

# SIFT 特征检测算法整理

## 一. 基本思想

SIFT 作为一种著名的图像特征提取算法，由 David Lowe 在 1999 年所发明，2004 年完善总结并发表。它的基本思想为在构建的尺度空间中寻找极值点（尺度空间极值检测），提取出其位置（关键点定位）、尺度、以及一个主方向，并在此基础上以关键点所在的尺度及其周围梯度构建描述子。

该算法具有**尺度不变性**、**旋转不变性**及**光照不变性**，受仿射变换、噪音等因素的影响较小。其描述子具有的信息量较大，因而对于后续的特征匹配过程而言有很强的辨识性。但是与此同时，不断地下采样和插值等操作带来的**巨大计算量**使得 SIFT 算法很难直接应用于一些实时性要求较高的领域。而对于边缘光滑的目标，很难准确提取特征。

## 二. 尺度空间极值检测

在观察物体时，所感知到的图像的亮度、对比度等信息受物体的光照、位置等因素影响较大。因此，在分析图像时，需要一种图像处理算子将这些影响剥离，使处理后的结果与灰度、对比度、平移、尺度、仿射变换等因素无关。在这一思路下，Iijima 于 1962 年提出了尺度空间（scale space）模型，并在之后经 Witkin 和 Koenderink 等人的推广逐渐得到关注，在计算机视觉邻域得到了广泛应用。

其基本思想为，模拟相机拍摄物体时由近到远的图像变化过程，在图像信息处理模型中引入一个尺度参数，以改变尺度参数的形式获得多尺度下的图像序列（图像的模糊程度随尺度增大而增大），并提取其主轮廓。之后在此基础上，实现边缘、角点检测和不同分辨率上的特征提取等。这一方法改单尺度下的单一图像分析为不同尺度下的动态分析，相对而言更加容易规避许多影响因素而获得图像本质信息。具体为在所有可能的尺度下搜索**稳定、不变特征**，以检测不同尺度下具有不变性的位置，作为候选位置，进而进一步考察其细节。

### 2.1 尺度空间构建

要在所有尺度下搜索，进行特征检测，首先要构建图像的尺度空间。一个图像的尺度空间，实际上是一个尺度连续函数，定义为一个变化尺度的高斯函数与原图像的卷积。尺度空间描述一般指图像在不同尺度下的描述。

尺度空间函数为

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y)$$

其中\*为卷积， $G(x, y, \sigma)$ 为以 $\sigma$ 为方差的高斯函数

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}$$

为更高效地检测不变特征，Lowe 在 1999 年提出，通过相差相邻乘系数 $k$ 的两个尺度的高斯差分（difference-of-Gaussian, DoG 算子），即 $D(x, y, \sigma)$ ，使用尺度空间极值的方式来检测不变特征。

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

一方面，在计算图像平滑 $L(x, y, \sigma)$ 后， $D$ 只需计算图像减法即可，相对而言更加高效；另一方面，DoG 提供了尺度规范下高斯拉普拉斯（Laplacian of Gaussian, LoG 算子，即 $\sigma^2 \nabla^2 G$ ）的近似解，有研究指出，相比许多梯度、角点函数，**LoG 的最大、最小值产生了最稳定的图像特征**。因此，选择了这一函数来检测尺度空间极值。

为更快求取 $\sigma^2 \nabla^2 G$ ，式中 $\sigma \nabla^2 G$ 可由 $\partial G / \partial \sigma$ 的有限差分近似表示，其中使用相邻尺度 $k\sigma$ 和 $\sigma$ 进行差分

$$\sigma \nabla^2 G = \frac{\partial G}{\partial \sigma} \approx \frac{G(x, y, k\sigma) - G(x, y, \sigma)}{k\sigma - \sigma}$$

移项得

$$\sigma^2 \nabla^2 G \approx \frac{G(x, y, k\sigma) - G(x, y, \sigma)}{k - 1}$$

即

$$G(x, y, k\sigma) - G(x, y, \sigma) \approx (k - 1) \sigma^2 \nabla^2 G$$

其中 $k$ 为常值，因而不影响极值的位置。理论上 $k$ 接近于1时，近似误差接近于0，但是事实上这一因素对极值检测毫无影响。

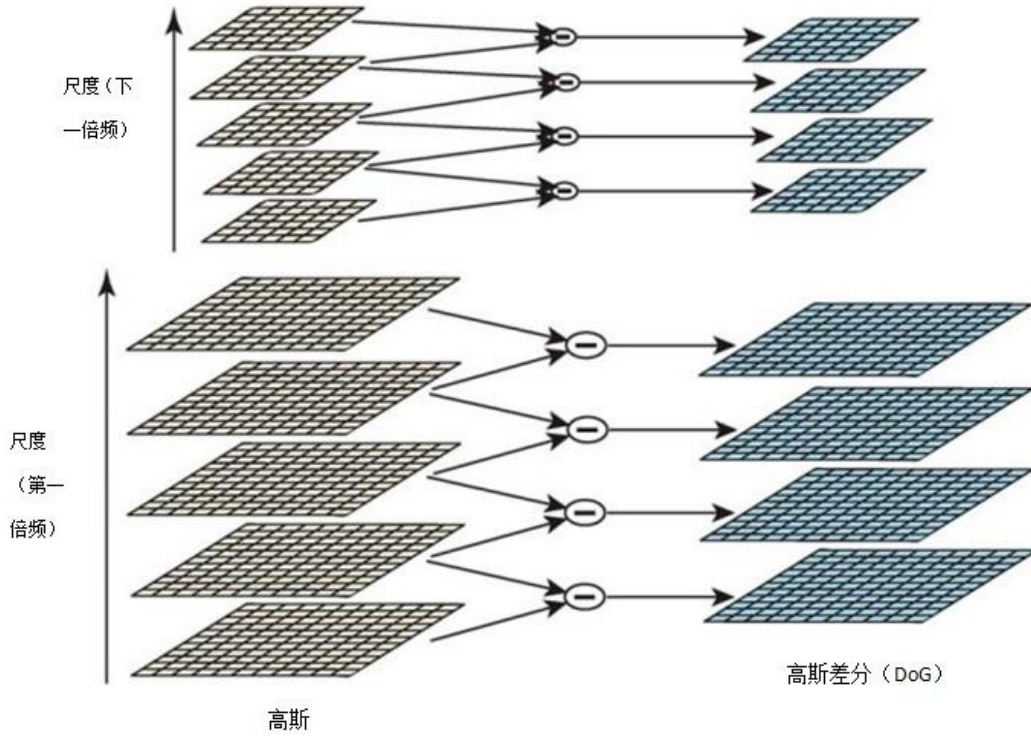


图1 尺度空间的构建

构建图像的尺度空间的具体的方式如图1, 相同大小, 不同模糊程度图像构成一组 (octave), 不同的 octave 构成尺度空间。初始图像与不同参数 $\sigma_i$ 的高斯函数卷积所得尺度空间如图像左侧所示, 相邻的高斯图像相减产生了右侧的差分高斯 (DoG) 图像。每个 octave 后, 高斯图像以取各行列上的第偶数个像素的方式被降采样 (长、宽各为上一 octave 一半)。将每个 octave 的尺度分为整数个间隔, 由于取极值时是在相邻3幅图像内取中间那幅图像中的极值 (见下节), 因此为保证取极值时覆盖整个 octave, 那么在 $s$ 层图像中取得极值的话就需要有 $s + 2$ 幅 DoG 图像; 而 DoG 是由相邻尺度高斯图像相减而来, 因此需要每个 octave 内有 $s + 3$ 幅高斯图像。

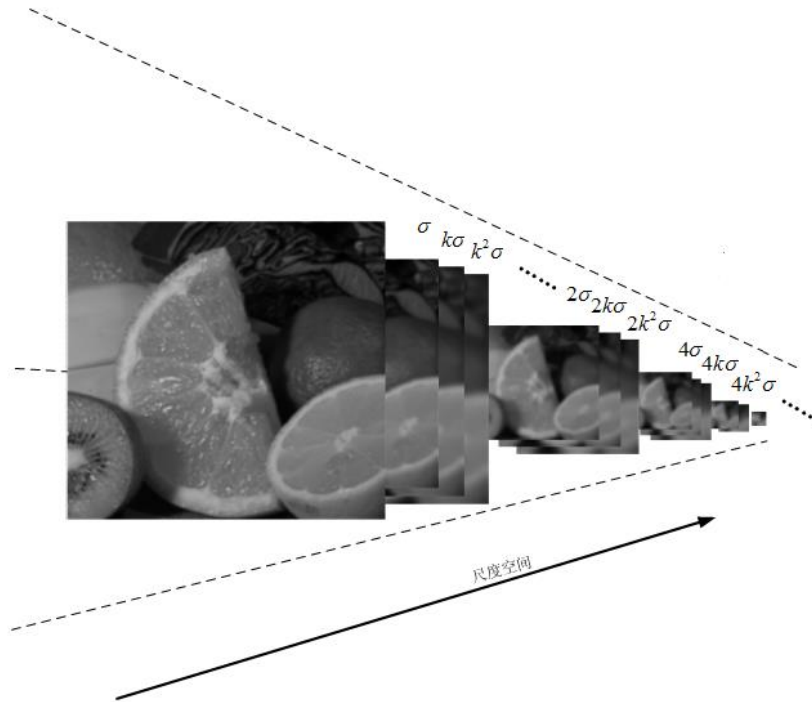


图 2 不同尺度的高斯系数

如图 2，取第 0 个 octave 的第 0 幅图像的高斯系数为  $\sigma_0$ ，第  $p$  个 octave 中第 0 幅图像的高斯系数为  $2^p \sigma_0$ ；而在同一 octave 内，相邻图像的尺度间距为  $k$ 。为了保证整个尺度空间在取极值时连续，取  $k = 2^{1/s}$ ，那么第  $p$  个 octave 中第  $q$  幅图像的高斯系数为

$$k^q 2^p \sigma_0 = 2^{p+\frac{q}{s}} \sigma_0$$

## 2.2 极值位置检测

如图 3，在 DoG 中进行极值检测，通过将一个像素（记为叉）与其临近的  $3 \times 3$  区域内的 26 个像素（其中有上下相邻尺度的 18 个像素和同一尺度下 8 个像素）进行对比，只选取比所有近邻都大/小的点。由于在前几次检测中大多数的样本点会被排除，因此其所需的资源相对较小。

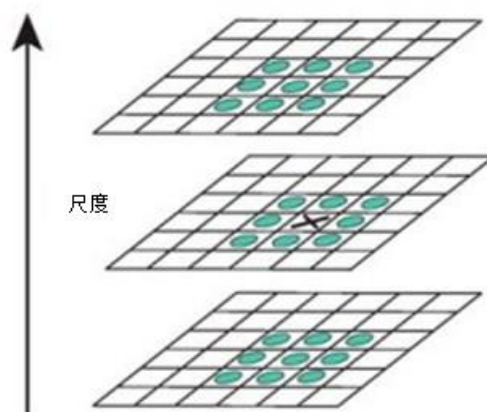


图 3 局部极值检测

另一方面，图像的尺度采样频率（即  $s$  的选取）和空间域采样频率（即  $\sigma$  的选取）都会影响到图像处理效率和结果的完整性，实际应用中，我们必须使用一个兼顾效率和完整性的方案。

在 Lowe 的论文中选择了  $s = 3$  和  $\sigma = 1.6$  的参数。

### 三. 关键点准确定位

在上一节中已经通过极值筛选得到了离散空间中的关键点的候选值, 为更加精确地确定关键点的位置和尺度, 需要通过插值的方式来精确定位, 并在这一过程中淘汰掉低对比度 (对噪声敏感) 或定位在边缘的一些点。

#### 3.1 子像素插值 (Sub-pixel Interpolation) 精确定位

利用已知的离散空间点插值得到的连续空间极值点的方法叫做子像素插值 (Sub-pixel Interpolation)。在 Lowe 的论文中, 使用了 Brown 提出的一种通过局部样本点的三维二次方程, 来决定最大值插值的方法。与此同时, 低对比度的不稳定候选点将被排除。

按 Brown 提出的方法, 对尺度空间 DoG 函数进行曲线拟合。将  $D(x, y, \sigma)$  以候选点为原点进行 Taylor 展开, 展开到二次项

$$D(\mathbf{x}) = D + \frac{\partial D}{\partial \mathbf{x}}^T \mathbf{x} + \frac{1}{2} \mathbf{x}^T \frac{\partial^2 D}{\partial \mathbf{x}^2} \mathbf{x}$$

其中  $D$ 、 $\frac{\partial D}{\partial \mathbf{x}}$  以及  $\frac{\partial^2 D}{\partial \mathbf{x}^2}$  是尺度空间函数及其导数在候选点的取值,  $\mathbf{x} = (x, y, \sigma)^T$  是相对于候选点的偏差补偿。

函数的极值点  $\hat{\mathbf{x}}$ , 就是将函数相对于  $\mathbf{x}$  的导数置零的点, 那么

$$\hat{\mathbf{x}} = - \left( \frac{\partial^2 D}{\partial \mathbf{x}^2} \right)^{-1} \frac{\partial D}{\partial \mathbf{x}}$$

此时的极值为

$$D(\hat{\mathbf{x}}) = D + \frac{1}{2} \frac{\partial D}{\partial \mathbf{x}}^T \hat{\mathbf{x}}$$

$D$  及其各阶导数由候选点与其邻点之间进行差分来近似, 这样可以在较低的计算成本下得到  $3 \times 3$  的线性结果 (即二阶导数)。当偏差补偿  $\hat{\mathbf{x}}$  的任一维度的值大于 0.5 (即  $x, y, \sigma$  中的有至少一个值  $> 0.5$ ), 就意味着极值会更接近其他的候选点, 这种情况下候选点将会改变, 插值将执行取代原候选点。

最终将偏差补偿  $\hat{\mathbf{x}}$  与原候选点相加, 以得到极值位置的插值估计。假定各点像素值取值在  $[0, 1]$  内, 那么  $D(\hat{\mathbf{x}}) < 0.03$  的点将会被视为低对比度的不稳定点, 被排除。

#### 3.2 消除边缘响应

就稳定性而言, 只排除低对比度的关键点是不够的。即使位置在边缘的点具有很差的决策性和噪声稳定性, DoG 函数在边缘也会有很强的响应。因此在这一步需要将这些位于边缘的点排除。

一维函数曲线相应位置的曲率可以表示函数在该点的变化趋势; 而在二维函数的曲面上, 不同方向的曲率也表示了函数在该点该方向上的陡峭程度。对于边缘上的点而言, 将会沿边沿产生极大的曲率, 而垂直边缘方向的曲率则很小。我们称远大于另一个的曲率为主曲率, 可以通过检测曲率的方式检测边缘上的点。

在 Lowe 的论文中借用了 Harris 的方法, 使用 Hessian 矩阵的特征值来检测边缘。矩阵的特征值对应着在特征向量方向上的投影, 该值越大就反应函数在该方向变化趋势越缓慢, 也就是曲率越大。即 Hessian 矩阵的特征值和特征值所在特征向量方向上的曲率成正比。

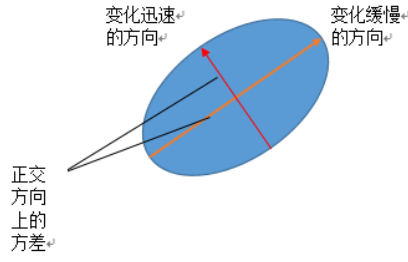


图 4 特征值与变化趋势

候选点的 Hessian 矩阵  $\mathbf{H}$  可由该点与其所在位置、尺度的邻点进行二阶差分求得

$$\mathbf{H} = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix}$$

为减小计算量，并不将  $\mathbf{H}$  的特征值计算出来，只需得知它们的比值即可检测边缘。以  $\lambda_1$  表示较大的特征值，以  $\lambda_2$  表示较小的特征值，那么

$$\lambda_1 + \lambda_2 = \text{trace } \mathbf{H} = D_{xx} + D_{yy}$$

$$\lambda_1 \lambda_2 = \det \mathbf{H} = D_{xx} D_{yy} - D_{xy}^2$$

由于  $\det \mathbf{H}$  必然为正，因此当  $\lambda_1$  及  $\lambda_2$  符号不同时，候选点不为极值，舍去。设  $\gamma$  为  $\lambda_1$  和  $\lambda_2$  的比值，那么

$$\frac{\text{trace}^2 \mathbf{H}}{\det \mathbf{H}} = \frac{(\lambda_1 + \lambda_2)^2}{\lambda_1 \lambda_2} = \frac{(\gamma + 1)^2}{\gamma}$$

易知上式在  $\gamma = 1$ ，即两特征值相等时可取最小值，因此检测上式是否小于指定值就可以检测  $\gamma$  是否小于阈值（ $\gamma$  若达到阈值，则判定为边缘），即

$$\frac{\text{trace}^2 \mathbf{H}}{\det \mathbf{H}} < \frac{(\gamma_0 + 1)^2}{\gamma_0}$$

Lowe 的论文中取  $\gamma_0 = 10$ ，若上式的值大于该阈值，则判定为边缘，将候选点排除。

#### 四. 指定关键点方向

在完成对候选点的筛选之后，利用图像的局部属性为每一个关键点分配一个方向。特征描述子的建立与这个方向相关，以实现 SIFT 特征的旋转不变性。在关键点所在尺度下，计算相应位置梯度的大小  $m(x, y)$  和方向  $\theta(x, y)$

$$m(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2}$$

$$\theta(x, y) = \tan^{-1} \left( \frac{L(x, y+1) - L(x, y-1)}{L(x+1, y) - L(x-1, y)} \right)$$

将关键点周围区域的采样点的梯度加权组成梯度方向直方图，直方图有 36 个 bin (即柱子)，覆盖  $360^\circ$  的方向。每个采样点按所属方向加入直方图的不同 bin 中，所添加高度值为梯度 \* 权值。所用权值为以  $\sigma_0 = 1.5\sigma$  ( $\sigma$  为关键点所在尺度) 为参数的高斯函数的函数值。按尺度采样的  $3\sigma$  原则，用来采集样本点的邻域窗口半径为  $r = 4.5\sigma$ 。

即，每个 bin 的高度为

$$\sum_{\text{梯度在此方向的样本点}} \text{样本点梯度值} \times \text{高斯函数加权值}$$

如图 5，按步骤即可得到梯度方向直方图。在这个方向直方图中，最大峰值对应的 bin 就是关键点的方向。（图 5 为简化，只画出了 8 个方向）

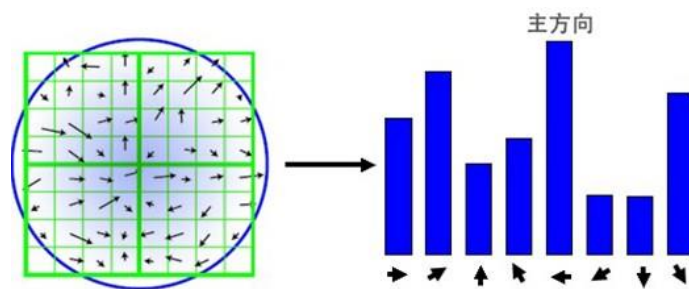


图5 关键点梯度方向直方图

对于高度达到峰值的 80%的 bin，将会在该方向上另外建立一个关键点。若有多个相似量级的 bin，则可以在该尺度和位置上建立多个关键点。只有 15%的点会被指定多个方向，但它们对匹配的稳定性有很大意义。

## 五. 局部图像描述子

前面提及的部分（尺度空间极值检测、关键点定位、定向分配）为每个关键点都分配了位置、尺度和方向。这些参数利用一个可重复的局部二维坐标系来描述局部图像区域，进而为这些参数提供不变性。我们需要做的下一步这是计算出**局部图像区域**的描述子，它要在其高度独特性的同时对于其余因素（如光照或三维视角等）的变化尽可能地保持其不变性。

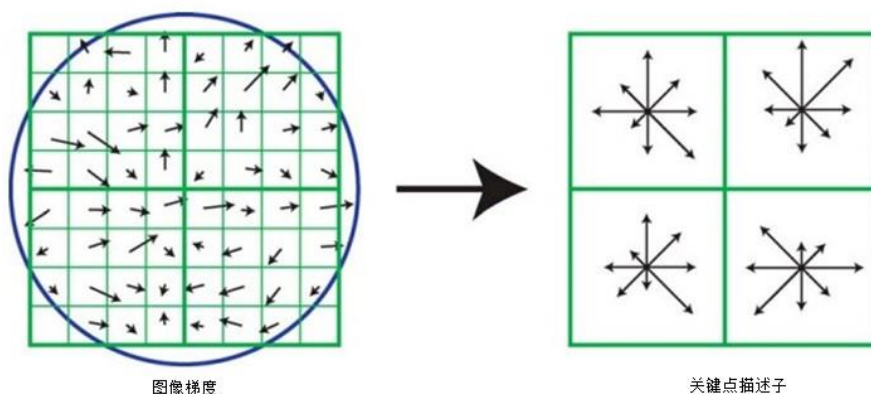


图6 关键点描述子计算

如图 6 左图，在关键点所在的尺度，将采样点周边半径为 $r$ 的邻域（Lowe 论文中为 $16 \times 16$ 区域）内的采样点按 $d \times d$ （Lowe 论文中为 $4 \times 4$ ）为一个子区域。与之前类似，按照计算关键点方向时已经预计算出的梯度，以高斯函数加权（高斯参数 $\sigma = 1.5r$ ）在子区域内建立梯度方向直方图，直方图中 bin 的数量为 $h$ （Lowe 论文中为 8 个 bin）。如图 6 的右图，按直方图的方向、量级在每个子区域内画出箭头。每个直方图中的每个 bin 为描述子提供一个参数，这样一个关键点的特征描述子就是一个长度为 $\frac{r^2}{d^2} \times h$ 的**向量**（论文中是 $\frac{16^2}{4^2} \times 8 = 128$ 维的特征向量）。

为了消除光照影响，将描述子向量进行归一化，即整个向量被标准化为单位长度。

为了达到方向不变性，描述子的坐标和梯度方向**都随关键点方向进行旋转**，在旋转以及插值后计算梯度方向，并计算描述子向量。

## 参考

- [1] Lowe D G. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004, 60(2): 91-110.
- [2] zdd. SIFT 算法详解[EB/OL]. <https://blog.csdn.net/zddbblog/article/details/7521424>, 2018-4-28.
- [3] losteng. SIFT 中的尺度空间和传统图像金字塔[EB/OL]. <https://blog.csdn.net/losteng/article/details/50815322>, 2016-3-6.
- [4] 一只有恒心的小菜鸟. SIFT 特征提取分析[EB/OL]. <https://www.cnblogs.com/YiXiaoZhou/p/5893835.html>, 2016-9-21.