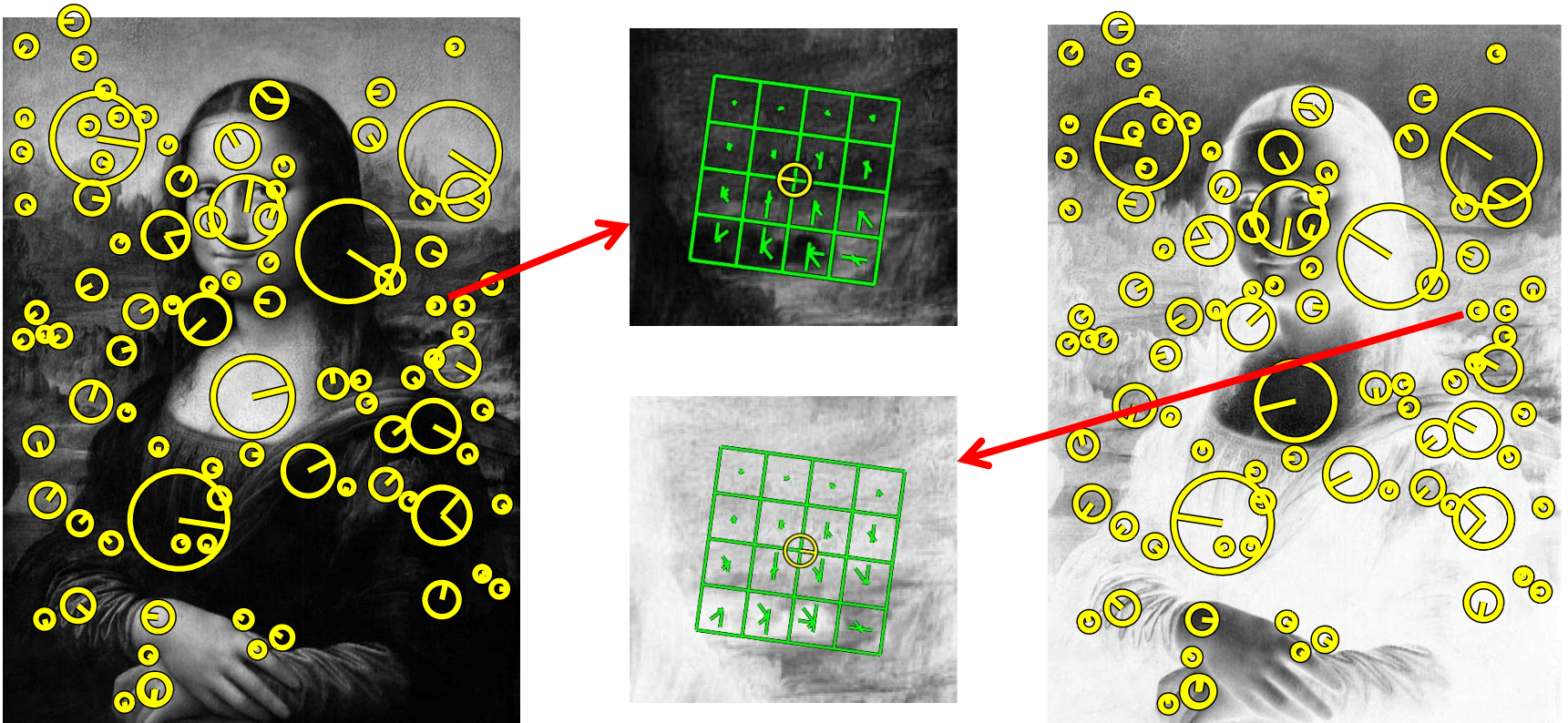
- 将图像区域缩放到固定大小，均分为 4×4 的小块
- 逐块计算梯度方向直方图（8-D）
- 将各块直方图拼接，得到128维的梯度方向直方图
- 最后，对直方图进行 L_2 归一化处理
 - ✓ 实际实现时，会将归一化的特征向量的模设为512，并裁减最大值为255，这样每一维用一个字节即可存储。



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。**

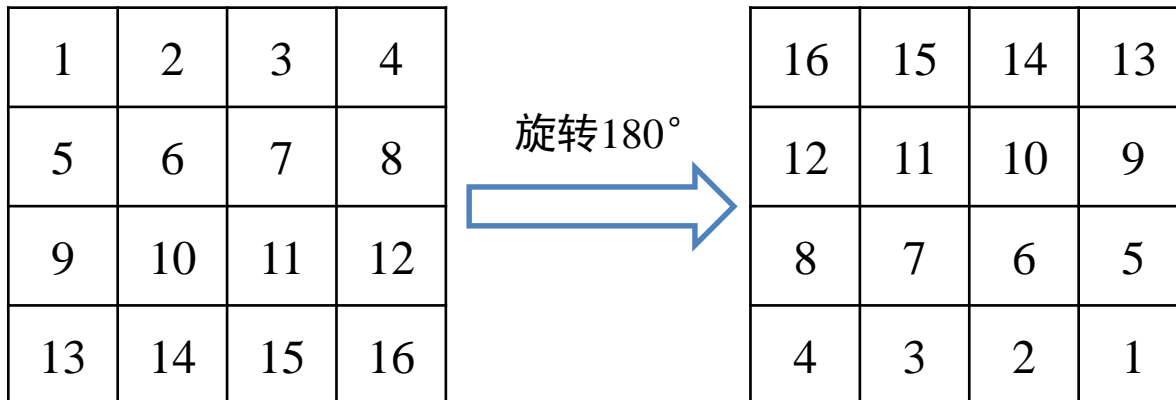
图像反色对SIFT特征提取的影响

- 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





局部视觉特征描述

□ 局部关键点检测

■ 角点检测

✓ Harris 角点检测子

■ 块检测

✓ 高斯差分(DoG) 检测子

✓ MSER 检测子

□ 局部区域描述

■ SIFT

■ LIOP

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

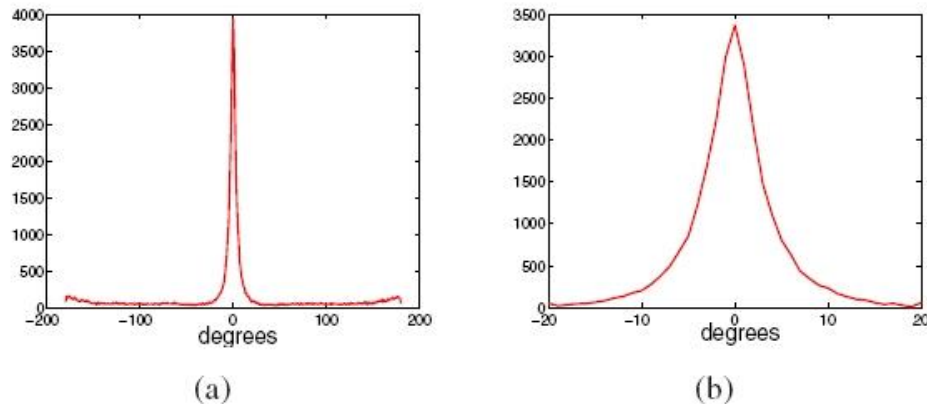


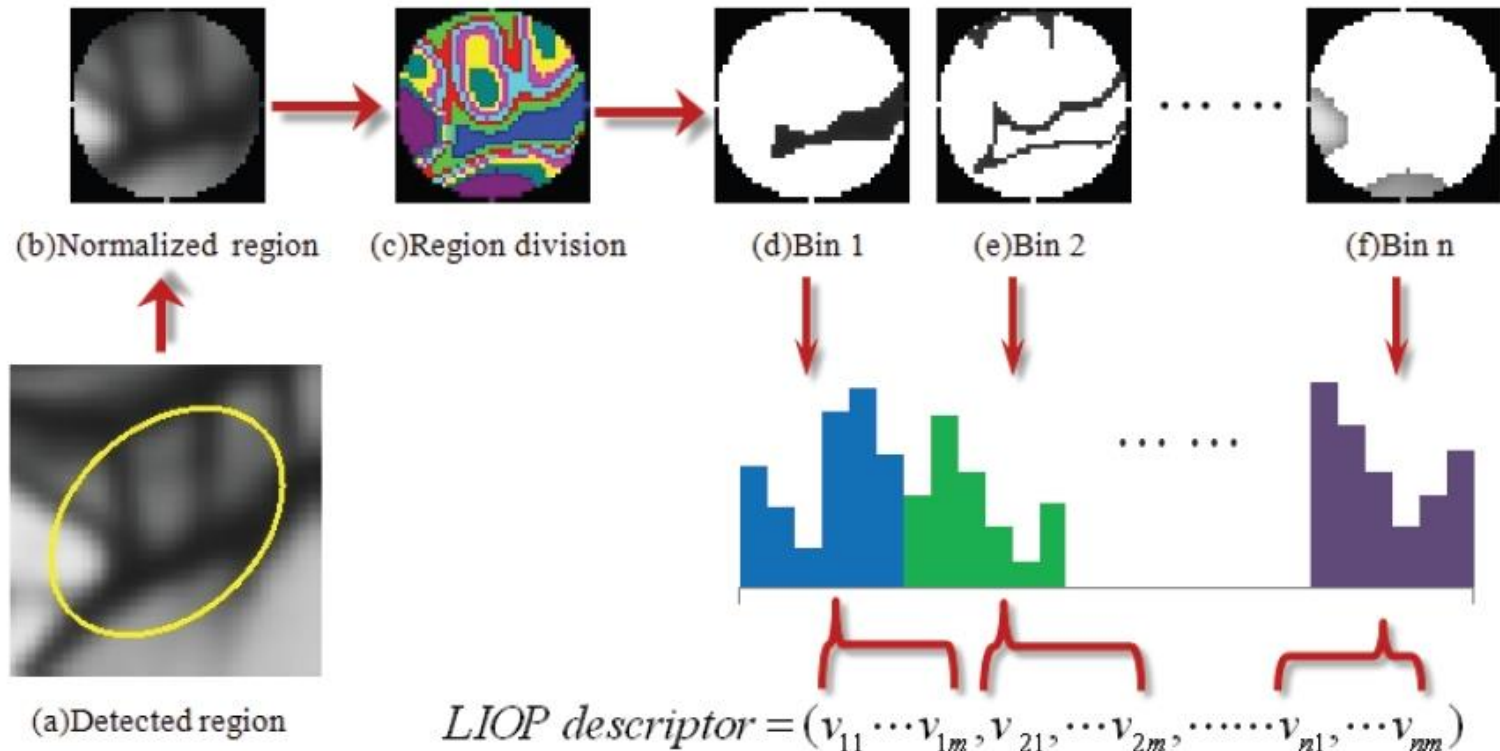
Figure. Orientation assignment errors. (a) Between corresponding points, only 63.77% of errors are in the range of $[-20,20]$. (b) Between corresponding points that are also matched by SIFT descriptors.

- 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

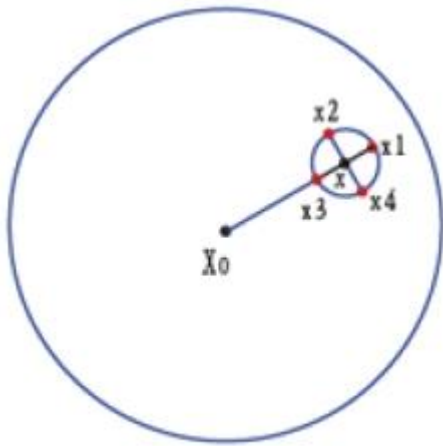
- 基本假设：图像区域中像素灰度的**相对大小关系**保持不变
- 利用像素的灰度值排序，将图像划分为 n 个不规则的区域
- 每个区域提取一个 m 维的特征向量，最后将所有区域的特征进行拼接



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



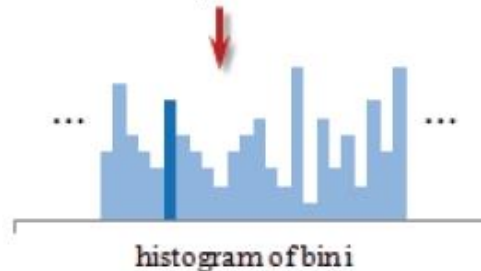
(a) Original patch

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



基于局部特征的图像表达

- 图像表达基本框架
- 局部视觉特征描述
 - 局部关键点检测
 - ✓ 角点检测：Harris
 - ✓ 块检测：DoG, MSER
 - 局部区域描述
 - ✓ SIFT
- 特征编码与聚合
 - Fisher Vector
 - Bow
 - VLAD
 - PQ



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

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

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

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

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

- 计算对数似然函数关于各参数的梯度值：

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

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

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

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

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

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



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:

$$b_i \propto \sum_{t=1}^T \gamma_t(i) \text{ for the } i\text{-th visual word}$$

- **Sparseness** is ensured with large N
 - ✓ Suitable for the inverted index

□ VLAD (vector of locally aggregated descriptor)

- The gradient with respect to the mean vector of GMM

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

- Disadvantage
 - ✓ **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

Of all the sensory impressions proceeding from the eye, the visual experience is the most important. Our perception of the world is based on the visual information that reaches the eye. For a long time, the visual system has been a topic of interest to scientists. Hubel and Wiesel's work on the retina and the cerebral cortex has shown that the visual system is a complex system that processes information in a hierarchical manner. The visual system is a complex system that processes information in a hierarchical manner. The visual system is a complex system that processes information in a hierarchical manner.

Retrieve

China is forecasting a trade surplus of \$90bn (£51bn) to \$100bn this year, a threefold increase over 2004's \$32bn. The Chinese government said the surplus would rise to \$70bn in 2005, a rise in imports from the US, which has long complained that China's exports are unfairly helped by a deliberately undervalued yuan.

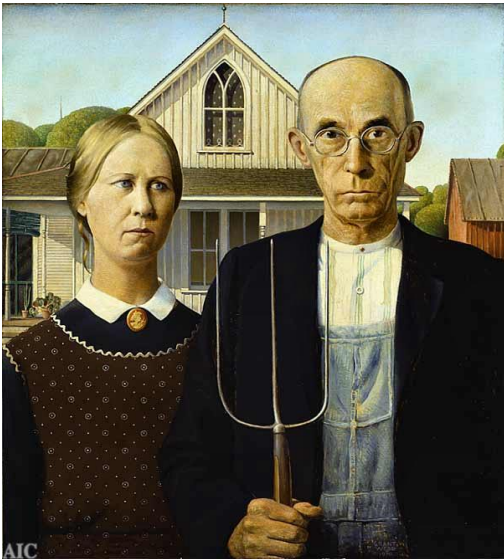
视觉词袋模型 (Bag of Visual Words)

- 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?

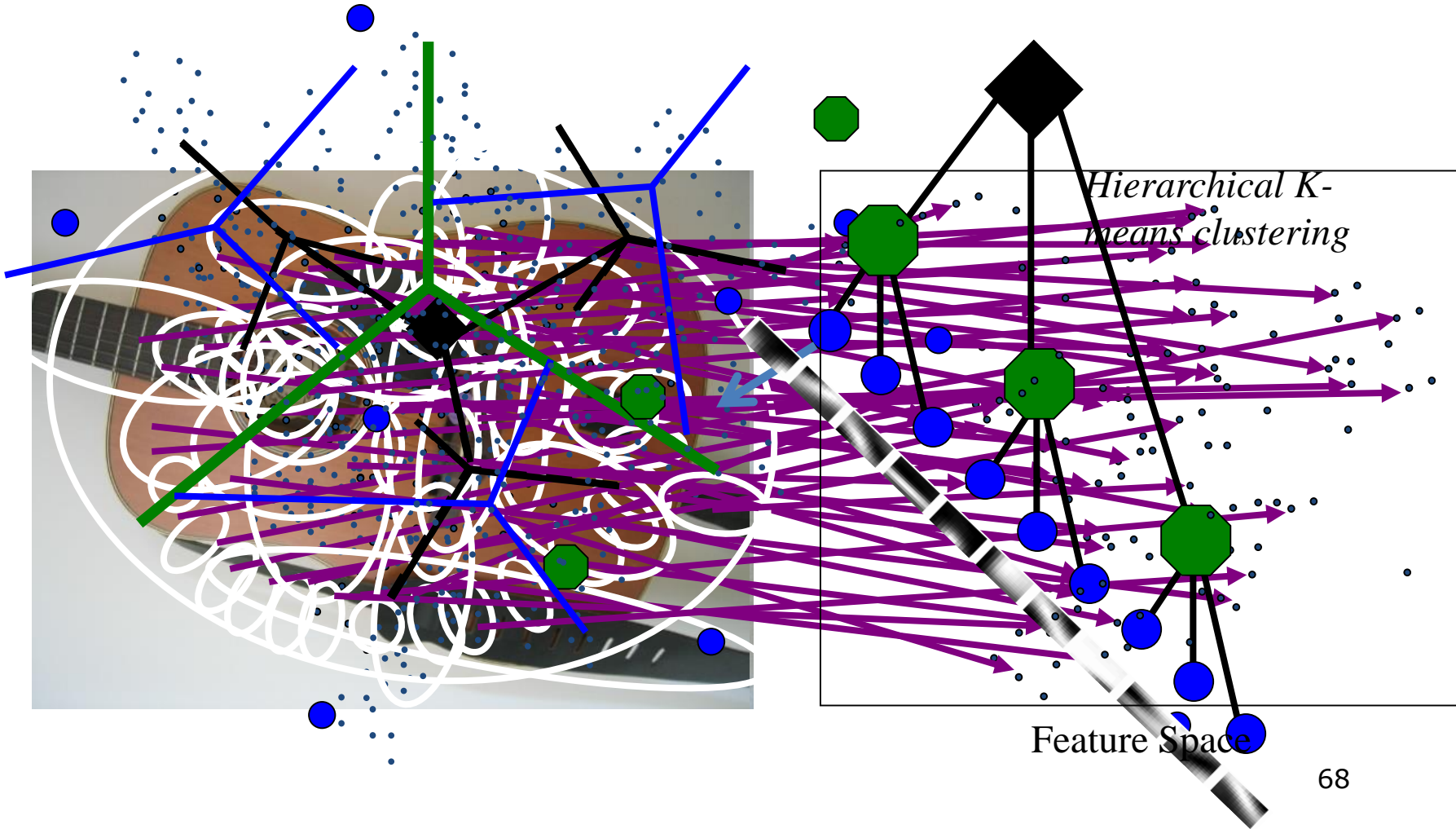
- ✓ A2: 基于视觉码本, 通过矢量量化, 将局部视觉特征量化到最近的视觉单词。
- ✓ 有损压缩 (lossy compression), 表达紧凑。

■ 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

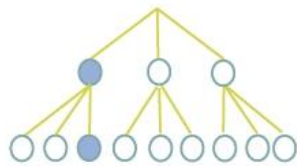
□ 通过矢量量化，实现特征匹配

- 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

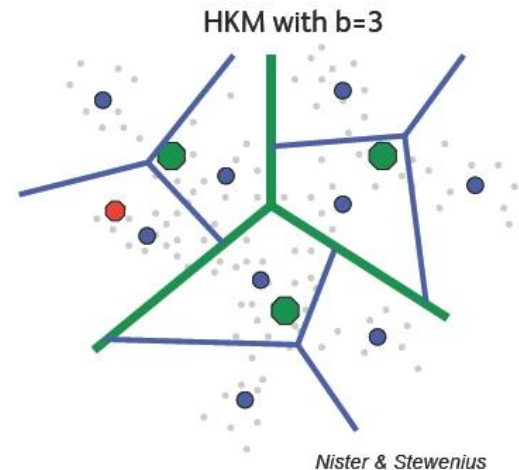
- Hierarchical k-means [Nister 06]

- ▶ K-means tree of height h



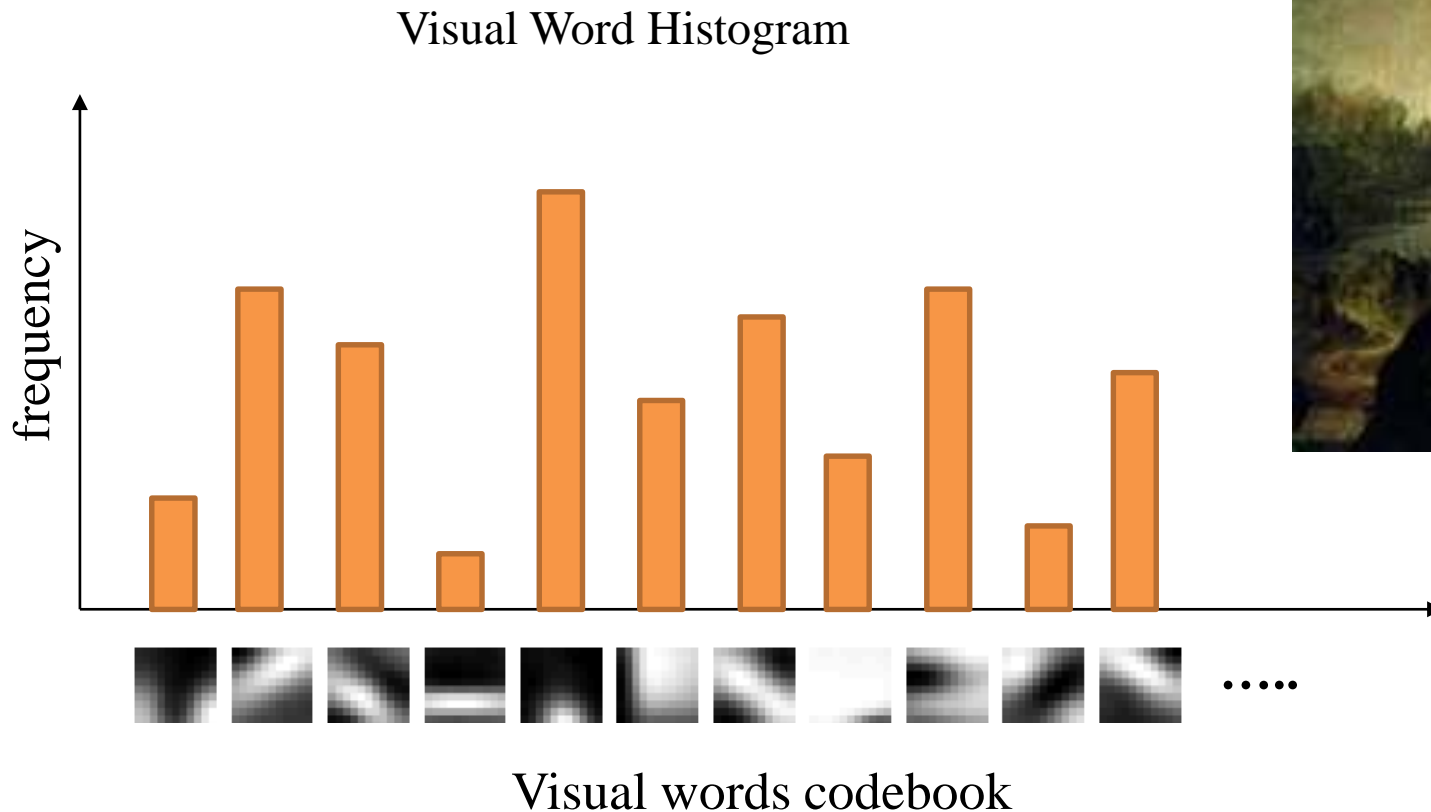
- ▶ Branching factor b : $k = b^h$
- ▶ Assignment Complexity:

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



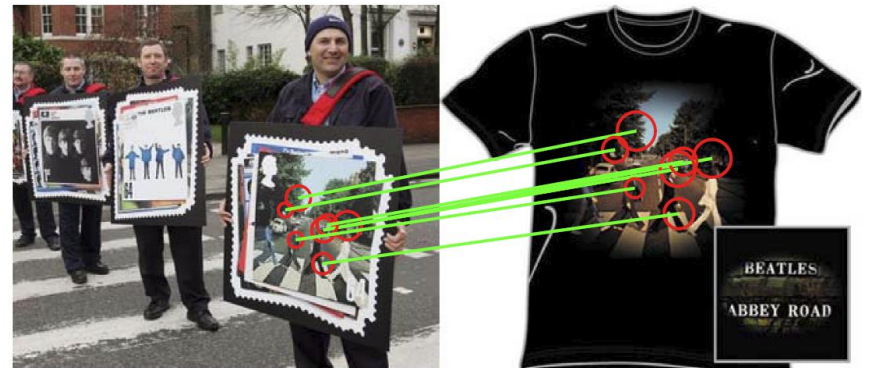
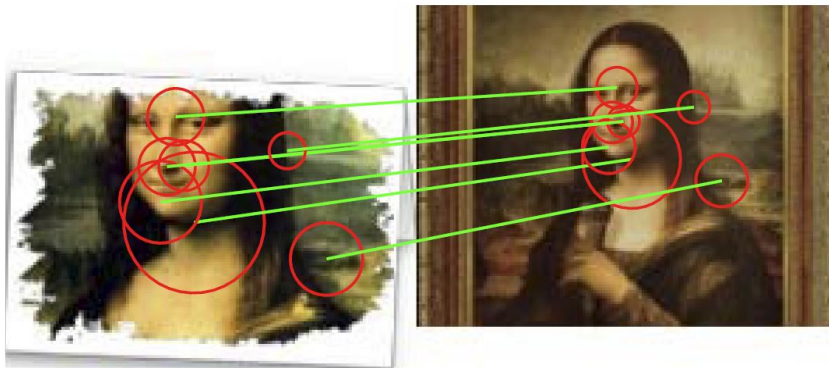
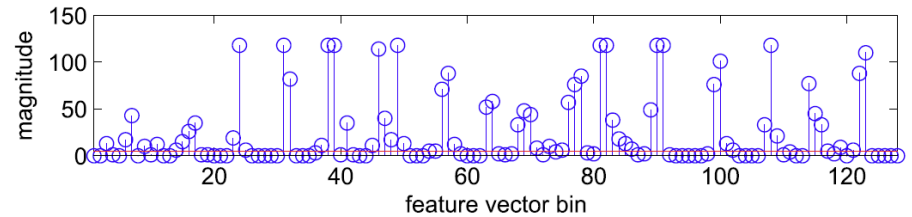
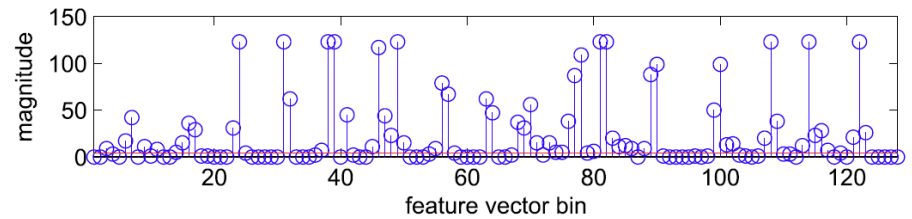
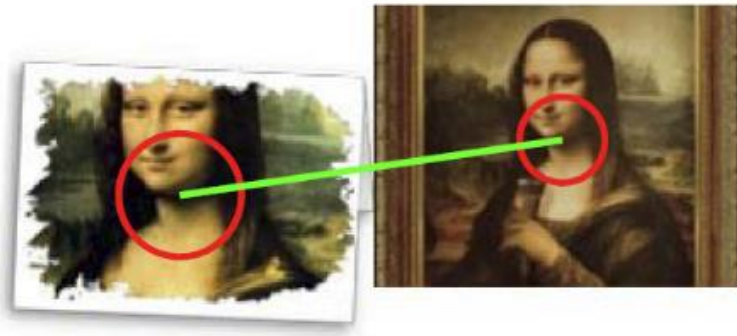
视觉词袋模型

- 将一副图像表达为视觉单词的分布直方图



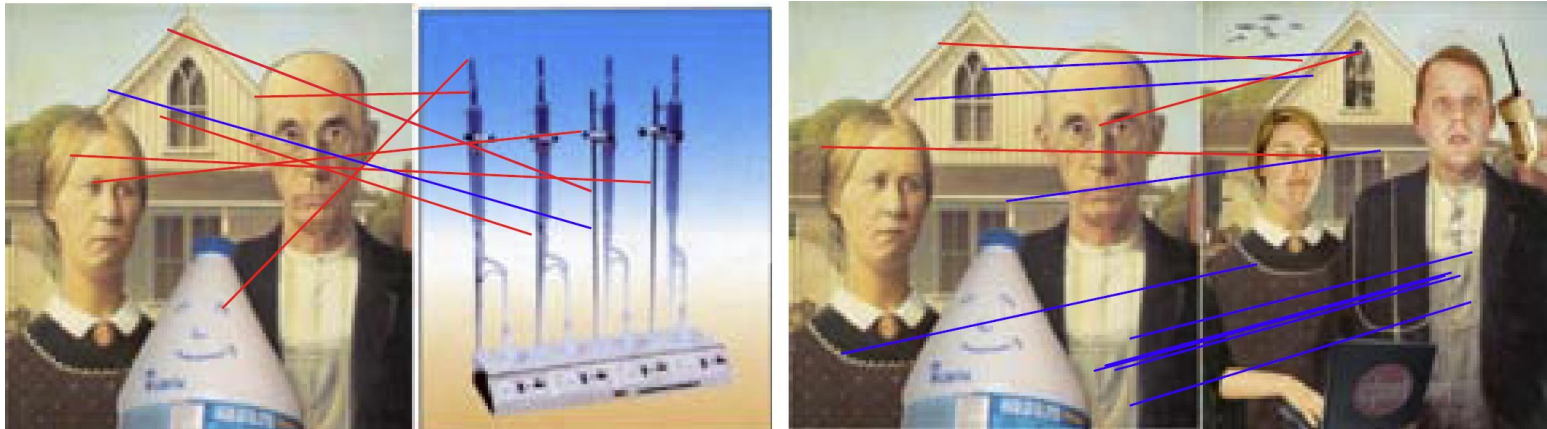
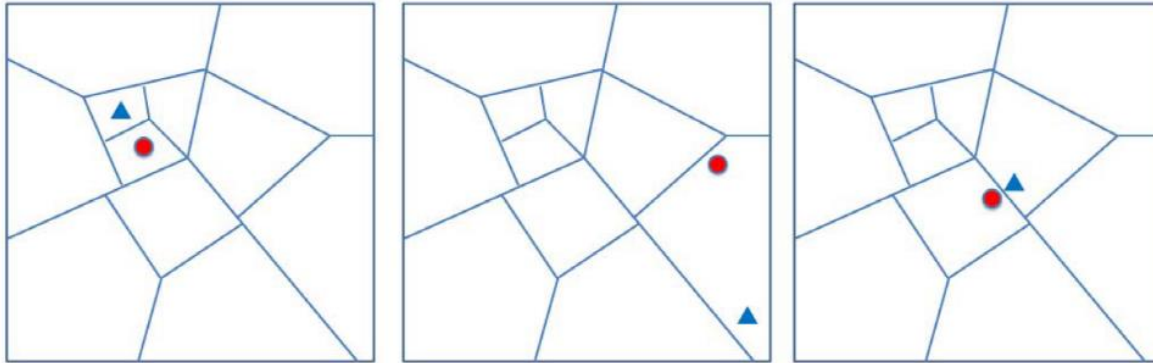
基于词袋模型的图像匹配-I

- 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



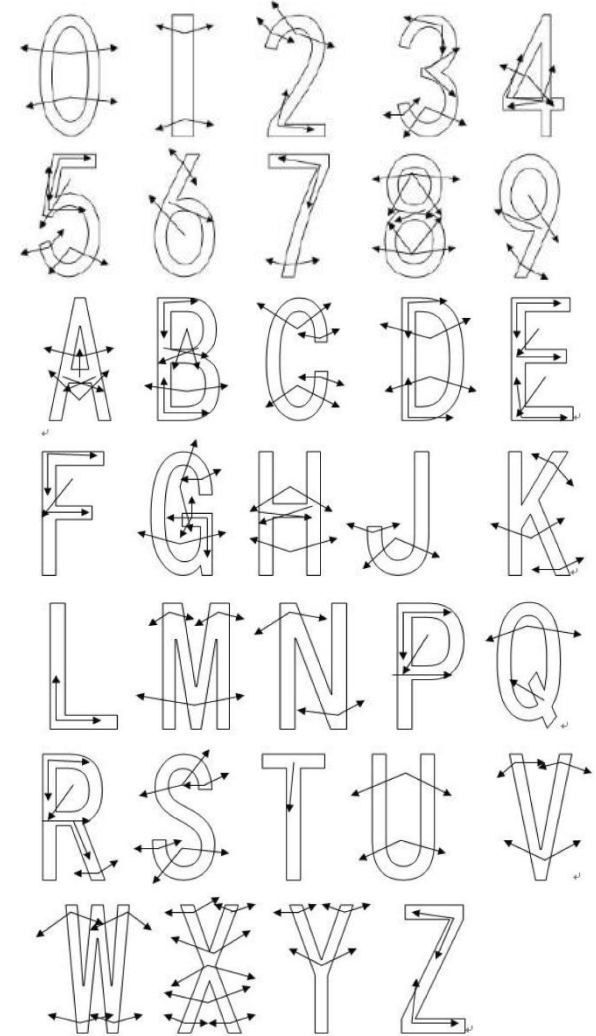
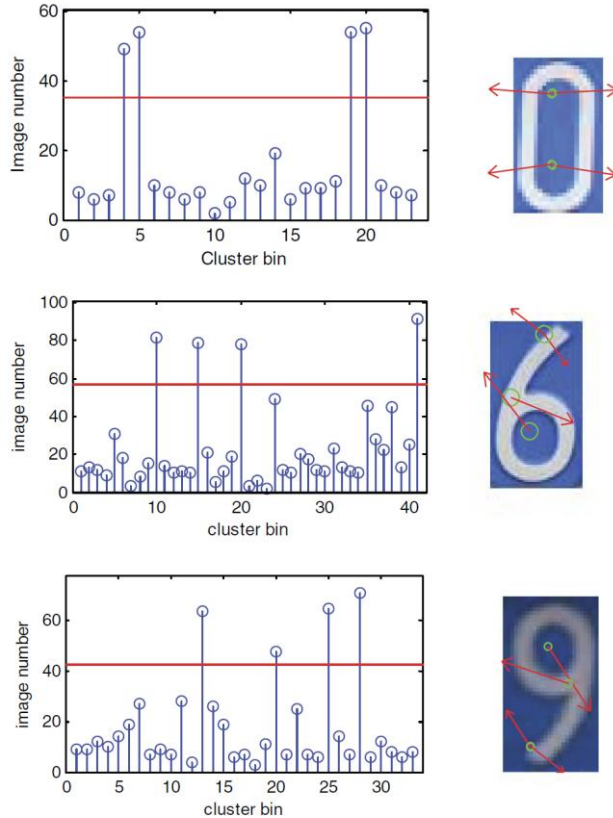
基于词袋模型的图像匹配-II

- 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.

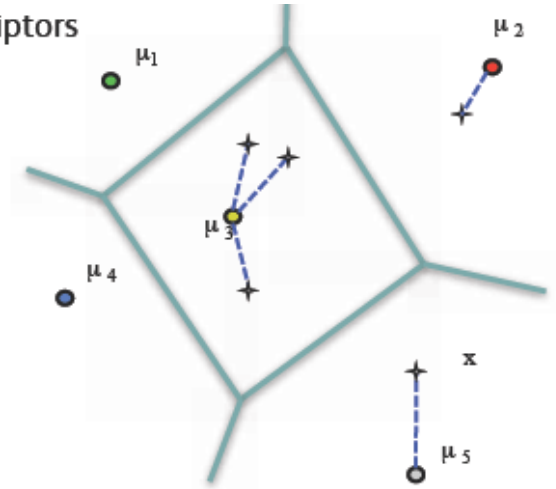
局部聚合描述子向量

(VLAD: Vector of Locally Aggregated Descriptor)

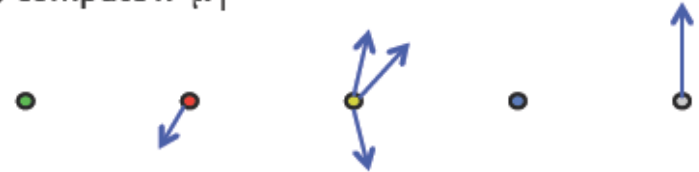
Given a codebook $\{\mu_i, i = 1 \dots N\}$,
 e.g. learned with K-means, and a set of
 local descriptors $X = \{x_t, t = 1 \dots T\}$:

- 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

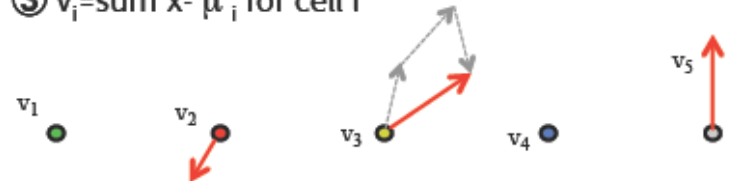
① assign descriptors



② compute $x - \mu_i$

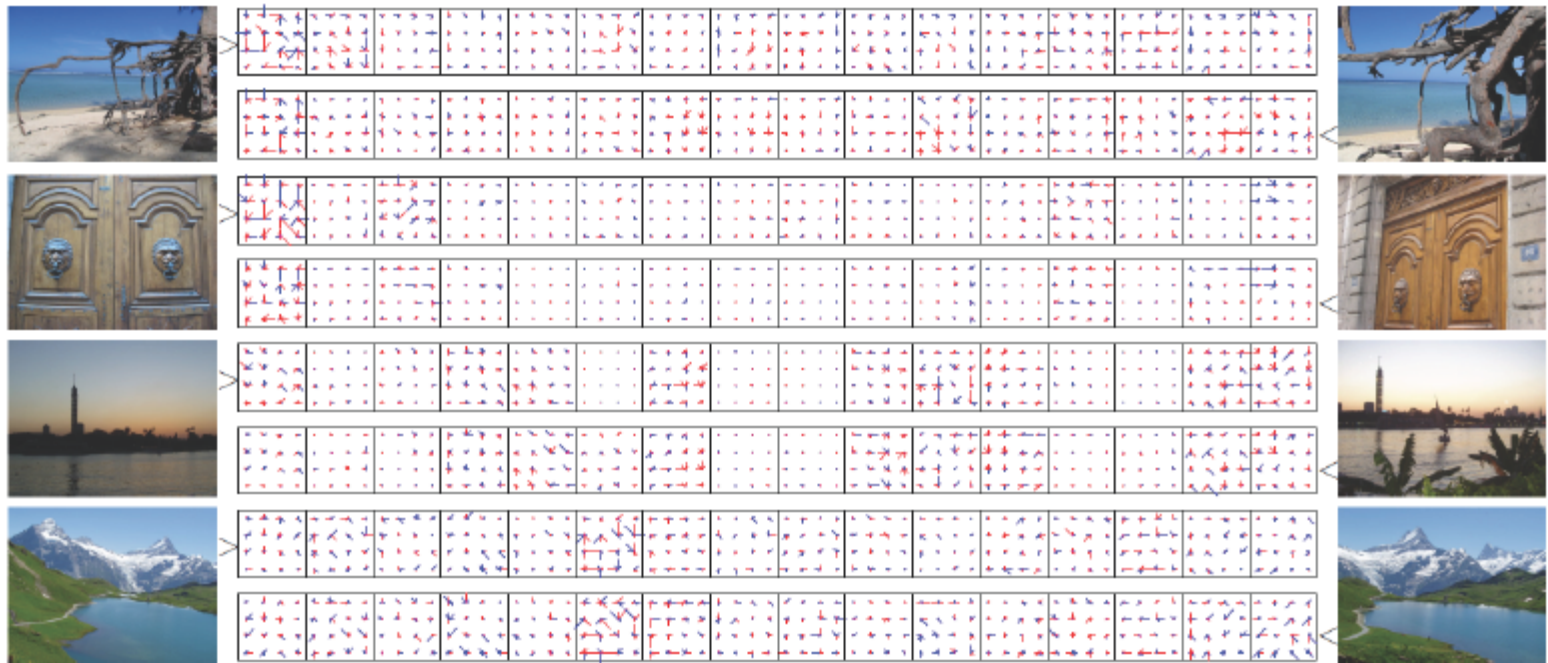


③ $v_i = \sum x - \mu_i$ for cell i



局部聚合描述子向量：实例

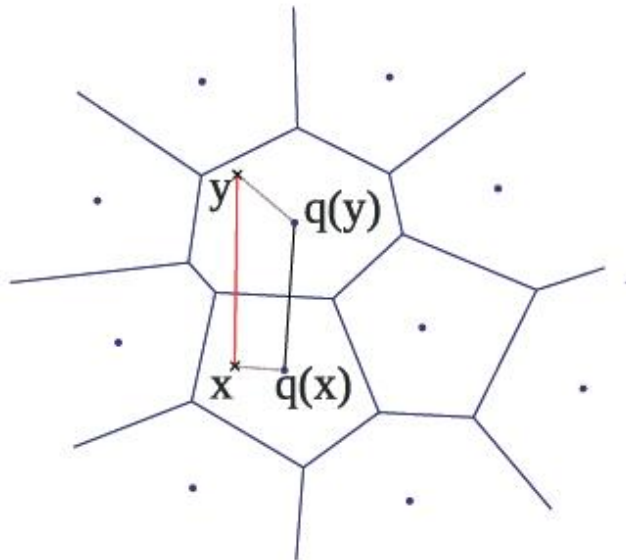
A graphical representation of $v_i = \sum_{x_t: \text{NN}(x_t) = \mu_i} x_t - \mu_i$



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

$$\underbrace{x_1, \dots, x_{D^*}, \dots, x_{D-D^*+1}, \dots, x_D}_{u_1(x)} \quad \underbrace{\hspace{10em}}_{u_m(x)}$$
$$\rightarrow q_1(u_1(x)), \dots, q_m(u_m(x)),$$

■ 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}
⋮			
d_{K1}			d_{KK}

乘积量化 (Product Quantization)

□ 误差分析

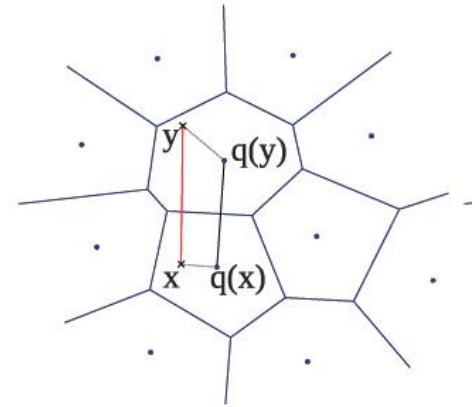
- 均方距离误差(MSDE)的上界是量化器的均方差(MSE)

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

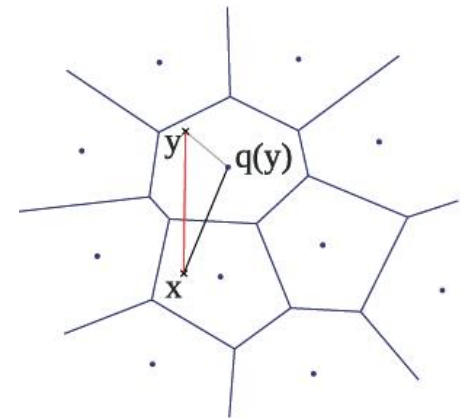
$$(d(x, y) - d(x, q(y)))^2 \leq d(y, q(y))^2.$$

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

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



symmetric case



asymmetric case