

第 5 章 尺度不变特征

Harris 角点不具有尺度不变性,但在许多情况下,需要在尺寸不同但内容相似的图像上进行特征点的检测与匹配。因此,本章首先讲述尺度不变的关键点检测理论与技术,即 Harris-Laplace 检测器,以此为基础,进一步介绍经典的尺度不变特征转换(Scale Invariant Feature Transform, SIFT)和结合定向 FAST 角点检测与旋转 BRIEF 描述子的特征(Oriented FAST and Rotated BRIEF, ORB)特征。

5.1 尺度不变理论基础

5.1.1 尺度不变思路

Harris 检测器使用固定大小的窗口来检测特征点。如果 Harris 检测器能够根据图像的尺度适当调整窗口大小,以确保窗口内的图像内容保持一致,就可以检测出相同的特征点。

如图 5-1 所示,(a)(b)两图中的两条曲线具有相同的形状和不同的尺度。(b)图曲线上的圆形窗口能够捕获整个角部分,而同样大小的窗口在(a)图曲线上只能获得一段弧线,因此只有选择一个更大的圆形窗口才能获取相同的图像信息。那么,如何独立地为每幅图像找到正确缩放的窗口,以便在不同尺度下的同一位置获得相同的图像内容呢?下面以图 5-2 为例说明。

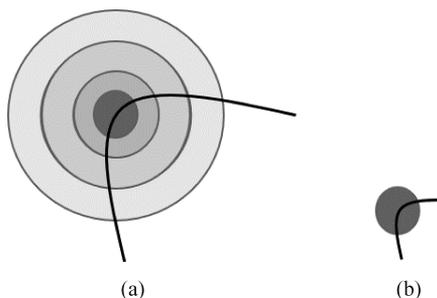


图 5-1 检测特征点时尺度变化对窗口选择的影响示意图

图 5-2 中(a)图与(b)图为不同尺度下的同一幅图像,分别记为 F_1 与 F_2 。二者的关系可利用尺度变换函数 $T_s(\cdot)$ 表示为

$$F_2 = T_s(F_1, \gamma) \quad (5-1)$$

或



图 5-2 不同尺度下的同一幅图像

$$F_1 = T_s(F_2, 1/\gamma) \quad (5-2)$$

其中, $\gamma > 0$ 表示图像 F_2 相对于图像 F_1 的缩放倍数。在图像 F_1 上确定一个关键点 P , 以 P 为圆心, 选择半径为 r_1 的圆形区域 R_1 。同样地, 在图像 F_2 上选取同一关键点, 以 P 为圆心, 选择半径为 r_2 的圆形区域 R_2 。需注意, 关键点 P 在图像 F_1 和图像 F_2 中的坐标位置不同。我们的目标是, 当给定区域 R_1 时, 确定区域 R_2 , 使得区域 R_2 包含的图像内容与区域 R_1 一致。

显然, 如果两个圆形区域半径满足 $r_2 = \gamma r_1$, 那么区域尺寸(对于圆形区域可用其半径表达)与图像尺度是协变的。在这种情况下, R_1 和 R_2 两个区域包含的图像内容必然一致, 因此称 R_1 和 R_2 为对应区域。然而, 在确定对应区域时, 图像间的尺度关系是未知的。选择区域 R_2 时, 只依据图像 F_2 , 而无法获取尺度比例 γ 、半径 r_1 和图像 F_1 的信息。因此, 确定 R_1 和 R_2 , 只能根据它们所包含的内容来进行, 也就是说, R_1 和 R_2 是相互独立确定的。为了应对这个问题, 一种简单的想法是构造一个函数 $C(\cdot)$, 将图像内容映射为一个响应量, 该响应量随着区域尺寸的变化而变化。通过分析响应量变化曲线来确定区域尺寸。对于区域 R_1 , 响应量 c_1 表示为

$$c_1 = C(F_1, P, r_1) \quad (5-3)$$

类似地, 对于区域 R_2 , 响应量 c_2 表示为

$$c_2 = C(F_2, P, r_2) \quad (5-4)$$

然后, 若已知 c_1 , 比较 c_1 与 c_2 , 当 $c_2 = c_1$ 时, 说明 R_2 和 R_1 是对应区域。

现在的问题是如何构造函数 $C(\cdot)$, 为此, 我们来讨论函数 $C(\cdot)$ 应满足的属性。根据上面的方案描述, 在不同尺度上具有相同图像内容的情况下, 函数 $C(\cdot)$ 的输出不受影响。由此, 函数 $C(\cdot)$ 必须具有尺度不变性, 也就是给定的 F_1 、 P 和 r_1 , 对于任意的 $\gamma > 0$, $C(T_s(F_1, \gamma), P, \gamma r_1) = C(F_1, P, r_1)$ 。这意味着, 如果将图像 F_1 缩放 γ 倍得到 F_2 , 那么 c_2 随着 r_2 的变化曲线应该是 c_1 随着 r_1 的变化曲线在横坐标方向上按比例压缩(或拉伸)的结果, 压缩比例也为 γ 。图 5-3 展示了 $\gamma = 1/2$ 时所期望的两种内容响应曲线的情况。因此, 内容响应量随着区域尺寸的变化曲线应该与图像尺度是协变的^[7]。



图 5-3 与图像尺度协变的内容响应曲线示例



为确定 r_2 , 需以 c_1 作为参考值, 即满足方程 $C(F_2, P, r_2) = c_1$ 。选择 c_1 , 应使未知数 r_2 必须有唯一解。其次, c_1 应具有可区分性。综合这两个因素, c_1 的最佳选择 c_1^* 就是曲线 $c_1 \sim r_1$ 的极值点, 即

$$c_1^* = \max_{r_1} C(F_1, P, r_1) \quad (5-5)$$

因此, 这也要求函数 $C(\cdot)$ 随着区域尺寸的变化只有一个明显的(尖锐的)峰值, 而如果峰值不明显或存在多个相似峰值, 响应函数 $C(\cdot)$ 就不符合要求。如图 5-4 所示, 其中(c)图响应曲线只有一个极大值点, 是比较符合要求的情况。

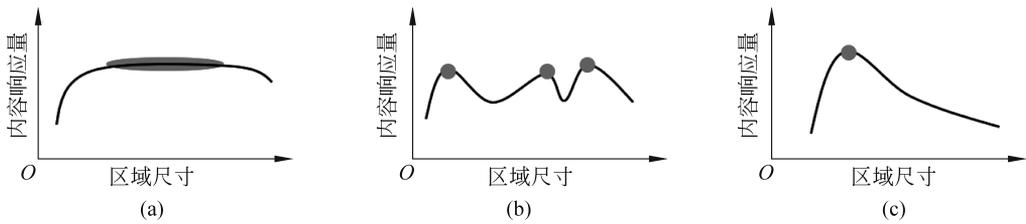


图 5-4 多种尺度不变区域的响应函数示例

5.1.2 高斯拉普拉斯算子

为了更进一步说明如何设计响应函数, 考虑一个重要的计算机视觉任务: 斑点(Blob)检测。Blob 是指近似圆形的区域, 其内部像素值相对均匀, 但与周围像素有明显的颜色和灰度差异。在图像中, 斑点广泛存在, 例如, 一棵树可以看作是一个斑点, 一片草地也是, 甚至一栋房子都可以被视为一个斑点。因此, 斑点检测是许多图像处理和识别任务的重要预处理步骤, 而一些斑点检测方法也可以应用于尺度不变特征检测中。

由于斑点区域边界的像素值会出现明显变化, 因此这些边界也可以看作一种边缘。回顾第 2 章介绍的边缘检测方法, 可以使用高斯一阶偏导核来检测边缘。图 5-5(a)展示了使用这种方法进行一维信号边缘检测的示例。在边缘区域, 信号与高斯一阶偏导核进行卷积后形成一个波峰, 边缘位置对应波峰的峰值点。继续求导后, 波峰将变成一种类似“涟漪”的

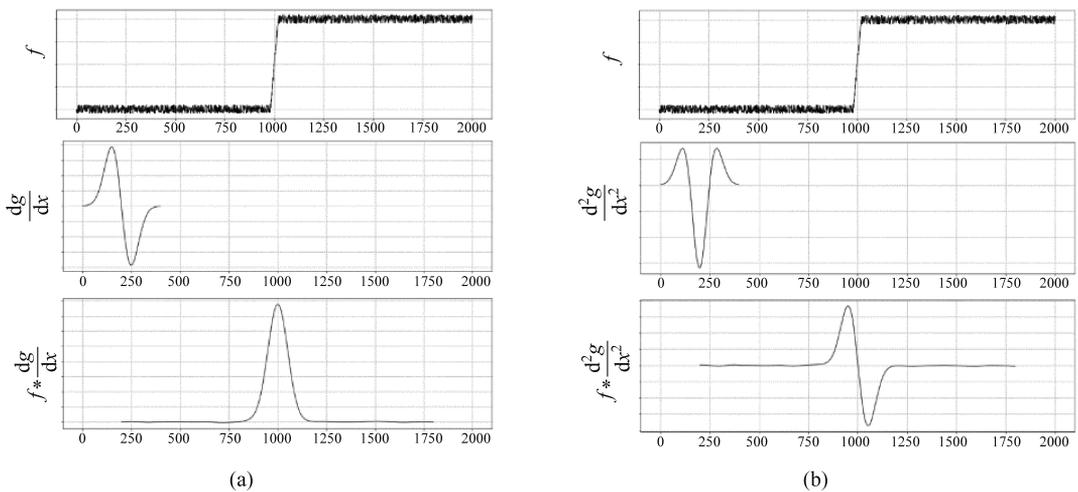


图 5-5 两种边缘检测方法



形状,此时,边缘点可以通过寻找极大值和极小值之间的过零点来确定。这个结果可以直接通过图像与高斯二阶偏导核(LoG 卷积核)进行卷积来获得,如图 5-5(b)所示。

一维理想斑点可以用图 5-6 第 1 行所示的方波信号来表示,它有两个边缘。只要确定了边缘的位置和它们之间的距离,就可以确定斑点的位置和尺寸。因此,可以使用 LoG 卷积核来检测斑点,该卷积核的设置涉及两个参数:标准差和窗口大小。通常情况下,可以将窗口的一半宽度设置为标准差的 3 倍,因此通常只需提供标准差 σ 即可。图 5-6 第 2 行展示了对第 1 行中每个斑点信号应用尺度 $\sigma=1$ 的 LoG 卷积核进行卷积后的结果。可以观察到,当斑点的尺寸较大时,卷积后的信号具有两个近似对称的“涟漪”;随着斑点尺寸的减小,这两个“涟漪”逐渐靠近并最终融合在一起;最终,当斑点尺寸与 LoG 函数曲线上两个波峰之间的跨度趋近一致时,卷积后的信号将在斑点的中心位置形成一个极值点。换句话说,如果 LoG 卷积后的信号幅度在斑点的中心位置达到最小值或最大值,那么 LoG 卷积核的尺度 σ 与斑点的尺寸“匹配”,因此可以根据该 LoG 卷积核的 σ 估计出斑点的尺寸。由此就可以得出确定斑点空间尺寸的一种方法:使用不同尺度的 LoG 卷积核对斑点进行卷积,找到一个尺度 σ ,使得卷积后的响应信号幅度在斑点的中心位置达到最小值或最大值。

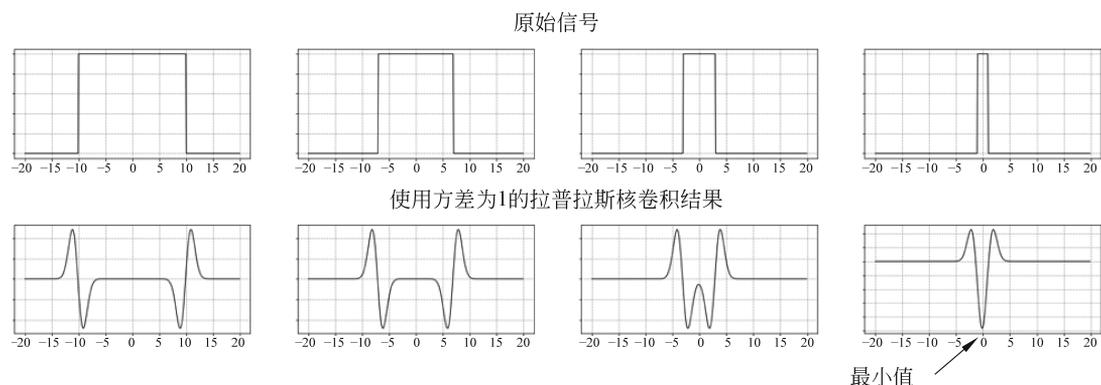


图 5-6 使用 LoG 算子检测一维斑点示意图

然而,直接应用这种思路仍然存在问题。如图 5-7 所示,对图中尺寸为 16 的一维斑点采用不同 σ 的 LoG 卷积核进行卷积时,随着 σ 的增大,输出信号的幅值出现了衰减,当 $\sigma=8$ 时,信号几乎变成了一条直线,没有出现预期的极值点。

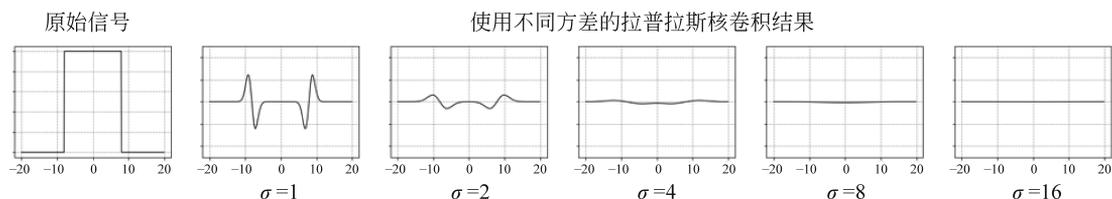


图 5-7 LoG 算子检测斑点存在的问题

为什么会产生这样的现象呢?回顾第 3 章给出的 LoG 卷积核函数,重写如下:

$$\nabla^2 g = g(x_1, x_2, \sigma) \frac{x_1^2 + x_2^2 - 2\sigma^2}{\sigma^4} \quad (5-6)$$

其中, $g(x_1, x_2, \sigma)$ 是二元高斯函数,可表示为



$$g(x_1, x_2, \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{x_1^2 + x_2^2}{2\sigma^2}} \quad (5-7)$$

利用式(5-6)对图像信号 $f_1(x_1, x_2)$ 进行卷积, 可得信号 $f_2(x_1, x_2)$ 为

$$f_2(x_1, x_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(\tau_1, \tau_2, \sigma) \frac{\tau_1^2 + \tau_2^2 - 2\sigma^2}{\sigma^4} f_1(x_1 - \tau_1, x_2 - \tau_2) d\tau_1 d\tau_2 \quad (5-8)$$

假设 $f_{1,\sigma}(x_1, x_2) = f_1(\sigma x_1, \sigma x_2)$, 式(5-8)可简化为

$$f_2(\sigma x_1, \sigma x_2) = \frac{1}{\sigma^2} (g(x_1, x_2, 1)(x_1^2 + x_2^2 - 2)) * f_{1,\sigma}(x_1, x_2) \quad (5-9)$$

式(5-9)表明, 随着卷积核尺度 σ 的增大, 信号 $f_2(x_1, x_2)$ 的幅值会衰减, 衰减因子为 $1/\sigma^2$ 。

为了解决不同 σ 的 LoG 卷积核可能引起的信号衰减问题, 可以将式(5-6)的 LoG 卷积核函数乘以 σ^2 , 以消去式(5-9)中的衰减因子, 该操作也称为尺度规范化, 于是 LoG 卷积核函数变为

$$\nabla_{\text{norm}}^2 g = \sigma^2 \nabla^2 g = g(x_1, x_2, \sigma) \frac{x_1^2 + x_2^2 - 2\sigma^2}{\sigma^2} \quad (5-10)$$

对 LoG 卷积核进行尺度规范化后, 对图 5-7 中的一维斑点使用不同的 σ 值进行测试, 结果如图 5-8 所示。响应信号的幅值不再出现衰减, 并且在 $\sigma=8$ 时出现了一个明显的极大值。

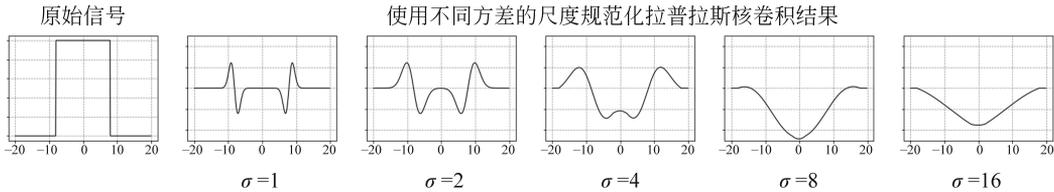


图 5-8 尺度规范化 LoG 算子的测试结果

对于二维图像中的斑点, 与前述一维斑点检测思路相似, 使用尺度规范化后的二维 LoG 卷积核进行检测。图 5-9 展示了二维 LoG 卷积核的可视化图像。

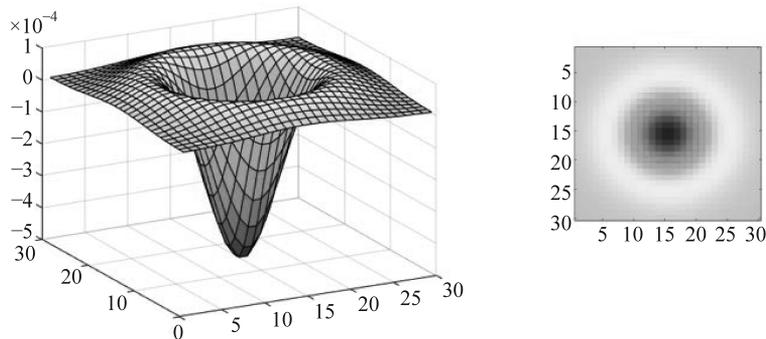


图 5-9 二维 LoG 卷积核的三维可视化图及其在水平面的投影效果

以图 5-10 中的(a)图为例, 这是一个理想的二维斑点图像, 记为 $f_1(x_1, x_2)$ 。以图像中心为原点, 斑点的半径为 r , 斑点内的像素值为 0, 斑点外的像素值为 255, 即



$$f_1(x_1, x_2) = \begin{cases} 0, & x_1^2 + x_2^2 \leq r^2 \\ 255, & x_1^2 + x_2^2 > r^2 \end{cases} \quad (5-11)$$

使用尺度规范化后的 LoG 卷积核对该斑点图像进行卷积,由式(5-10)可知,当 $x_1^2 + x_2^2 \leq 2\sigma^2$ 时,LoG 函数 $\nabla_{\text{norm}}^2 g \leq 0$, 否则 $\nabla_{\text{norm}}^2 g > 0$ 。因此,只有当 σ 满足 $2\sigma^2 = r^2$ 时,原点处卷积的响应值才会达到极大值,此时卷积核的尺度称为特征尺度,记为 σ^* , 满足

$$\sigma^* = \frac{r}{\sqrt{2}} \quad (5-12)$$

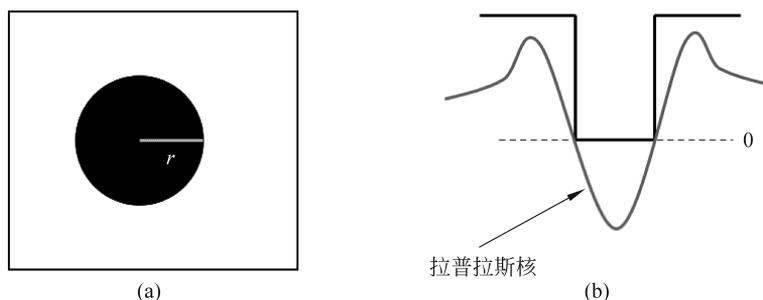


图 5-10 使用规范化后的 LoG 卷积核对该斑点图像进行卷积所得响应取极大值时的情况示意图

使用 LoG 卷积核检测斑点时,需要引入一个被视为尺度的参数 σ ,通过连续变化尺度参数 σ 可获得一个多尺度的图像序列,构成所谓的图像尺度空间,然后在尺度空间上搜索特征尺度。因此,也称其为尺度空间的斑点检测器,搜索过程主要有以下两个步骤。

- ① 利用尺度规范化后的 LoG 卷积核函数产生卷积模板,在不同尺度上对图像进行卷积。
- ② 在尺度空间中找到 LoG 响应的极值,其所对应的空间尺度为特征尺度。

对图 5-11 中的葵花图案,通过图像的 LoG 卷积结果探测出在葵花中心位置的暗色斑点。接着,在该位置上应用不同尺度的 LoG 卷积核,以搜索特征尺度,确定斑点的半径。在特征尺度下,且 LoG 卷积核的中心与斑点的中心重合时,LoG 卷积的响应值最高。基于这一特性,即使随后图像尺寸发生变化,也能独立确定斑点的位置和半径,保持检测区域的内容一致,实现尺度不变的斑点区域检测。

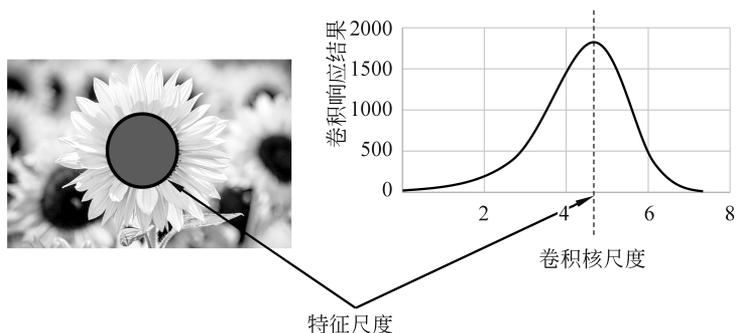


图 5-11 二维斑点检测示意图



5.1.3 Harris-Laplace 检测器

根据 5.1.3 节的介绍,可以使用 LoG 卷积核提取具有尺度不变性的区域,而不仅仅局限于斑点的检测。只要某个尺度的 LoG 卷积核在图像的某个位置产生极大值,那么以这个点为圆心、半径为 $\sqrt{2}\sigma$ 的区域就被视为一个尺度不变区域。无论图像尺寸如何变化,LoG 卷积核都能够检测到相同的点和包含相同图像内容的区域。

LoG 卷积可以作为 5.1.1 节中所需的内容响应函数,因其满足前面提出的内容响应函数属性。首先,LoG 卷积核满足内容响应函数的要求,具有尺度不变性。对图像 $f_1(x_1, x_2)$,使用尺度规范化后的 LoG 卷积核进行卷积得到 $f_2(x_1, x_2)$:

$$f_2(x_1, x_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(\tau_1, \tau_2, \sigma) \frac{\tau_1^2 + \tau_2^2 - 2\sigma^2}{\sigma^2} f_1(x_1 - \tau_1, x_2 - \tau_2) d\tau_1 d\tau_2 \quad (5-13)$$

假设对信号 $f_1(x_1, x_2)$ 进行尺度拉伸,得到 $f_{1,1/\gamma}(x_1, x_2) = f_1(x_1/\gamma, x_2/\gamma)$,利用尺度为 σ/r LoG 卷积核对信号 $f_{1,1/\gamma}(x_1, x_2)$ 进行卷积,再经变换后可得信号 $\tilde{f}_2(x_1, x_2)$,即

$$\tilde{f}_2(x_1, x_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(\tau_1, \tau_2, \gamma\sigma) \frac{\tau_1^2 + \tau_2^2 - 2\gamma^2\sigma^2}{\gamma^2\sigma^2} f_{1,1/\gamma}(x_1 - \tau_1, x_2 - \tau_2) d\tau_1 d\tau_2 \quad (5-14)$$

对比式(5-13)与式(5-14),可以得到

$$\tilde{f}_2(x_1, x_2) = f_2\left(\frac{x_1}{\gamma}, \frac{x_2}{\gamma}\right) \quad (5-15)$$

因此,LoG 卷积响应量的变化与图像尺度的变化是协变的。其次,当 LoG 的尺度与信号 $f_1(x_1, x_2)$ 的尺度匹配时,LoG 的响应会且仅会产生一个明显的极值。

将第 4 章介绍的 Harris 检测器与 LoG 卷积核结合,便能得到具有尺度不变性的特征点检测器——Harris-Laplace 检测器。该检测器先采用各种尺度的规范化 LoG 卷积核对输入图像进行卷积,在尺度空间生成一组响应图像。然后在每个响应图像上使用 Harris 角点检测器检测角点在空间的位置(指像素点坐标)。最后,搜索尺度空间找到 Harris 角点检测器的响应极值点,进而确定其是否为所选特征点。此时,检测到的角点区域是尺度不变的区域。通过这种方式,筛选的角点便具有了尺度不变性。

5.2 SIFT

5.2.1 高斯差分尺度空间

根据前面学习的内容,可以借助 LoG 卷积核函数确定尺度不变区域,并与特征点检测器结合,以获取具有尺度不变性的特征点。然而,LoG 卷积核的计算成本较高,在 SIFT 算法中,通常更多地使用一种与之近似的卷积核,即差分高斯(Difference of Gaussians, DoG)卷积核。DoG 核函数通过采用两个不同尺度的高斯核函数进行差分运算而得,定义如下

$$\text{DoG}(\sigma) = G(x, y, k\sigma) - G(x, y, \sigma) \quad (5-16)$$

其中, k 表示相邻尺度高斯核之间的比例因子。通过将 DoG 卷积核直接应用于原始图像,等同于使用尺度相差 k 倍的 2 个高斯核分别对原始图像进行卷积,然后计算它们之间的差值。



1994 年, Lindeberg 证明了 DoG 核函数近似等于 LoG 核函数, 这种近似关系可以通过热扩散方程推导得到^[8]。图 5-12 展示了一维 LoG 与 DoG 核函数的曲线图, 它们的函数曲线非常相似, 因此它们都可用于实现具有尺度不变性的特征点检测。

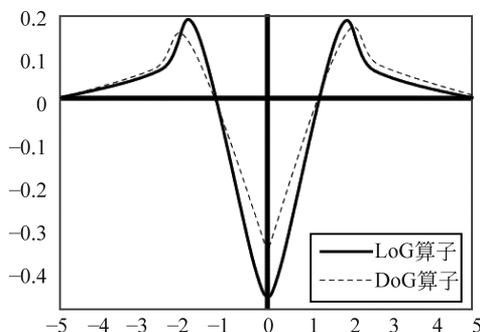


图 5-12 一维 LoG 与 DoG 算子的函数曲线图

在 SIFT 算法中, 首要任务是构建高斯差分金字塔, 在尺度空间中对图像进行多尺度表示。在此过程中, 图像经过一系列不同尺度的高斯平滑操作, 产生了一组平滑图像, 可用于分析和处理不同尺度下的图像结构。在图像尺度空间中, 随着高斯卷积核尺度的逐渐增大, 图像的平滑效果逐渐加强, 同时细节信息逐渐减少, 可以模拟人眼从近到远观察目标时目标在视网膜上的呈现过程。因此, 尺度空间有助于更好地捕捉图像的本质特征。

其次, 在图像尺度空间中进行特征提取具备灰度和对比度不变性。在拍摄目标时, 光照条件的变化导致输出图像的亮度水平和对对比度发生变化, 但图像的内容是一致的。因此, 提取的图像特征应确保不受图像灰度水平和对对比度变化的影响, 也就是要具备灰度不变性和对比度不变性, 这对下游图像处理任务非常重要。

最后, 在视觉感知中, 当观察者和物体之间的相对位置发生变化时, 视网膜感知到的图像的位置、大小、角度和形状也会发生变化。因此, 在尺度空间中提取的图像特征需要具备不受图像的位置、大小、角度和仿射变换影响的性质, 也就是要具备几何不变性。

尺度空间一般用高斯金字塔的形式表示, 如图 5-13 所示, 图像尺度空间一般用高斯金字塔的形式表示。高斯金字塔是通过将图像与多个不同采样因子和尺度因子的高斯核函数进行卷积而构建的, 可表示为

$$L(x, y, \sigma, p) = p * G(x, y, \sigma) * I(x, y) \quad (5-17)$$

其中, $I(x, y)$ 表示原始图像, p 表示采样因子, $G(x, y, \sigma)$ 是高斯核函数, σ 是尺度空间因子, 即标准差, 它决定了图像的模糊程度。在较大尺度下 (σ 值较大), 高斯金字塔中的相应图像呈现出图像的概貌信息, 而在较小尺度下 (σ 值较小), 则呈现出图像的细节信息。在构建尺度空间时, 高斯金字塔被划分为多个图像组 (Octave), 每个组内包含多个图像层 (Interval)。在同一组内, 每一层的图像都是原始图像经过与尺度按 k 倍增加的高斯卷积核的卷积运算得到的结果。假设高斯金字塔总共有 O 个组, 每个组内包含 S 层图像, 那么每一组中的尺度因子 k 应为 $2^{\frac{1}{S-3}}$, 而不是 $2^{\frac{1}{S}}$ 。这是因为, 在每组高斯金字塔图像中, 对第一幅和最后一幅图像无法实行极值检测 (用于找到尺度不变的关键点), 此外, 极值检测之前需要计算相邻高斯金字塔图像之差, 因此, 实际可实行极值检测的图像只有 $S-3$ 幅, 第 O_{i+1} 组的第 1 幅图

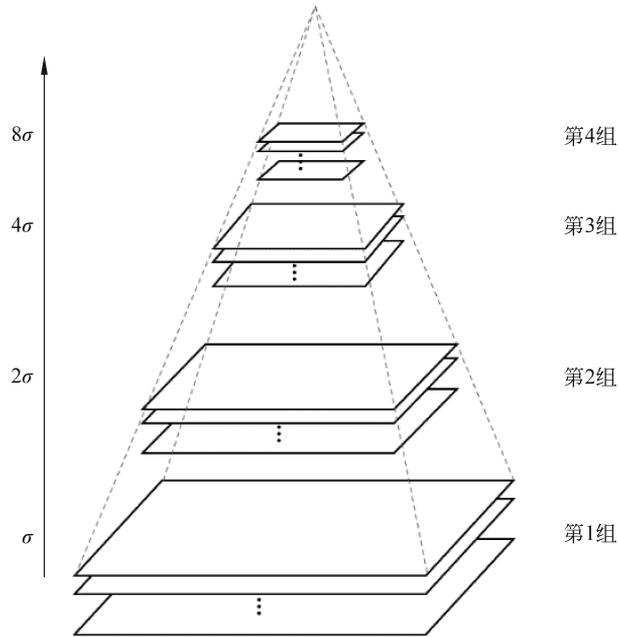


图 5-13 高斯金字塔示意图

像则由第 O_i 组的第 $S-3$ 幅图像(即倒数第 3 幅图像)进行二倍下采样得到,这样可保证尺度连续。

通过对高斯金字塔中相邻尺度的图像进行相减,获得相应尺度的 DoG 卷积结果,从而得到高斯差分金字塔,如图 5-14 所示。

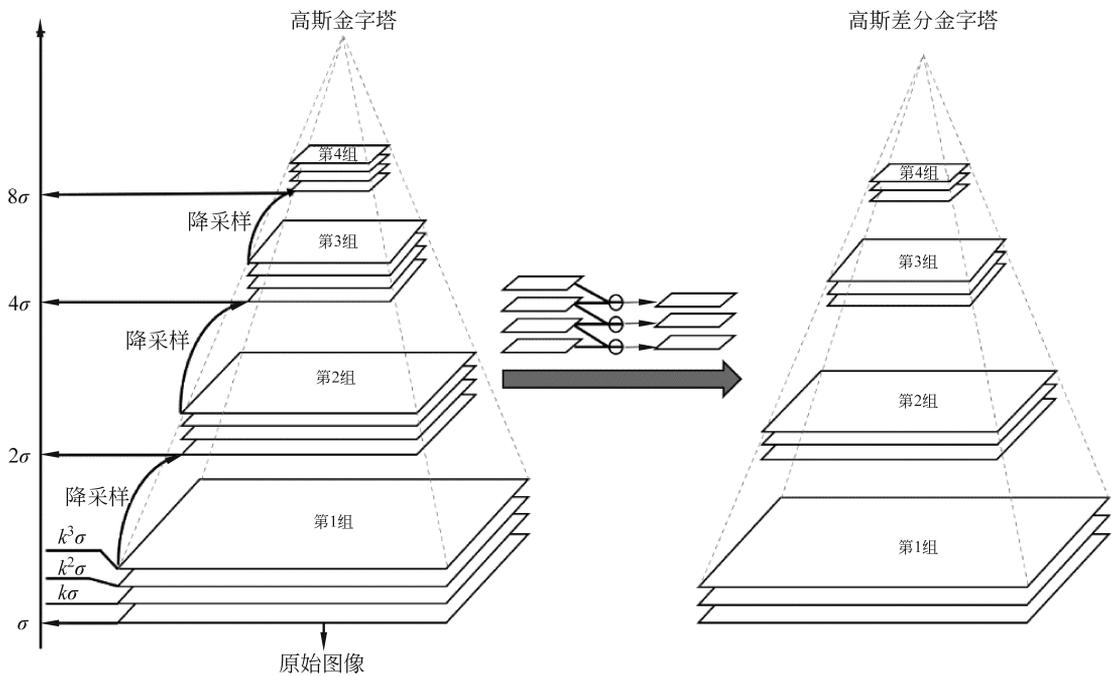


图 5-14 高斯差分金字塔构建示意图



5.2.2 SIFT 特征点检测

为了获得尺度不变的特征点,需要在 5.2.1 节构建的高斯差分金字塔中检测极值点。具体地,对于高斯差分金字塔的某一图像层,将其每个像素点与同一层的相邻 8 个像素点以及它上一层和下一层的 9 个相邻像素点(总共 26 个相邻像素点)进行比较。如图 5-15 所示,如果标有叉号的像素点的 DoG 值在所有相邻 26 个像素点中都是最大或最小的,那么该点将被视为一个局部极值点,并记录下它的位置和对应的尺度。需要注意的是,这些检测到的极值点是在离散空间中的极值点,在连续空间中可能并不是真正的极值点,如图 5-16 所示。此外,由于 DoG 值对噪声和边缘比较敏感,因此需要对检测到的极值点进行进一步的筛选和验证。Lowe 的论文中提到,可以通过拟合三元二次函数来精确获取关键点的位置和所在尺度,具体细节可以参考相关文献^[9]。

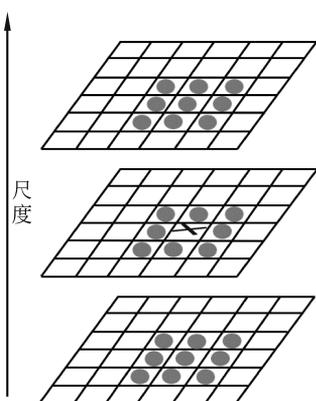


图 5-15 DoG 空间局部极值检测示意图

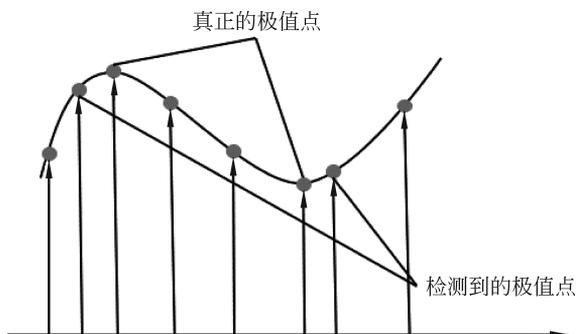


图 5-16 离散空间极值点与连续空间极值点

经过上述步骤获得的尺度空间中的局部极值点作为关键点。随后,通过使用图像的局部特性为每个关键点分配一个稳定的方向,称为主方向。在具体计算中,根据关键点 (x^*, y^*) 的尺度 σ^* 找到最接近尺度 σ^* 的高斯金字塔图像 $L^*(x, y)$ 。在图像 $L^*(x, y)$ 上,以关键点 (x^*, y^*) 为中心,取一个正方形窗口(其边长通常取 $3 \times 1.5\sigma^*$ 个像素,当然需要四舍五入)。接下来,在窗口内的每个像素点计算梯度方向,这可以通过第 2 章中的图像梯度计算方法来实现。然后,建立这些局部梯度方向的统计直方图,横轴是梯度方向,范围是 $0 \sim 360^\circ$,每 10° 为一格,总共 36 个方向,如图 5-17 所示。纵轴是加权的梯度,对属于某个梯度方向的所有局部梯度模值进行高斯加权处理,这意味着距离中心关键点更近的像素的梯度方向对直方图的贡献更大。高斯加权的圆形窗的尺度通常设置为特征点尺度 σ^* 的 1.5 倍。

关键点的主方向 θ_m^* 被确定为梯度直方图的峰值所在的方向,这样可以确保主方向的稳定性。如果存在其他局部梯度峰值,其峰值高度大于最高峰值的 80%,那么这些方向被认为是该特征点的辅方向。一个特征点可能具有多个方向,包括一个主方向和多个辅方向。为了更精确地定位峰值位置,通常会使用抛物线插值来拟合梯度方向直方图中的多个峰值。

后续生成的局部特征表达都是相对于主方向的,因此即使目标在另一个图像中发生旋