读书笔记一：导论大纲

前言

作为CV（Computer Vision）方向的研究生，一直想要系统地了解和学习一下CV的基本知识，理清CV发展的脉络。而David Forsyth的《Computer Vision-A Modern Approach(2nd edition)》是CV领域的经典教材之一，介绍了许多CV领域的基本知识和现代技术，按照图像形成、早期（低层）视觉、中层视觉、高层视觉以及CV的实际应用来组织内容，脉络清晰。本人断断续续花了一年左右的时间把这本书看了个大概，觉得有必要把阅读此书过程中做的一些读书笔记以思维导图的形式记录下来，写成一个系列，方便日后浏览和复习。笔记是在翻译的基础上加入了一些个人的理解和认识，不妥之处欢迎评论交流！

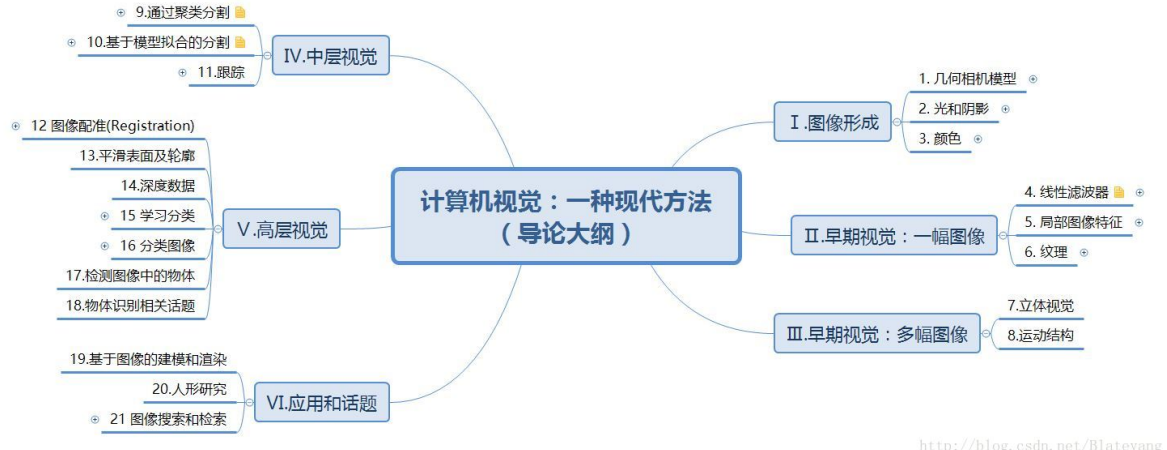
导论大纲

下图为书中作者建议的适合计算机科学、电子工程及其他工科高年级或研一学生学习的计算机视觉一学期导论课。

Week	Chapter	Sections	Key topics
1	1, 2	1.1, 2.1, 2.2.x	pinhole cameras, pixel shading models, one inference from shading example
2	3	3.1–3.5	human color perception, color physics, color spaces, image color model
3	4	all	linear filters
4	5	all	building local features
5	6	6.1, 6.2	texture representations from filters, from vector quantization
6	7	7.1, 7.2	binocular geometry, stereopsis
7	8	8.1	structure from motion with perspective cameras
8	9	9.1–9.3	segmentation ideas, applications, segmentation by clustering pixels
9	10	10.1–10.4	Hough transform, fitting lines, robustness, RANSAC,
10	11	11.1-11.3	simple tracking strategies, tracking by matching, Kalman filters, data association
11	12	all	registration
12	15	all	classification
13	16	all	classifying images
14	17	all	detection
15	choice	all	one of chapters 14, 19, 20, 21 (application topics)

21 Image Search and Retrieval

按此大纲列出的全书大体架构如下



后续同系列博文安排

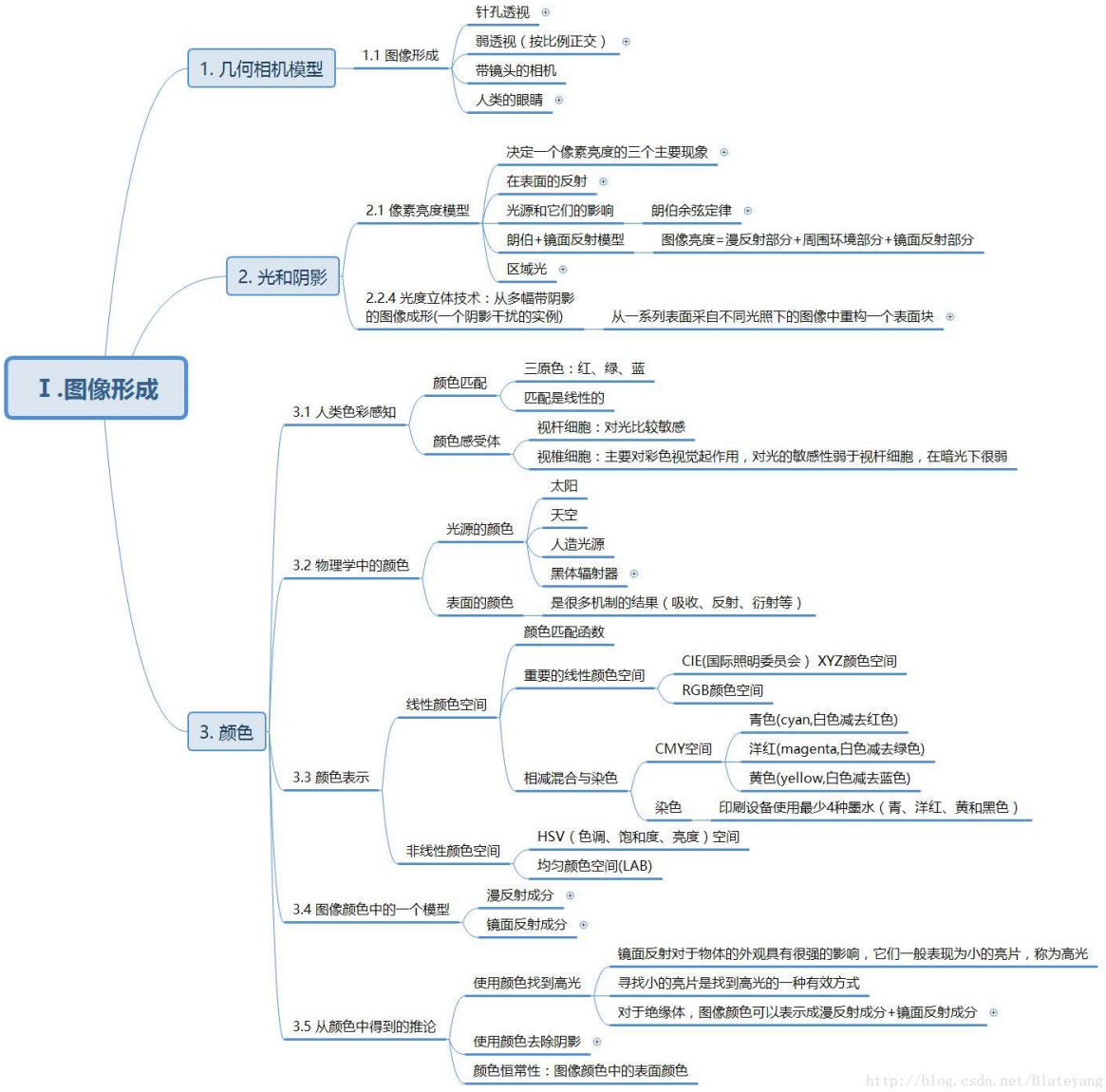
后面我会贴出自己看过的以下内容的思维导图和笔记整理

1. 图像形成：几何相机模型、光和阴影、颜色
2. 早期视觉一幅图像：线性滤波器、局部图像特征、纹理

3. 中层视觉：通过聚类分割、基于模型拟合的分割、跟踪
4. 高层视觉：图像配准、学习分类、分类图像
5. 图像搜索和检索

读书笔记二：图像形成

本篇思维导图



相关读书笔记

1.几何相机模型

1.1 图像形成

1.针孔透视（其实就是小时候玩的小孔成像）

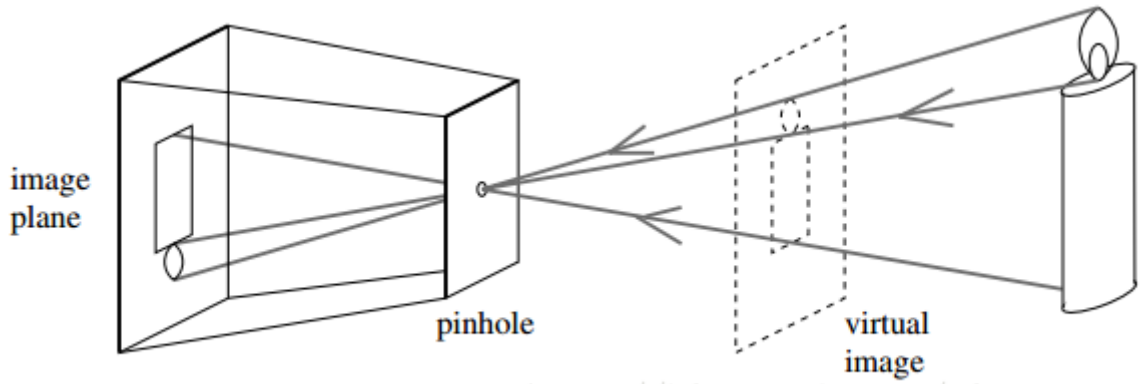


FIGURE 1.2: The pinhole imaging model.

2.仿射投影 弱透视投影

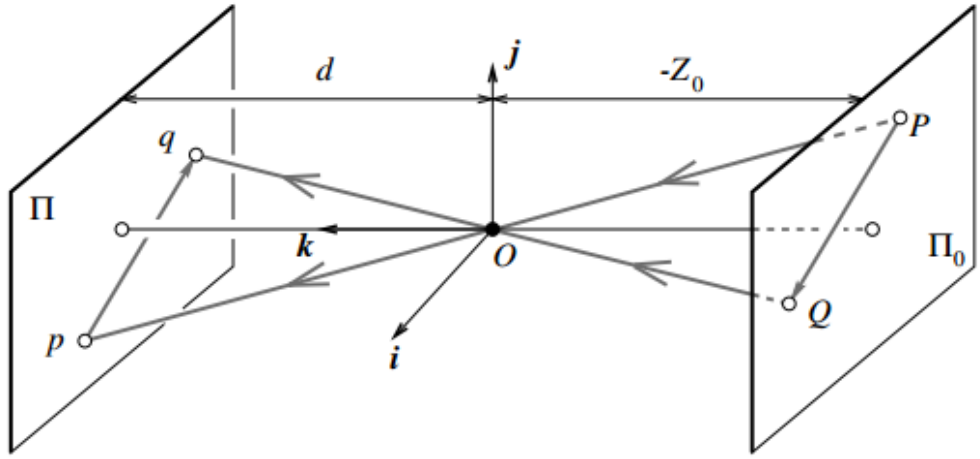
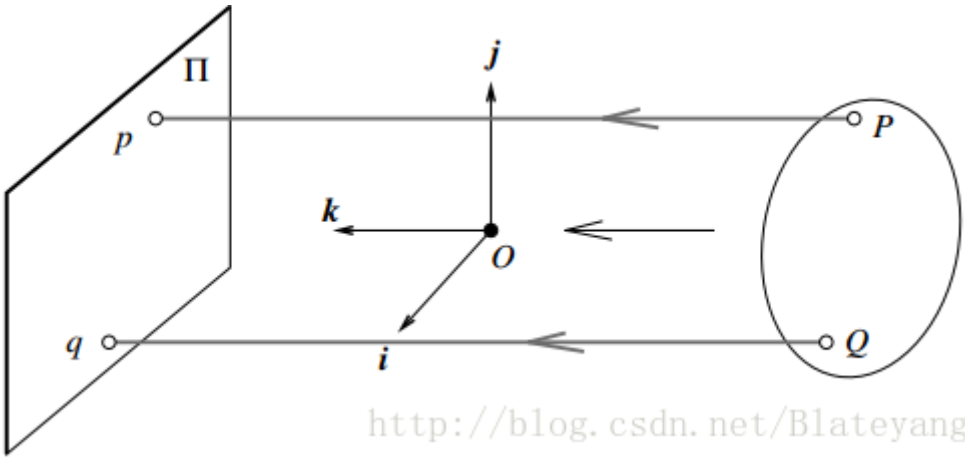


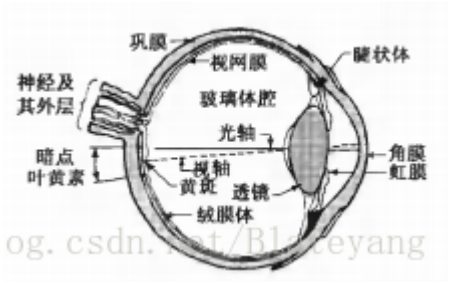
FIGURE 1.5: Weak-perspective projection. All line segments in the plane Π_0 are projected with the same magnification.

在上图中，当d=-Z0时，称为正交投影



3.带镜头的相机

4.人类的眼睛



可以看作是一个相机模型

2.光和阴影

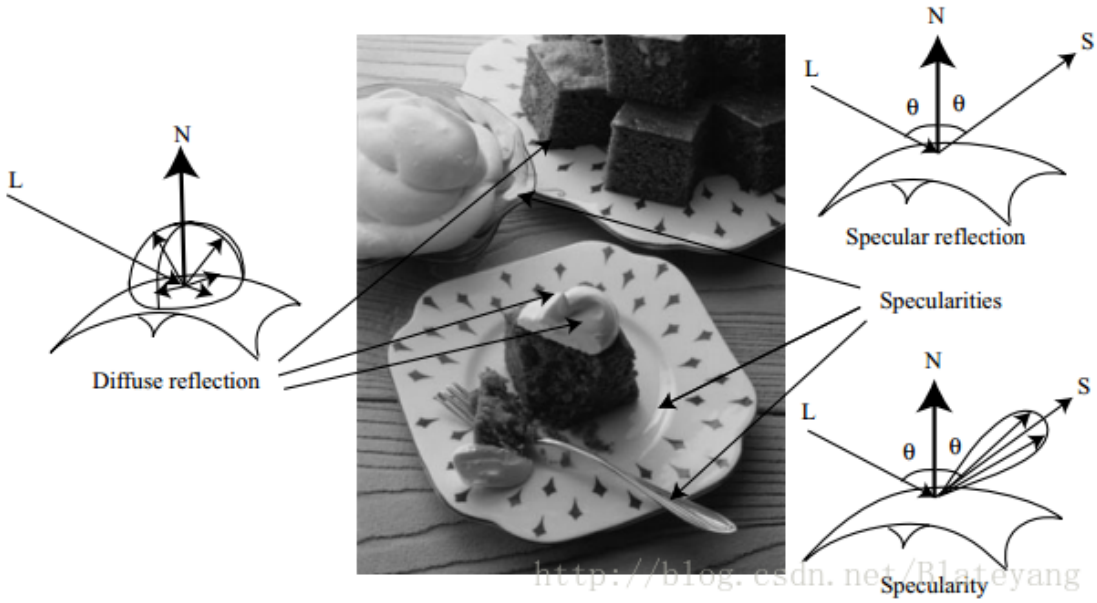
2.1 像素亮度模型

1.决定一个像素亮度的三个主要现象

- 光照：照射到物体表面的光-
- 表面反射：光从表面反射到相机的部分
-
- 相机对光的响应：

$$Icamera(x) = kIpatch(x)$$

2.在表面的反射

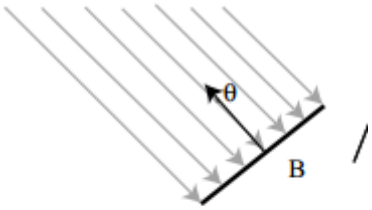


3.光源和它们的影响

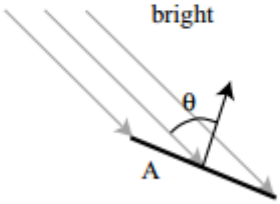
朗伯余弦定律：

```
I=ρI0cosθ
```

,θ是入射角与表面法线的夹角， ρ是反射率



Diffuse reflection, bright



Diffuse reflection, dark

4.朗伯+镜面反射模型

图像亮度=漫反射部分+镜面反射部分+周围环境部分

5.区域光

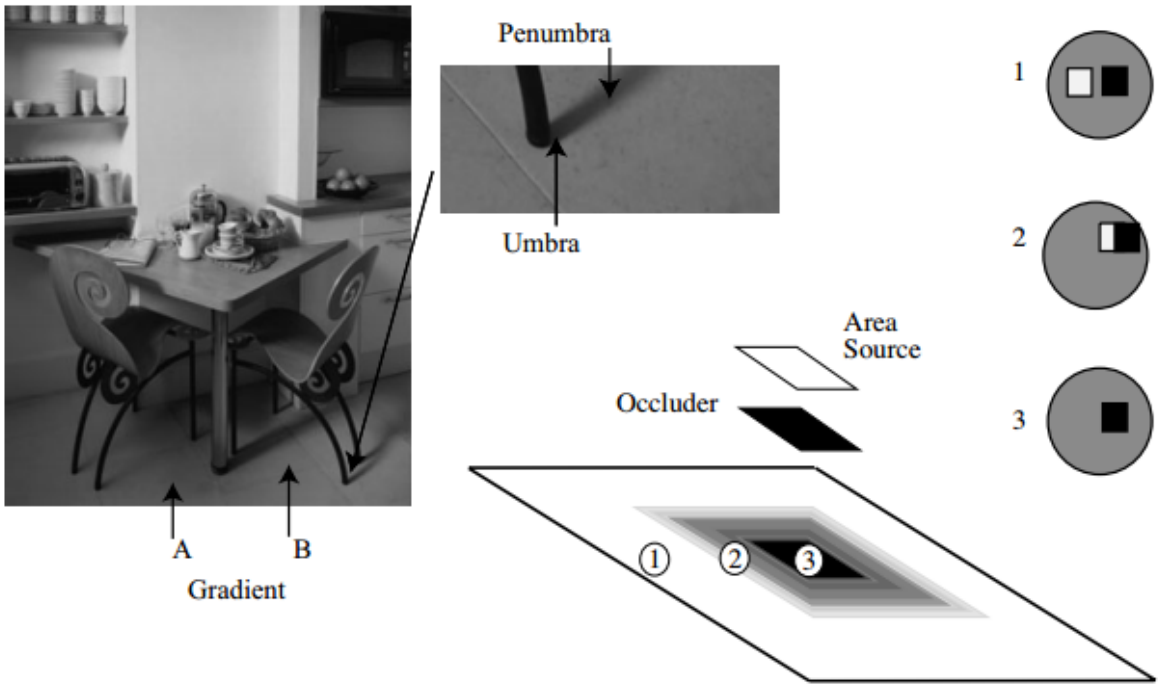
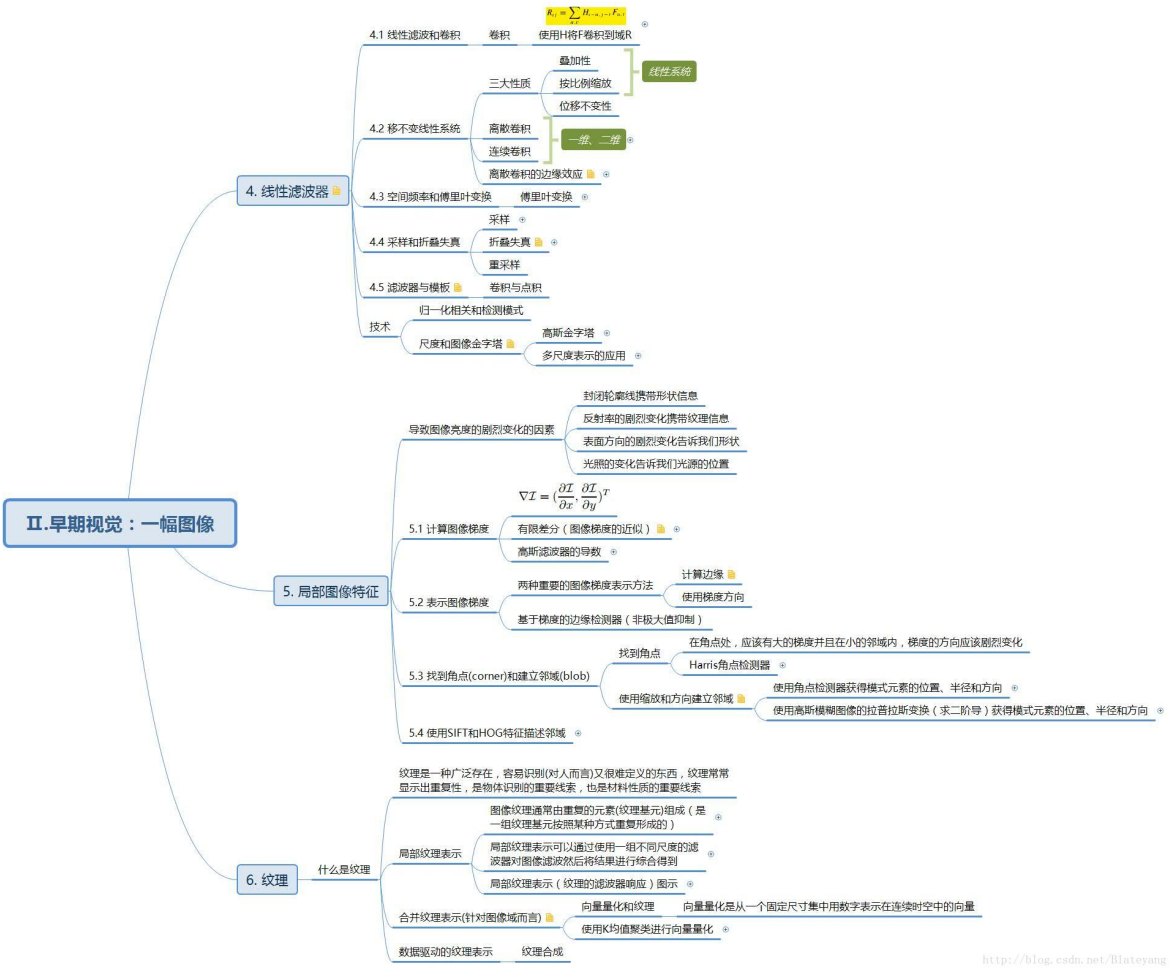


FIGURE 2.3: Area sources generate complex shadows with smooth boundaries, because from the point of view of a surface patch, the source disappears slowly behind the occluder. **Left:** a photograph, showing characteristic area source shadow effects. Notice that A is much darker than B; there must be some shadowing effect here, but there is no clear shadow boundary. Instead, there is a fairly smooth gradient. The chair leg casts a complex shadow, with two distinct regions. There is a core of darkness (the *umbra*—where the source cannot be seen at all) surrounded by a partial shadow (*penumbra*—where the source can be seen partially). A good model of the geometry, illustrated **right**, is to imagine lying with your back to the surface looking at the world above. At point 1, you can see all of the source; at point 2, you can see some of it; and at point 3, you can see none of it. *Peter Anderson © Dorling Kindersley, used with permission.*

读书笔记三：早期视觉（一幅图像）

本篇思维导图



线性滤波器

线性滤波器的主要策略是用不同的加权模式计算像素加权和，以寻找不同的图像模式

1.线性滤波和卷积

- 二维卷积公式：

$$Rij = \sum u, vHi - u, j - vFu, v$$

,使用卷积核H将图像F卷积到域R

几种常见的线性滤波：平均平滑、高斯平滑、导数和有限差分

2.移不变线性系统

2.1 三大性质

- 叠加性-
- 按比例缩放
- 位移不变性

2.2 离散卷积和连续卷积的性质

- 对称的:

$$(g * h)(x) = (h * g)(x)$$

- 可结合的:

$$(f * (g * h)) = ((f * g) * h)$$

2.3 离散卷积的边缘效应

在边缘处计算有些像素位置的卷积值时，需要虚拟并不存在的图像值,采用的策略有

- 忽略这些点
- 使用常数填充
- 使用其他方法填充图像

3 空间频率和傅里叶变换

二维傅里叶变换：$$\mathcal{F}(g(x,y))(u,v)=\int_{-\infty}^{\infty}\int g(x,y)e^{-i2\pi(ux+vy)}dx\,dy$$

效果图：

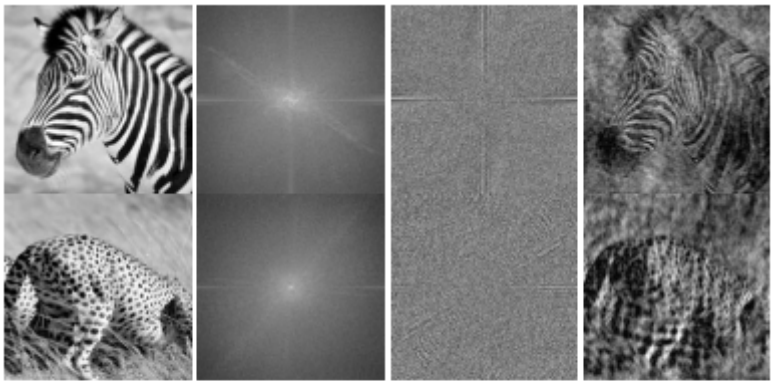


FIGURE 4.6: The second image in each row shows the log of the magnitude spectrum for the first image in the row; the third image shows the phase spectrum scaled so that $-\pi$ is dark and π is light. The final images are obtained by swapping the magnitude spectra. Although this swap leads to substantial image noise, it doesn't substantially affect the interpretation of the image, suggesting that the phase spectrum is more important for perception than the magnitude spectrum.

4.采样和折叠失真

- 采样

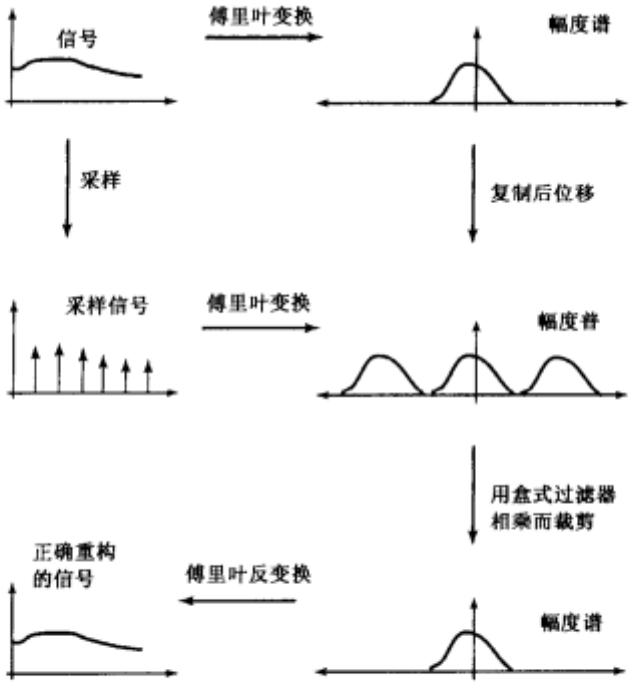


图 7.10　采样信号的傅里叶变换由原始信号的傅里叶变换根据采样频率偏移后的复制组合构成。有两种可能性：如果偏移的部分互不相交叠（这里的情况），原始信号能够根据采样信号重构（裁剪傅里叶变换中的一块，进行反变换）。如果偏移部分相交叠（见图7.11），则由于相交叠区域被叠加，无法获得单独的傅里叶变换，信号是折叠失真的

- 折叠失真

  原始信号中的高频空间元素在采样信号中会表现为低频元素，这种效应称为折叠失真

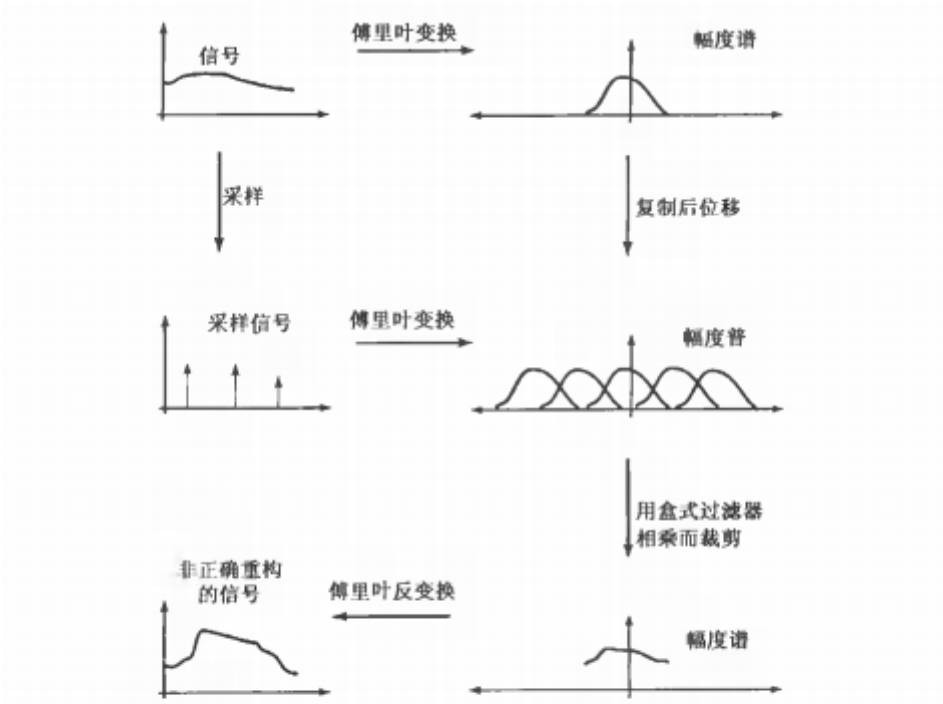


图 7.11 采样信号的傅里叶变换由原始信号的傅里叶变换根据采样频率偏移后的复制组合构成。有两种可能性：如果偏移的部分互不相交叠（见图7.10），原始信号能够根据采样信号重构（裁剪傅里叶变换中的一块，进行反变换）。如果偏移部分相交叠（这里的情况），相交叠区域被叠加，无法获得单独的傅里叶变换，信号是折叠失真的。这也说明高频信号具有向低频信号折叠失真的趋势

- 重采样

5.滤波器与模板

滤波器对类似滤波器的模式元素有很强的响应

6.技术

6.1 归一化相关和检测

6.2 尺度和图像金字塔

图像金字塔是以金字塔形状排列的分辨率逐步降低，且来源于同一张原始图的图像集合

1. 高斯金字塔

   在高斯金字塔中，每一层使用一个对称的高斯核进行平滑，同时进行重采样以获得下一层

2. 多尺度表示

   *多尺度搜索：小图像模式

   *空间搜索：在两张图片中找到对应点，由粗到精的匹配

   *特征跟踪：跟踪特征到较粗糙尺度，接收在较粗糙尺度下能找到对应的精细尺度特征

局部图像特征

导致图像亮度剧烈变化的因素：

- 封闭轮廓线携带形状信息-
- 反射率的剧烈变化携带纹理信息-
- 表面方向的剧烈变化也告诉我们形状-
- 光照的变化告诉我们光源的位置

1 计算图像梯度

$$\nabla I = (\partial I/\partial x, \partial I/\partial y)T$$

- 有限差分（图像梯度的近似）

$$\frac{\partial \mathcal{I}}{\partial x} = \lim_{\delta x \to 0} \frac{\mathcal{I}(x+\delta x, y) - \mathcal{I}(x,y)}{\delta x} \approx \mathcal{I}_{i+1,j} - \mathcal{I}_{i,j}.$$

注：简单的有限差分往往会对噪声有较强的相应，解决的方法是先平滑图像再对它进行差分（可在像素尺度上抑制噪声）

高斯滤波器的导数

$$\partial(G\sigma * *I)/\partial x = (\partial G\sigma/\partial x) * *I$$

2 表示图像梯度

- 两种重要的图像梯度表示方法：计算边缘、使用梯度方向-
- 基于梯度的边缘检测器

3 找到角点（corner）和建立邻域(blob)

3.1 找到角点

在角点处，应该有大的梯度并且在小的邻域内，梯度的方向应该剧烈变化

Harris角点检测器

- 原理：在角点窗口，图像高斯平滑后的二阶导数矩阵H的两个特征值都比较大
-
- 公式：寻找下式的局部极大值

  det(H)−k(trace(H)/2)^2,k是一个固定的尺度参数

  特点：检测器不受平移和旋转影响

3.2 通过缩放和方向建立邻域

- 使用角点检测器获得模式元素的位置、半径和方向
- Assume a fixed scale parameter k
  Apply a corner detector to the image $\mathcal{I}$
  Initialize a list of patches
  For each corner detected
  Write (x_c, y_c) for the location of the corner
  Compute the radius r for the patch at (x_c, y_c) as
  $r(x_c, y_c) = \underset{\sigma}{\text{argmax}}\ \nabla^2_\sigma \mathcal{I}(x_c, y_c)$
  by computing $\nabla^2_\sigma \mathcal{I}(x_c, y_c)$ for a variety of values of σ,
  interpolating these values, and maximizing
  Compute an orientation histogram $H(\theta)$ for gradient orientations within a radius kr of (x_c, y_c).
  Compute the orientation of the patch θ_p as
  $\theta_p = \underset{\theta}{\text{argmax}}\ H(\theta)$. If there is more than one theta that maximizes this histogram, make one copy of the patch for each.
  Attach (x_c, y_c, r, θ_p) to the list of patches for each copy

  Algorithm 5.2: Obtaining Location, Radius and Orientation of Pattern Elements Using a Corner Detector.

- 使用高斯模糊图像的拉普拉斯变换（求二阶导）获得模式元素的位置、半径和方向

Assume a fixed scale parameter k
Find all locations and scales which are local extrema of
  $\nabla^2_\sigma \mathcal{I}(x, y)$ in location (x, y) and scale σ forming a list of triples (x_c, y_c, r)
For each such triple
  Compute an orientation histogram $H(\theta)$ for gradient orientations within
    a radius kr of (x_c, y_c).
  Compute the orientation of the patch θ_p as
    $\theta_p = \underset{\theta}{\text{argmax}}\ H(\theta)$. If there is more than one θ that
    maximizes this histogram, make one copy of the patch for each.
  Attach (x_c, y_c, r, θ_p) to the list of patches for each copy

Algorithm 5.3: Obtaining Location, Radius, and Orientation of Pattern Elements Using the Laplacian of Gaussian.

4 使用SIFT和HoG特征描述邻域

- SIFT(Scale Invariant Feature Transform)特征（是经典的尺度不变特征）

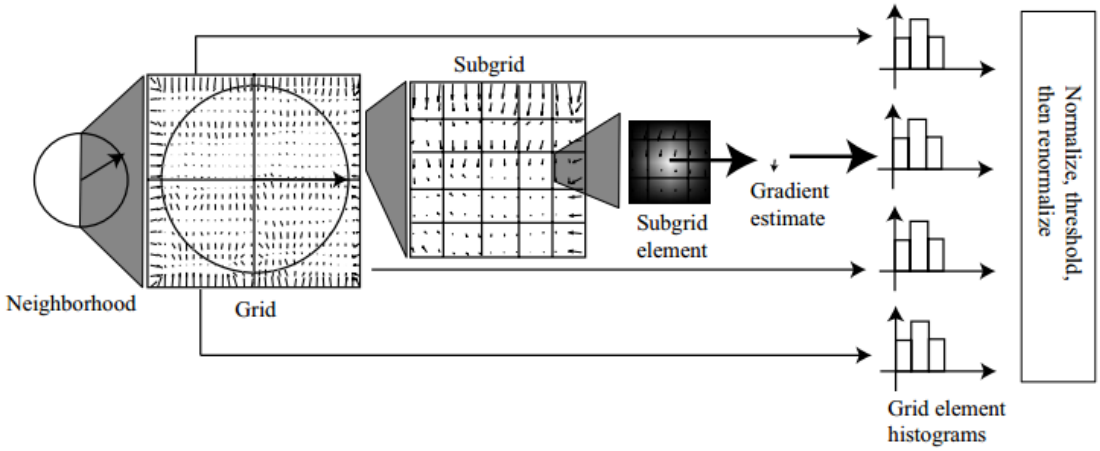


FIGURE 5.14: To construct a SIFT descriptor for a neighborhood, we place a grid over the rectified neighborhood. Each grid is divided into a subgrid, and a gradient estimate is computed at the center of each subgrid element. This gradient estimate is a weighted average of nearby gradients, with weights chosen so that gradients outside the subgrid cell contribute. The gradient estimates in each subgrid element are accumulated into an orientation histogram. Each gradient votes for its orientation, with a vote weighted by its magnitude and by its distance to the center of the neighborhood. The resulting orientation histograms are stacked to give a single feature vector. This is normalized to have unit norm; then terms in the normalized feature vector are thresholded, and the vector is normalized again.

- HOG(Histogram Of Gradient orientations)特征（是SIFT特征的一个重要变体）

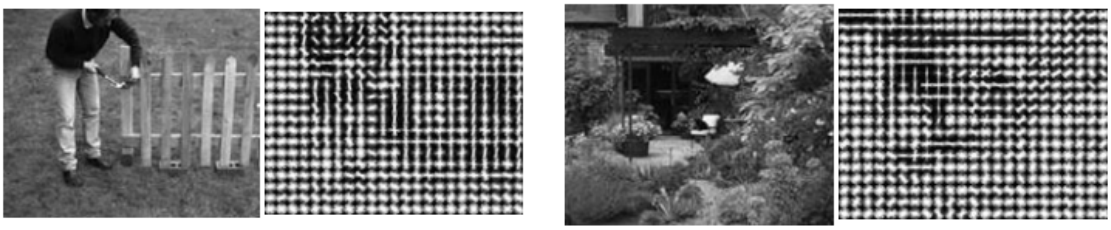


FIGURE 5.15: The HOG features for each the two images shown here have been visualized by a version of the rose diagram of Figures 5.7–5.9. Here each of the cells in which the histogram is taken is plotted with a little rose in it; the direction plotted is at right angles to the gradient, so you should visualize the overlaid line segments as edge directions. Notice that in the textured regions the edge directions are fairly uniformly distributed, but strong contours (the gardener, the fence on the **left**; the vertical edges of the french windows on the **right**) are very clear. This figure was plotted using the toolbox of Dollár and Rabaud. *Left:* © *Dorling Kindersley, used with permission. Right: Geoff Brightling* © *Dorling Kindersley, used with permission.*

纹理

1 什么是纹理

纹理是一种广泛存在，容易识别(对人而言)又很难定义的东西，纹理常常显示出重复性，是物体识别的重要线索，也是材料性质的重要线索

2 局部纹理表示

- 图像纹理通常由重复的元素(纹理基元)组成（是一组纹理基元按照某种方式重复形成的）
- 局部纹理表示可以通过使用一组不同尺度的滤波器对图像滤波然后将结果进行综合得到

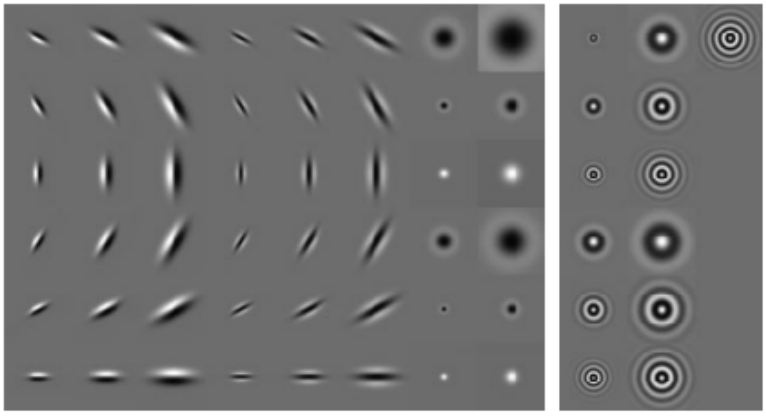


FIGURE 6.4: **Left** shows a set of 48 oriented filters used for expanding images into a series of responses for texture representation. Each filter is shown on its own scale, with zero represented by a mid-gray level, lighter values being positive, and darker values being negative. The left three columns represent edges at three scales and six orientations; the center three columns represent stripes; and the right two represent two classes of spots (with and without contrast at the boundary) at different scales. This is the set of filters used by Leung and Malik (2001). **Right** shows a set of orientation-independent filters, used by Schmid (2001), using the same representation (there are only 13 filters in this set, so there are five empty slots in the image). The orientation-independence property means that these filters look like complicated spots.

- 注：不同尺度的纹理滤波器通常是点状和条状- 局部纹理表示（纹理的滤波器响应）图示

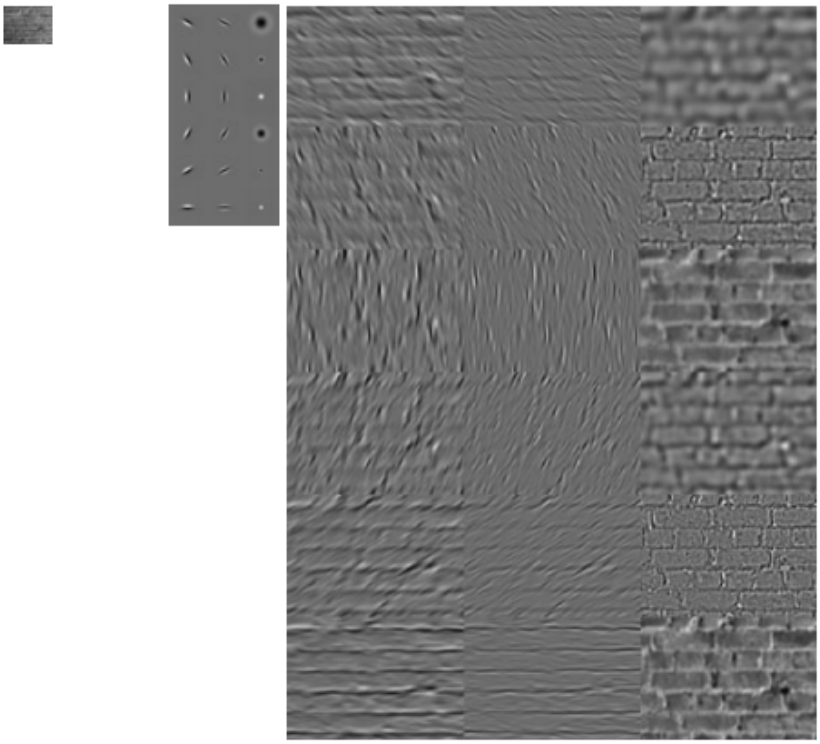


FIGURE 6.5: Filter responses for the oriented filters of Figure 6.4, applied to an image of a wall. At the **center**, we show the filters for reference (but not to scale, because they would be too small to resolve). The responses are laid out in the same way that the filters are (i.e., the response map on the top left corresponds to the filter on the top left, and so on). For reference, we show the image at the **left**. The image of the wall is small, so that the filters respond to structures that are relatively large; compare with Figure 6.6, which shows responses to a larger image of the wall, where the filters respond to smaller structures. These are filters of a fixed size, applied to a small version of the image, and so are equivalent to large-scale filters applied to the original version. Notice the strong response to the vertical and horizontal lines of mortar between the bricks, which are at about the scale of the bar filters. All response values are shown on the same intensity scale: lighter is positive, darker is negative, and mid-gray is zero.

注：在一点的纹理表示应该包含邻近滤波器输出的综合，而不仅仅是它们自身的滤波器输出

3 合并纹理表示（纹理识别）

- 向量量化和纹理：向量量化是从一个固定尺寸集中用数字表示在连续时空中的向量- 使用K均值聚类进行向量量化

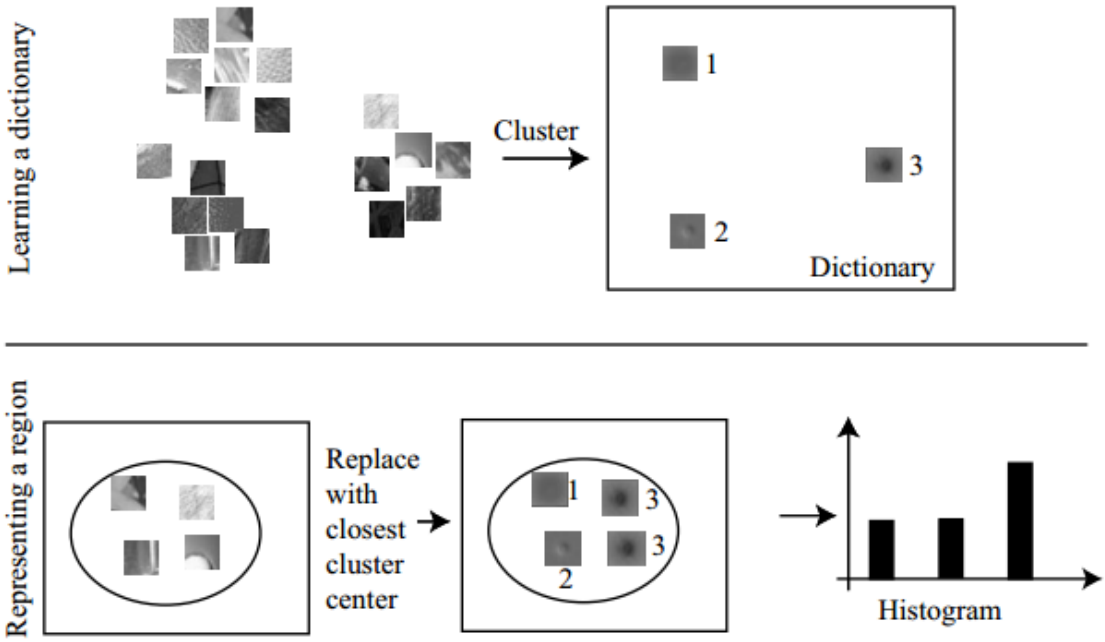
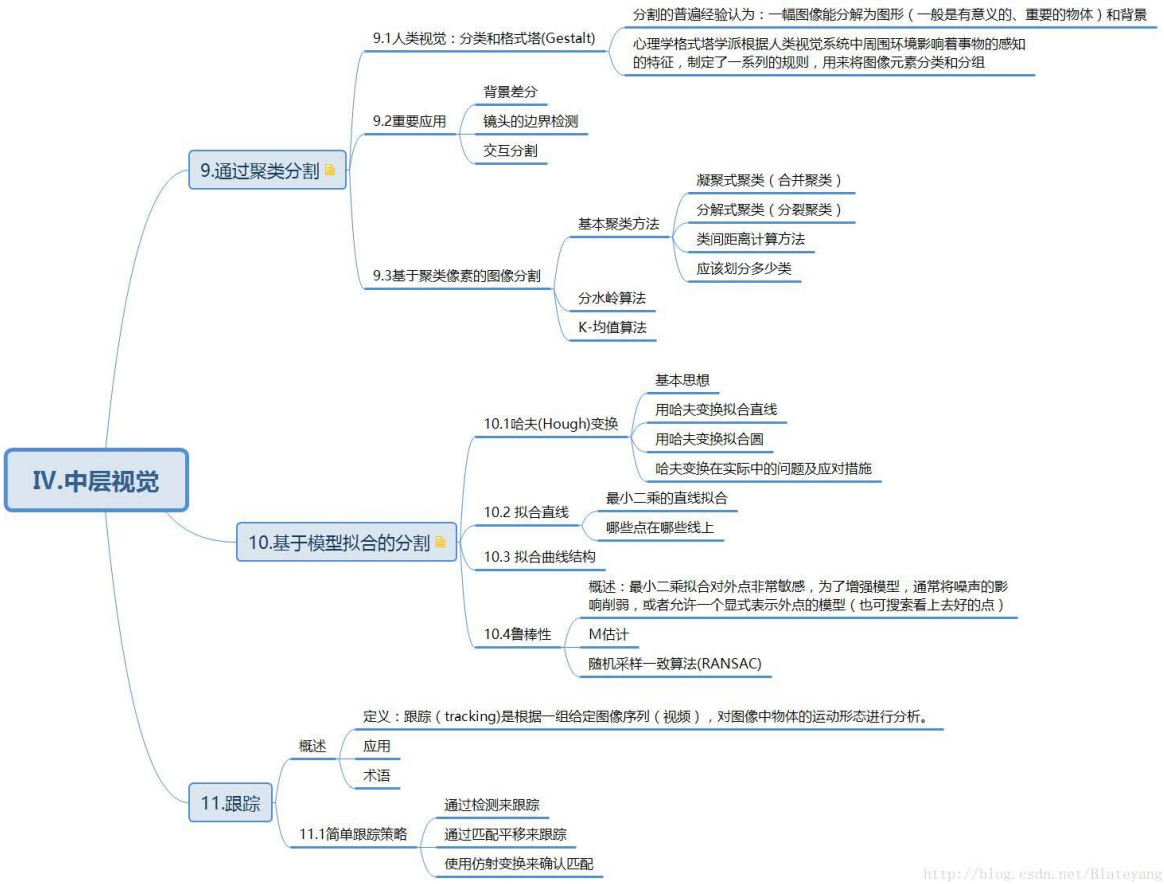


FIGURE 6.8: There are two steps to building a pooled texture representation for a texture in an image domain. First, one builds a dictionary representing the range of possible pattern elements, using a large number of texture patches. This is usually done in advance, using a training data set of some form. Second, one takes the patches inside the domain, vector quantizes them by identifying the number of the closest cluster center, then computes a histogram of the different cluster center numbers that occur within a region. This histogram might appear to contain no spatial information, but this is a misperception. Some frequent elements in the histogram are likely to be textons, but others describe common ways in which textons lie close to one another; this is a rough spatial cue. *This figure shows*

4 数据驱动的纹理表示（纹理合成）

#《计算机视觉-一种现代方法（第2版）》读书笔记四：中层视觉

本篇思维导图



注：中层视觉主要关注的是图像中的几何结构以及特定对象和目标，应用领域包括目标分割和跟踪

通过聚类分割

分割的目的是为了得到一幅图片中有用部分的一个精简的表示，其具体的理论和方法取决于应用的需求

1 人类视觉：分类和格式塔(Gestalt)

- 分割的普遍经验认为：一幅图像能分解为图形（一般是有意义的、重要的物体）和背景
- 心理学格式塔学派根据人类视觉系统中周围环境影响着事物的感知的特征，制定了一系列的规则，用来将图像元素分类和分组-
- 元素集合分组的一些规律性质

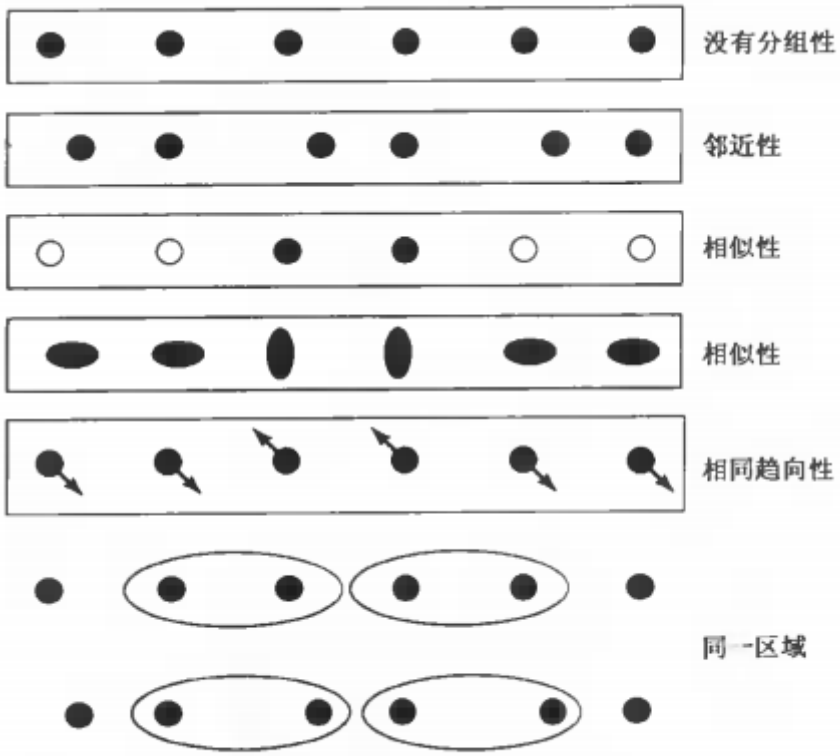


图 14.4　Gestalt 性质指导分类的例子(文中有更详细的描述)

2 重要应用

2.1 背景差分

- 在很多应用中，物体总是出现在一个相对稳定的背景中-
- 在这些应用中，通常可以通过从图像中减去背景图像的估计值，然后从结果中寻找绝对值比较大的部分来获得有用的分割-
- 背景差分算法

算法 14.1　背景差分

形成一个背景估计值 $\mathcal{B}^{(0)}$。对每一帧 $\mathcal{F}$

更新背景图片估计值，一般通过下面公式 $\mathcal{B}^{(n+1)}=\frac{w_a\mathcal{F}+\sum_i w_i\mathcal{B}^{(n-i)}}{w_c}$ 其中，选择好权重值 w_a, w_i, w_c。

从帧中减去背景图片估计值，重新记录差值大于选定阈值的每一像素点的值。

end

注：使用运动平均方法估计背景像素点的值

2.2 镜头的边界检测

- 镜头：指基本显示的是同一物体的较短视频流-
- 关键帧：一个镜头通常可以用一个关键帧来表示，这种表示可以用于视频的检索或者概况视频内容以便用户进行浏览-
- 镜头边界检测算法

算法 14.2　采用帧间差异的镜头边界检测

对于图像流中的每一帧

　计算这一帧和上一帧之间的距离

　如果距离大于某个阈值，

　　将这一帧作为一个镜头边界

end

- 计算距离的几种准则：帧差分算法、基于直方图的算法、块比较算法、边缘差分算法

2.3 交互分割

人为指定一些分离区域，计算机在此基础上完成自动分割

3 基于聚类像素的图像分割

3.1 基本聚类方法

3.1.1 凝聚式聚类（合并聚类）

```
定义每个点为独立的一个类
直到聚类达到所要求的
    将类间距离最小的两类合并
end
```

3.1.2 分解式聚类（分裂聚类）

```
定义一个包含所有点的类
直到聚类达到所要求的
    将一个类分裂成两个类，条件时所产生的两个类的类间距离最大
end
```

3.1.3 类间距离计算方法

- 单连接聚类：选择两类之间最近的两个元素之间的距离作为类间距离-
- 全连接聚类：选择两类之间最远的两个元素之间的距离作为类间距离-
- 基于集团均值的聚类：选择聚类中元素间距离的平均值

3.1.4 应划分多少类

通过树状图（一种显示类间距离的层次结构表示），让用户据其做出一个适当的聚类选择

3.2 分水岭算法

分水岭算法可以较好地得到超像素（具有相似颜色或纹理的图像块）

3.3 K-均值算法

3.3.1 基本步骤

1. 假设聚类中心已知，并且分配每个点到最近的聚类中心
2. 假设分配已确定，选择一个新的聚类中心（每个中心是分布在这个类中各个点的平均值）集
   随机的选择聚类中心作为起始点，并轮流迭代执行这些步骤，直至收敛于目标函数的局部最小值。

3.3.2 实现细节

具体细节会略有差异，比如初始化聚类中心后，将其它点一次性就近分配到各聚类，然后重新计算聚类中心再不断调整样本点分配；也可一次将一个样本点依顺序归入就近的聚类，并立即重新计算该类的聚类中心，然后再计算下一个样本的归类，直到所有的样本都归到相应的类中。

基于模型拟合的分割

与基于聚类的分割不同的是，基于模型拟合的分割中模型是已知的，而且是从更大尺度的关系看问题，而不仅仅是样本到样本的关系。通常选择一个模型，然后确定一个拟合好坏的准则，来审视一组样本是否具有拟合该模型的属性。

1 哈夫（Hough）变换

1.1 基本思想

通过记录所有样本点能具有的所有结构，然后看看哪一个结构有最多的投票来把在同样结构上的点聚类

1.2 用哈夫变换拟合直线

哈夫变换最成功的应用是在直线检测上，利用点-线对偶性原理寻找参数空间网格中投票最多的网格来确定对应的直线

1.3 用哈夫变换拟合圆

原理与拟合直线大体上类似，只是点对应的二维极径极角空间被三维的圆心点x, y还有半径r空间取代，实际实现中是通过哈夫梯度法求解。

1.4 哈夫变换在实际中的问题及应对措施

问题：

- 量化误差（合适的网格尺寸很难选择）-
- 受噪声影响大

应对措施：

- 减少不相关样本（包括去噪）-
- 小心地选择网格（试算法）

2 拟合直线

- 最小二乘直线拟合-
- 增量直线拟合：通过沿着曲线走，对曲线上的点拟合直线，当残差足够大时截断曲线-
- k-均值直线拟合：通过把点分配到最近的直线然后重新拟合

3 拟合曲线结构

4 鲁棒性

最小二乘拟合对外点非常敏感，为了增强模型，通常将噪声的影响削弱，或者允许一个显式表示外点的模型（也可搜索看上去好的点）

4.1 M估计

- 最好的估计是在接近参数模型的最坏分布下还可以表现得非常好的估计。-
- 一个M估计可以认为是一种策略，其可以保证外点作用被衰减的概率要比它们产生平方误差的概率要高
-
- 一个M估计通过最小化后面的表达式来估计参数 Σiρ(ri(xi,θ);σ),ri(xi,θ)表示残留误差

4.2 随机采样一致算法(RANSAC)

基于在数据点中搜索好的点的思想，具体搜索一个随机采样，对其进行拟合，并判断其是否得到许多数据点支持

算法 15.4 RANSAC:用随机采样一致拟合直线

确定:

n——所需要的最少点数

k——需要的迭代次数

t——用来判断一个点是否拟合得很好的阈值

d——判断一个点是否拟合得很好所需要的邻近点数目

直到 k 次迭代完成

从数据中均匀随机的采样 n 个点

对这 n 个点进行拟合

对于在采样外的每一个点

用 t 比较点到直线的距离;如果距离小于 t,那么点是很靠近的

end

如果有 d 或更多个点靠近直线,那么是一个好的拟合。重新用这些点拟合直线

end

使用拟合误差作为准则,挑出最好的拟合

跟踪

1 概述

1.1 定义

跟踪（tracking)是根据一组给定图像序列（视频），对图像中物体的运动形态进行分析。

1.2 应用

- 运动捕捉-
- 从运动中识别-
- 监视-
- 定位

1.3 术语

- 状态（跟踪的基本假设为当前状态仅依赖于前一状态）-
- 观测：是对运动物体状态的测量（当前观测仅仅依赖当前状态）
  跟踪包括利用观测去推测状态，状态和观测的基本假设意味着跟踪问题的推理结构是个隐马尔可夫模型。

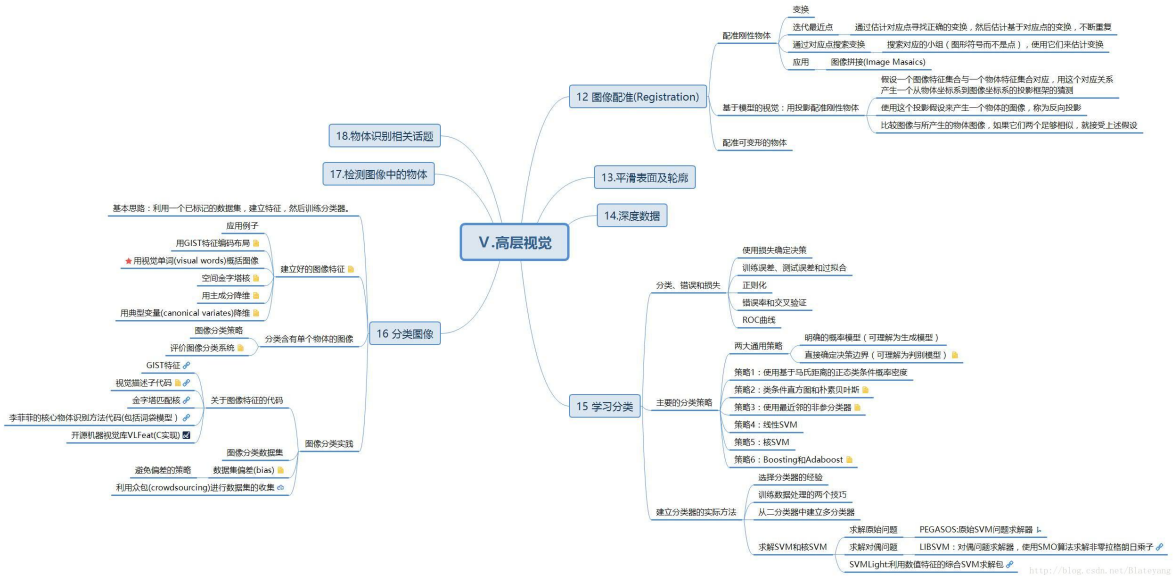
2 简单跟踪策略

- **通过检测来跟踪**：当视频中只有一个简单对象时，可以通过报告检测器在视频每一帧中响应的位置来跟踪对象；有多个物体或复杂情况可以采取轨迹跟踪-

- **通过匹配平移来跟踪**：在相邻帧间可以把对象看作是缓慢运动的box,通过在帧间邻近位置搜索最佳匹配的平移后的box来跟踪对象- 利用仿射变换来确认匹配：如果观测时间较长，对象有可能发生形变（如三维旋转），需要修改为基于仿射变换的匹配

#《计算机视觉-一种现代方法（第2版）》读书笔记五：高层视觉

本篇思维导图



图像配准(Registration)

1.配准刚性物体

- 变换：旋转(rotation)、平移(translation)、缩放(scale)

∑i[(sR(θ)xi+t)−yc(i)]2

迭代最近点：通过估计对应点寻找正确的变换，然后估计基于对应点的变换，不断重复

通过对应点搜索变换：搜索对应的局部区域（图形符号而不是点），使用它们来估计变换

应用：图像拼接(Image Masaics)

2.基于模型的视觉：用投影配准刚性物体

1. 假设一个图像特征集合与一个物体特征集合对应，用这个对应关系产生一个从物体坐标系到图像坐标系的投影框架的猜测
2. 使用这个投影假设来产生一个物体的图像，称为反向投影
3. 比较图像与所产生的物体图像，如果它们两个足够相似，就接受上述假设

3.配准可变形的物体

学习分类

1.分类、错误和损失

- 使用损失确定决策-
- 训练误差、测试误差和过拟合-
- 正则化-
- 错误率和交叉验证-
- ROC曲线

2.主要的分类策略

2.1 两大通用策略

- 明确的概率模型(可理解为生成模型，如朴素贝叶斯模型）-
- 直接确定决策边界（可理解为判别模型）

2.2 策略1：使用基于马氏距离的正态类条件概率密度

Assume we have N classes, and the kth class contains N_k examples, of which the ith is written as $\boldsymbol{x}_{k,i}$.
For each class k, estimate the prior, the mean and standard deviation for that class-conditional density.

$$p(k) = \frac{N_k}{\sum_i N_i}$$

$$\boldsymbol{\mu}_k = \frac{1}{N_k}\sum_{i=1}^{N_k} \boldsymbol{x}_{k,i};$$

$$\Sigma_k = \frac{1}{N_k - 1}\sum_{i=1}^{N_k} (\boldsymbol{x}_{k,i} - \boldsymbol{\mu}_k)(\boldsymbol{x}_{k,i} - \boldsymbol{\mu}_k)^T;$$

To classify an example $\boldsymbol{x}$,

Choose the class k with the smallest value of $\delta(\mathbf{x}; \boldsymbol{\mu}_k, \Sigma_k)^2 - p(k)$

where

$$\delta(\mathbf{x}; \boldsymbol{\mu}_k, \Sigma_k) = \frac{1}{2}\left((\mathbf{x} - \boldsymbol{\mu}_k)^T \Sigma_k^{-1} (\mathbf{x} - \boldsymbol{\mu}_k)\right)^{(1/2)}.$$

注：马氏距离表示数据的协方差距离，它是一种有效计算两个未知样本集间相似度的方法。与欧式距离不同的是它考虑到各种特征之间的联系，并且是尺度无关的。

该算法的几何解释：在考虑方差情况下将数据项划分到距类均值最近的类中（特别地，沿某一方向方差小的距类均值的距离有大权重，方差大的距类均值的距离权重小）

评价：在有很多训练数据和类别的低维问题可以尝试。马氏距离的适用场景相对较少，因为当特征向量是高维时求协方差矩阵比较困难。

2.3 策略2：类条件直方图和朴素贝叶斯

如果有足够的标记数据，就可以对类条件密度直方图建模，这在低维情况下是有用的。

- 利用贝叶斯公式直接算出后验概率然后进行比较

2.4 策略3：使用最近邻的非参数分类器

对一个类别未知的样本，可以假设其类别是在特征空间中距离这个样本最近的训练样本的类别，或找出距离待识别样本最近的几个，然后用这几个训练样本的类别进行投票来确定待识别样本的类别

算法 22.3 A(k,l)近邻分类器根据特征空间中距离待识别样本最近的训练样本所属的类别对待识别样本进行分类

对一个特征向量 $\boldsymbol{x}$

1. 寻找特征空间中距离 $\boldsymbol{x}$ 最近的 k 个训练样本 $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_k$；
2. 从 $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_k$ 对应的 $y_1, \cdots, y_k$ 中找出数量最多的一个类别 c，并设这个数量为 n；
3. 如果 $n > l$，则把 $\boldsymbol{x}$ 归为类别 c，否则拒识这个样本。

评价：这一策略总是有用的，当训练数据很多时与其他分类方法相比也保持有竞争力。

2.5 策略4：线性SVM

线性可分情况

$$\begin{aligned} \text{minimize} \quad & (1/2)\boldsymbol{w}\cdot\boldsymbol{w} \\ \text{subject to} \quad & y_i(\boldsymbol{w}\cdot\boldsymbol{x}_i+b)\geq 1, \end{aligned}$$

线性不可分情况

$$\begin{aligned} \text{minimize} \quad & \tfrac{1}{2}\boldsymbol{w}\cdot\boldsymbol{w}+C\textstyle\sum_{i=1}^{N}\xi_i \\ \text{subject to} \quad & y_i(\boldsymbol{w}\cdot\boldsymbol{x}_1+b)\geq 1-\xi_i \\ \text{and} \quad & \xi_i\geq 0. \end{aligned}$$

2.6 策略5：核SVM

$$\begin{aligned} \text{maximize} \quad & \textstyle\sum_i^N \alpha_i - \frac{1}{2}\sum_{i,j=1}^{N}\alpha_i(y_i y_j k(\boldsymbol{x}_i,\boldsymbol{x}_j))\alpha_j \\ \text{subject to} \quad & \alpha_i \geq 0 \\ \text{and} \quad & \textstyle\sum_{i=1}^{N}\alpha_i y_i = 0, \end{aligned}$$

2.7 策略6：Boosting和Adaboost

集合多个弱分类器构造强分类器

3.建立分类器的实际方法

3.1 选择分类器的经验

经验建议对大多数问题可以首先用线性SVM去尝试，如果效果不理想，接下来换用核SVM或boosting方法

3.2 训练数据处理的两个技巧

1. 数据增广：缩放、裁剪、旋转、翻转等
2.
3. bootstrapping(自助法)
    1. 基本思想：将被分类错误的正负样本插入到训练集中重新训练分类器，反复迭代
    2. 变体：hard negative mining 从负样本中选取出一些有代表性的负样本(分类器检测出的错误的正样本，被称作hard negative)，不断重新训练，使得分类器的训练结果更好

3.3 从二分类器中建立多分类器

- all-vs-all方法：为每类都建立一个分类器-
- one-vs-all方法：为每类和其余类建立一个分类器（比all-vs-all方法往往要更可靠和有效一些）

3.4 求解SVM和核SVM

- PEGASOS:原始SVM问题求解器
- LIBSVM：对偶问题求解器，使用SMO算法求解非零拉格朗日乘子
- SVMLight:利用数值特征的综合SVM求解包

分类图像

基本思路：利用一个已标记的数据集，建立特征，然后训练分类器

1. 建立好的图像特征

不同的特征构建适用于不同的情况。关键是建立那些能暴露类间变化并且抑制类内变化的特征。 任何一种特征表示形式都应该是对图像的旋转、平移或缩放鲁棒的，因为这些变换并不会影响图像的含义

1.1 应用例子

- 检测特定图像（如色情图片识别）
- 材料分类
- 场景分类

1.2 用GIST特征编码布局

GIST特征综合了一幅图像不同部分的梯度信息（尺度和方向），提供了关于一个场景的粗略描述

- 对于场景分类一个自然的线索就是图片的整体布局，GIST特征企图捕获的正是这种布局。
-
- GIST特征的计算过程

  1 用32个Gabor滤波器（4个尺度，8个方向）卷积图像，产生32幅特征图-

  2 把每个特征图等分成16个区域（4*4网格），求每个区域的特征值均值- *

  3 *连接所有32幅特征图的特征均值，形成一个16*32=512维的GIST特征

1.3 用视觉单词(visual words)概况图像

- 记录具有特点的局部图像块，用某些局部特征(如SIFT特征)描述这些局部邻域并进行向量量化得到视觉单词，然后通过直方图的形式对视觉单词集进行统计概括，如果在一幅图像中大多数单词与另一幅图像中的大多数单词匹配，它们的视觉单词直方图就会是相似的。
-
- 衡量直方图的相似性，普遍采用的是直方图的交距离：

$$K(h,g)=\sum imin(hi,gi)$$

1.4 空间金字塔核

是视觉单词直方图方法的一个重要变体，能产生可有效粗略编码空间布局的核

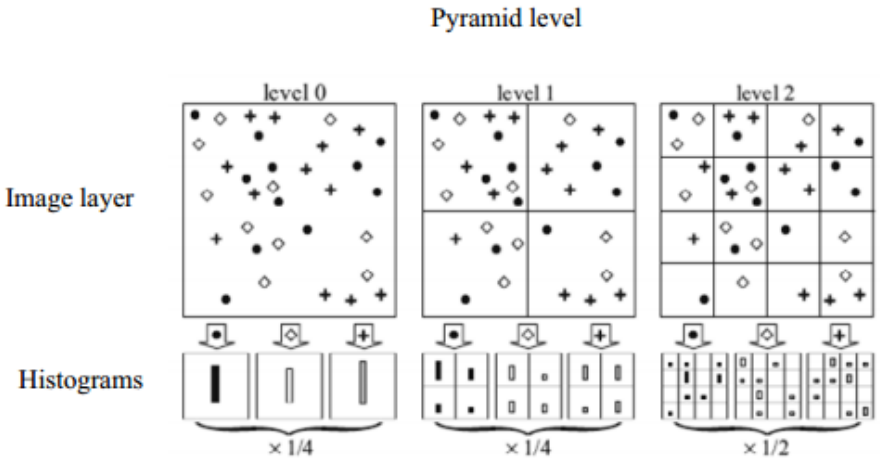


FIGURE 16.8: A simplified example of constructing a spatial pyramid kernel, with three levels. There are three feature types, too (circles, diamonds, and crosses). The image is subdivided into one-, four-, and sixteen-grid boxes. For each level, we compute a histogram of how many features occur in each box for each feature type. We then compare two images by constructing an approximate score of the matches from these histograms. *This figure was originally published as Figure 1 of "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," by S. Lazebnik, C. Schmid, and J. Ponce, Proc. IEEE CVPR 2006, © IEEE 2006.*

- 模型

$$\sum_{t \in \text{feature types}} \sum_{l \in \text{levels}} \sum_{i \in \text{grid boxes}} w_l \min(H^l_{\mathcal{I},t}(i), H^l_{\mathcal{I},t}(i)).$$

  - wl表示相匹配的grid权重，越精细的grid权重越大
  - 应用
    空间金字塔核在场景图像分类上做得很好，在标准图像分类任务上也要优于直方图交核；它能够很好地表示相对独立的物体或自然场景，但对于缺少纹理的物体或与背景相似的物体会遇到麻烦

1.5 用主成分(PCA)降维

Assume we have a set of n feature vectors $\boldsymbol{x}_i$ $(i = 1, \ldots, n)$ in $\mathbb{R}^d$. Write

$$\boldsymbol{\mu} = \frac{1}{n} \sum_i \boldsymbol{x}_i$$

$$\Sigma = \frac{1}{n-1} \sum_i (\boldsymbol{x}_i - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})^T$$

The unit eigenvectors of Σ—which we write as $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_d$, where the order is given by the size of the eigenvalue and $\boldsymbol{v}_1$ has the largest eigenvalue—give a set of features with the following properties:

- They are independent.
- Projection onto the basis $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$ gives the k-dimensional set of linear features that preserves the most variance.

Algorithm 16.1: Principal Components Analysis

PCA建立了在特定维数下最能表达原高维数据变化的新的特征集，但是并不能保证这个特征集能帮助我们实现有效分类。

1.6 用典型变量(canonical variates)降维

典型变量指能够明显地反映出不同类别样本间差异的线性特征，这些特征能使类间尽可能分开

Assume that we have a set of data items of g different classes. There are n_k items in each class, and a data item from the kth class is $\boldsymbol{x}_{k,i}$, for $i \in \{1, \ldots, n_k\}$. The jth class has mean $\boldsymbol{\mu}_j$. We assume that there are p features (i.e., that the $\boldsymbol{x}_i$ are p-dimensional vectors).
Write $\overline{\boldsymbol{\mu}}$ for the mean of the class means, that is,

$$\overline{\boldsymbol{\mu}} = \frac{1}{g}\sum_{j=1}^{g} \boldsymbol{\mu}_j,$$

Write

$$\mathcal{B} = \frac{1}{g-1}\sum_{j=1}^{g} (\boldsymbol{\mu}_j - \overline{\boldsymbol{\mu}})(\boldsymbol{\mu}_j - \overline{\boldsymbol{\mu}})^T.$$

Assume that each class has the same covariance Σ, which is either known or estimated as

$$\Sigma = \frac{1}{N-1}\sum_{c=1}^{g} \left\{ \sum_{i=1}^{n_c} (\boldsymbol{x}_{c,i} - \boldsymbol{\mu}_c)(\boldsymbol{x}_{c,i} - \boldsymbol{\mu}_c)^T \right\}.$$

The unit eigenvectors of $\Sigma^{-1}\mathcal{B}$, which we write as $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_d$, where the order is given by the size of the eigenvalue and $\boldsymbol{v}_1$ has the largest eigenvalue, give a set of features with the following property:

- Projection onto the basis $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$ gives the k-dimensional set of linear features that best separates the class means.

Algorithm 16.2: Canonical Variates

2. 分类含有单个物体的图像

2.1 图像分类策略

- 通用策略是计算特征，利用特征向量构建多分类器
- 典型方法
  - 使用HOG和SIFT特征的变体，结合颜色特征
  - 视觉单词字典（计算图像的视觉单词，进行向量量化，用视觉单词的直方图表示图像然后使用直方图交方法对其分类）
  - 空间金字塔核金字塔匹配核

3. 图像分类实践

3.1 关于图像特征的代码

- GIST特征
- 视觉描述子代码（用于计算基于多颜色SIFT特征的视觉单词）
- 金字塔匹配核
- 李飞飞的核心物体识别方法代码（包括词袋模型）
- 开源计算机视觉库VLFeat(C实现)
-

3.2 图像分类数据集

- Caltech系列数据集（Caltech-101、Caltech-256等）
- LabelMe数据集（是一个图像标注环境，可用来标注该数据集图像中的物体）
- ImageNet数据集（某种程度上成就了深度学习）
- SUN数据集（关于场景的最大数据集）

- 其他

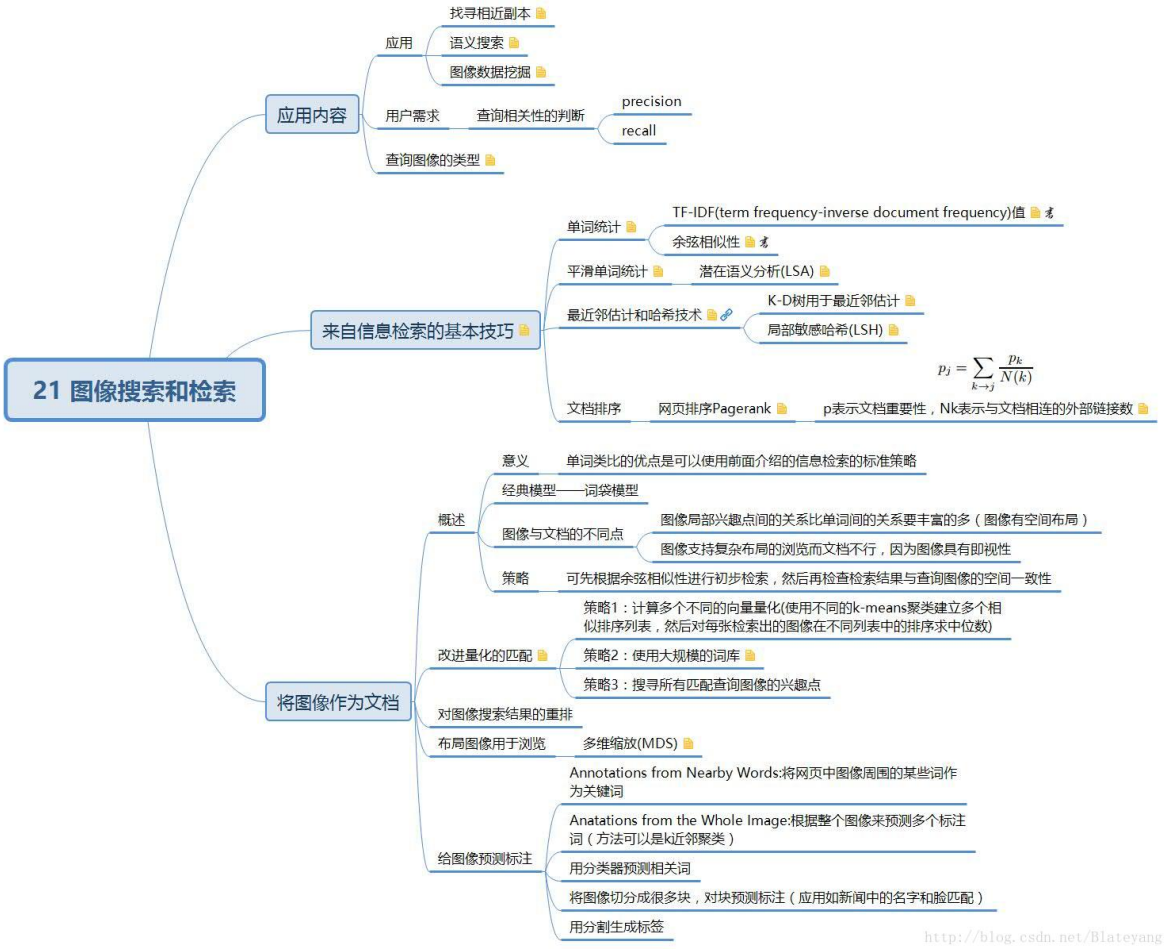
3.3 数据集偏差(bias)

指数据集的性质与真实世界的性质存在表示偏差

- 避免偏差的策略：
  - 从多种不同的途径收集大量数据
  - 在使用数据集评估复杂方法前使用基本方法小心地评估数据集
  - 采取与收集训练数据不同的策略来收集数据，并通过评估它们来量化偏差的影响

读书笔记六：应用之图像搜索和检索

本篇思维导图



1 应用内容

1.1 应用领域

- 找寻相近副本：相似图片搜索，可用在电子商务以及商标、摄影作品等版权保护中-
- 语义搜索：搜寻某种语义的图片-
- 图像数据挖掘：如对卫星图像进行数据挖掘以回答城市扩张程度、还剩下多少热带雨林等

1.2 用户需求

查询相关性的判断：准确率(precision)、召回率(recall)

1.3 查询图像的类型

有多种划分方式

一种划分

- 独特的物体：如埃菲尔铁塔
- 带限定的独特物体：如1950年的埃菲尔铁塔
- 非独特的物体：类别，如狗
- 带限定的非独特物体：修饰的类别，如趴在地上吐舌头的狗

另一种划分

- Specific:类的一个实例，如某个邻居家的猫- general:类的一个个体，如一只猫- subjective:情感或抽象的内容，如猫

2 来自信息检索的基本技巧

2.1 单词统计

基于传统向量空间模型进行相似匹配，其中有两个比较常见的术语，TD-IDF值和余弦相似性。

- TF-IDF(term frequency-inverse document frequency)值：词频-逆文档频率值，信息检索中常见的术语，详见-
- 余弦相似性：通过两向量的夹角大小来判断向量的相似程度，夹角越小，余弦值越大，向量越相似。详见

2.2 平滑单词统计

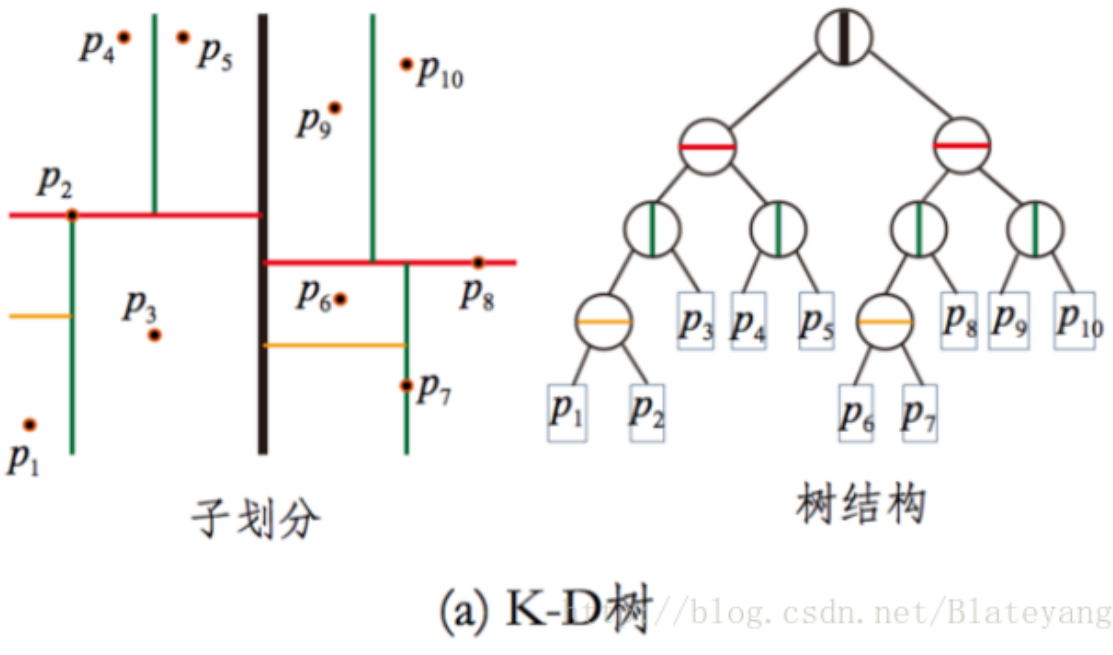
单词的多义性和同义性（不同单词表示相同意思）对基于传统单词统计的信息检索有很大阻碍，需要去除单词间的相关性，一种常用的技术是潜在语义分析(Latent Semantic Analysis,LSA)。

LSA(LSI)使用SVD来对单词-文档矩阵进行分解。SVD可以看作是从单词-文档矩阵中发现不相关的索引变量(因子)，将原来的数据映射到语义空间内。在单词-文档矩阵中不相似的两个文档，可能在语义空间内比较相似。

2.3 最近邻估计和哈希技术

2.3.1 K-D树用于最近邻估计

将特征向量以树结构的方法组织起来，使得在检索的时候其计算复杂度降到关于样本数目n的对数的复杂度，具体算法可看《统计学习方法》K-近邻一章，适合于低维的情况，特征维数过高也不太适用，一种改进是multiple randomized k-d trees。

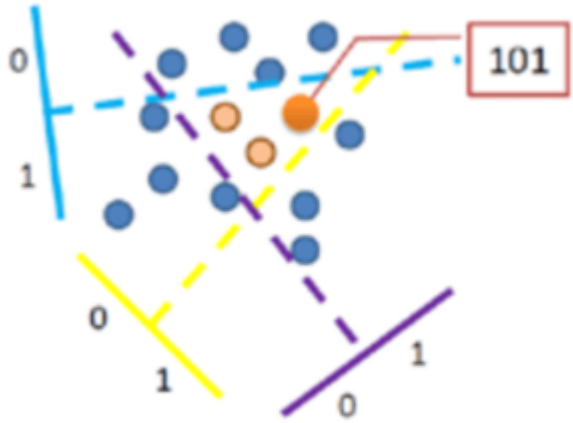


(a) K-D树 //blog.csdn.net/Blateyang

2.3.2 哈希技术之局部敏感哈希(LSH)

基于哈希的检索方法其关键之处在于设计一个有效的哈希函数集，使得原空间中的数据经过该哈希函数集映射后，在汉明空间其数据间的相似性能够得到较好的保持或增强。

基于哈希的检索过程包括：特征提取、哈希编码、汉明距离排序和重排(重排的目的是在相似样本集中进行精细检索)



(b) LSH

上图为一种局部敏感哈希方法的示意图，它用随机生成的超平面去分割数据集，并为它们编码， 最后将编码结果串接在一起形成哈希码。原空间中相似的数据在很大概率上会被编成相同的哈希码（落入同一个“桶”中）。

2.4 文档排序

网页排序Pagerank

谷歌搜索早期的核心算法，由Larry Page发明,它的主要思想是网页文档间含有大量的有向连接，重要的文档往往有很多链接指向它们.

$pj = \sum k- > jpk/N(k)$

p表示文档重要性，Nk表示与文档相连的外部链接数

3 将图像作为文档

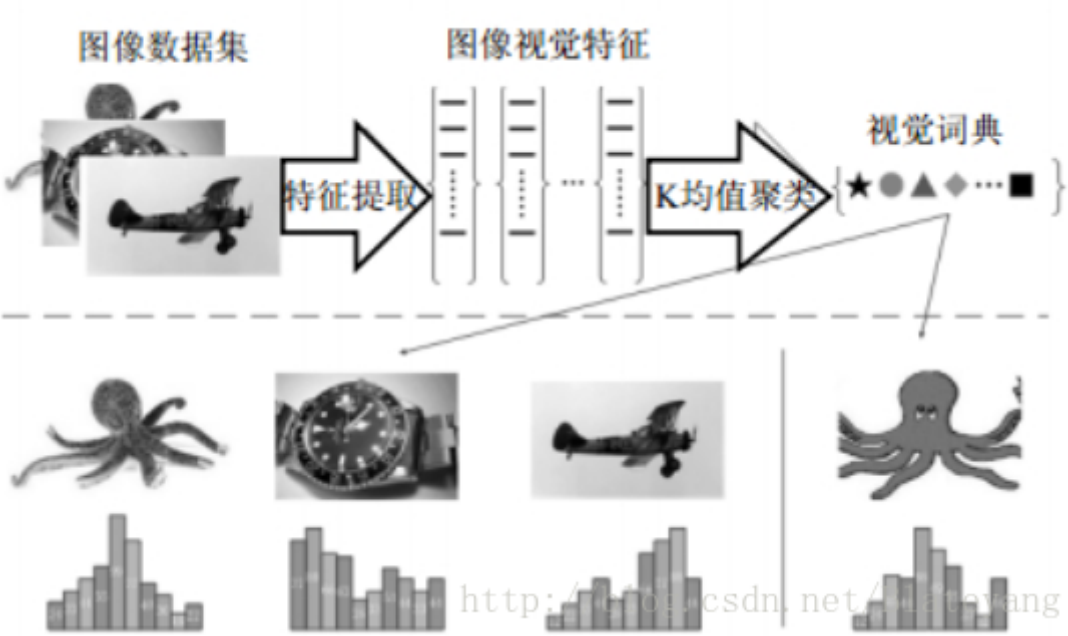
3.1 概述

3.1.1 意义

将图像类比成文档，将图像局部类比成单词的优点是可以使用前面介绍的信息检索的标准策略。

3.1.2 经典模型——词袋模型

词袋模型的基本思想是提取图像的局部特征描述子（如SIFT），然后将其聚类量化成视觉词(Visual words)，并对整幅图像的视觉词进行词频统计，乘上TD-IDF值组成加权词频向量，作为最终用于相似性匹配的特征向量，最后按计算出的相似性大小排序返回检索的结果。



词袋模型示意图

3.1.3 图像与文档的不同点

- 图像局部兴趣点间的关系比单词间的关系要丰富的多（图像有空间布局）-
- 图像支持复杂布局的浏览而文档不行，因为图像具有即视性

3.1.4 策略

可先根据余弦相似性进行初步检索，然后再检查检索结果与查询图像的空间一致性

3.2 改进量化的匹配

由于视觉单词中的向量量化会抑制一些可能重要的细节，可以尝试采用以下策略进行改进量化的匹配。
策略1：计算多个不同的向量量化(使用不同的k-means聚类建立多个相似排序列表，然后对每张检索出的图像在不同列表中的排序求中位数)
策略2：使用大规模的词库（利用层级k-means（树形结构）等方法提高搜索效率）
策略3：搜索所有匹配查询图像的兴趣点

3.3 对图像搜索结果重排

可根据空间一致性采用RANSAC（随机采样一致）等算法进行重排

3.4 布局图像用于浏览

可采用多维缩放(MDS)等技术（类似谷歌地球地图放大的效果）

3.5 给图像预测标注

- Annotations from Nearby Words:将网页中图像周围的某些词作为关键词
- Anatations from the Whole Image:根据整个图像来预测多个标注词（方法可以是k近邻聚类）
- 用分类器预测相关词
- 将图像切分成很多块，对块预测标注（应用如新闻中的名字和脸匹配）
- 用分割生成标签