Scale Invariant Feature Transform Note

FanmingL

2018 年 12 月 1 日

1 SIFT 特征提取

1.1 概述

SIFT 即尺度不变特征变换，顾名思义，它是一种对尺度变化能够保持不变的一种特征，实际上他也对旋转、亮度等保持不变性，在之后的处理流程中，可以一步步看出他是如何对尺度和旋转保持不变的。

1.2 DoG 空间极值检测

1.2.1 高斯尺度空间

我们都知道高斯模糊是由 2 维的高斯函数 $G(x,y,\sigma)=\frac{1}{2\pi\sigma^2}\exp(-\frac{x^2+y^2}{2\sigma^2})$ 与图像卷积得到的。即对图像 $L(x,y)$

$$L(x,y,\sigma)=G(x,y,\sigma)*I(x,y) \tag{1}$$

符号 $*$ 表示卷积操作，对于越大的 σ，$L(x,y,\sigma)$ 相比于原图，越模糊，但是图像大小还是一样的，于是用不同的 σ 对同一张图进行处理，我们就能得到多张图像大小一样，但是 σ 大小不一样的图像。$L(x,y,\sigma)$ 就是高斯尺度空间，σ 称为尺度空间因子。

1.2.2 差分高斯

为了能够检测出图像中的特征点，我们将高斯尺度空间相邻的两个图像相减得到差分高斯 DoG 的响应图像。设 k 为相邻两个 $L(x,y,\sigma)$ 的尺度空间因子的比值，DoG 可定义为

$$D(x,y,\sigma)=L(x,y,k\sigma)-L(x,y,\sigma) \tag{2}$$

1.2.3 图像金字塔

图1为多分辨率图像金字塔，高层的图像由低层的图像降采样而来。但是其本质是降采样，无法保持特征的尺度不变性。

1.2.4 高斯金字塔

如图2，高斯金字塔结合了图像金字塔和高斯尺度空间，高斯金字塔是对像素金字塔的每层加上高斯滤波构建高斯尺度空间得到。

可以看出高斯金字塔有多个组，每组又有多层，每组的每个层之间的高斯尺度是不一样的，每一层上一层的高斯尺度是这层的 k 倍，由于上一组图像的最底层是由下一组图像的尺度为 2σ 的图像进行因子为 2 的降采样得到的，所以假设每组有 S 层，那么 $k=2^{\frac{1}{S}}$，不妨设当前组数为 o，层数为 s，那么对应的 σ 为

$$\sigma(o,s)=\sigma_0\cdot 2^{o+\frac{s}{S}} \tag{3}$$

一般高斯金字塔的总组数在 $[1,\log_2\min(rows,cols)]$ 之间，其中 rows、cols 分别表示图像的行数和列数。

在 SIFT 算法提出作者的实验中，令 $\sigma_0=1.6, o_{min}=-1, S=3$，$o_{min}=-1$ 表示首先将原图像长宽各拓展一倍的结果。

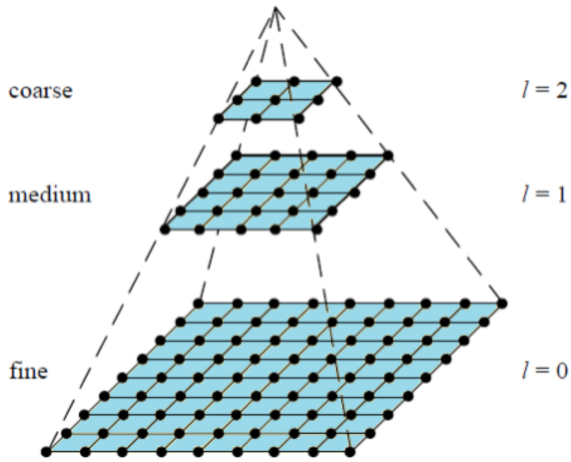

图 1: 多分辨率图像金字塔

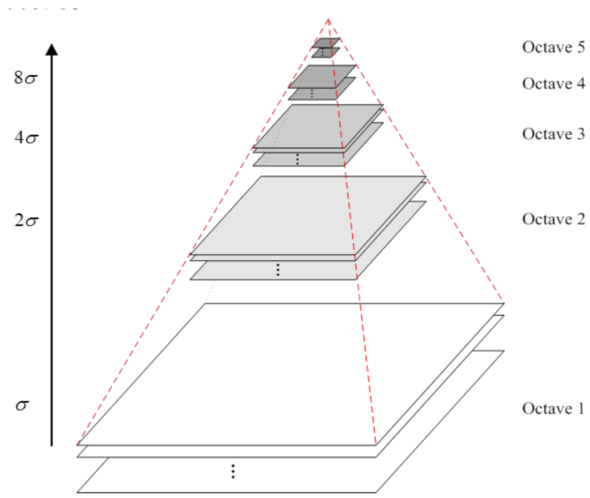

图 2: 高斯金字塔

1.2.5 DoG 金字塔

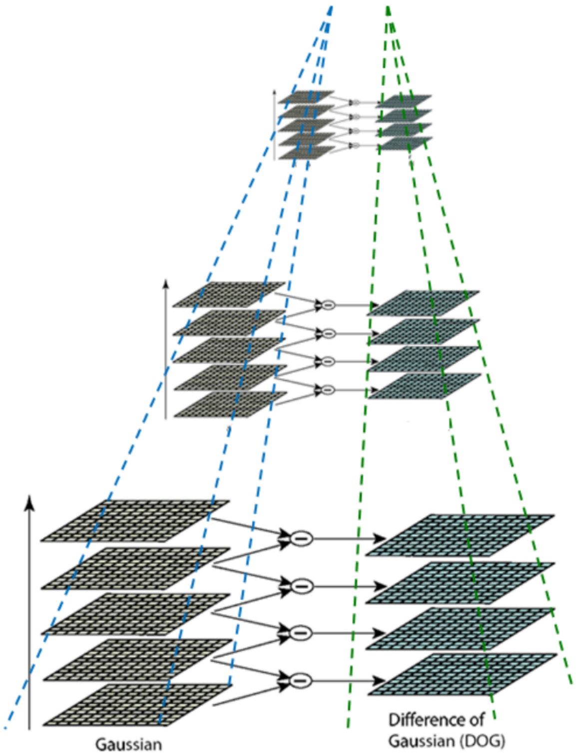

图 3: DoG 金字塔

如图3，DoG 金字塔由同一组的相邻两个图像相减得到，所以若每一组有 S 层，那么 DoG 金字塔的每一组有 $S-1$ 层。

1.2.6 DoG 空间的极值检测

如图4，构建完 DoG 金字塔之后，SIFT 算法的下一步是寻找 DoG 空间的极值点，若某个点 (图4里的黑叉) 比其同一层的相邻的 8 个点和其相邻层同样位置的 18 个点都大或者都小，那么可以认定其为极值点。

那么最上层和最下层是不能取出极值点的，为了弥补这个问题，每组图像我们额外用高斯模糊生成 3 幅图像，那么最后高斯金字塔的每组有 $S+3$ 层，DoG 金字塔每组有 $S+2$ 层，所以能够对 S 层取极值。

1.2.7 DoG 空间极值检测总结

首先确定高斯金字塔的组数 O 和每组的层数 $S+3$，并确定 σ_0，确定 $k=2^{\frac{1}{S}}$

然后构建第 0 组，即对原始图像的宽和高都增加一倍得到 I_0，用 $G(x,y,\sigma_0)$ 对 I_0 进行卷积，得到第 0 层，$I0*G(x,y,k\sigma_0)$ 得到第 1 层，以此类推。

然后构建第 1 组，对 I_0 降采样得到 I_1，用 $G(x,y,2\sigma_0)$ 对 I_1 进行卷积，得到第 0 层，$I0*G(x,y,2k\sigma_0)$ 得到第 1 层，以此类推。

$\vdots$

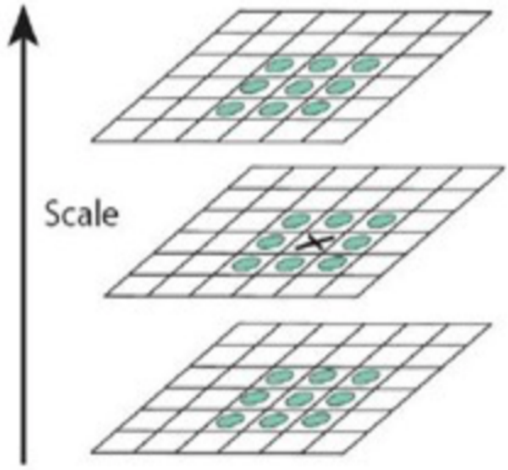

图 4: DoG 空间的极值检测

构建好高斯金字塔之后，同一组的相邻层进行差分，得到 DoG 金字塔，并在 DoG 空间中寻找极值点。

1.3 筛选极值点

1.3.1 剔除对比度低的特征点

对于某个特征点 x，假设其偏移 Δx，$D(x)$ 为高斯差分图像 (Dog 中的一层)，令 $D(x)$ 在该特征点处进行泰勒展开得到，

$$D(\Delta x + x) = D + \frac{\partial D^T}{\partial x}\Delta x + \frac{1}{2}\Delta x^T \frac{\partial^2 D}{\partial x^2}\Delta x \tag{4}$$

为了求得 $D(x)$ 的极值，对上式求偏导，并令偏导为 0，得到

$$\Delta x^* = -\left(\frac{\partial D}{\partial x^2}\right)^{-1} \frac{\partial D(x)}{\partial x} \tag{5}$$

回代至 $D(x)$ 的表示式中

$$D(x + \Delta x^*) = D + \frac{1}{2}\frac{\partial D^T}{\partial x}\Delta x^* \tag{6}$$

若 $|D(x + \Delta x^*)| > threshold$，保留该点，否则剔除掉。

1.3.2 剔除边缘响应点

由之前的 harris 角点检测一章我们知道，图像的 $D(x)$ 的二阶导数矩阵 Hessian 矩阵 H

$$H = \begin{bmatrix} D_{xx} & D_{yx} \\ D_{xy} & D_{yy} \end{bmatrix} \tag{7}$$

的两个特征值可以提现当前是平坦区域还是边缘。

设 H 的大的特征值为 α，小的特征值为 β，那么

$$\begin{aligned} Tr(H) &= \alpha + \beta \\ Det(H) &= \alpha \cdot \beta \end{aligned} \tag{8}$$

令 $\gamma = \frac{\alpha}{\beta}$ 则

$$\frac{Tr(H)^2}{Det(H)} = \frac{(\gamma+1)^2}{\gamma} \tag{9}$$

而若 γ 很大的时候，说明该特征点处于边缘（在 Harris 角点一章解释过）而 $\frac{(\gamma+1)^2}{\gamma}$ 当且仅当 $\gamma = 1$ 时去到最小值，且随之 γ 的增大递增，所以当

$$\frac{Tr(H)^2}{Det(H)} > \frac{(T_\gamma+1)^2}{T_\gamma} \tag{10}$$

时，剔除该特征点，否则保留，作者取 $T_\gamma = 10$。

1.4 计算特征点的主方向

对于筛选之后的特征点我们知道他的 x, y, σ，所以我们也可以知道该特征点所在的图像 $L(x, y) = G(x, y, \sigma) * I(x, y)$，以该特征点为中心，计算 $3 \times 1.5\sigma$ 为半径区域的像素的幅角和幅值，并绘制梯度方向直方图（这个在 HOG 特征中有详细的介绍）。这里方向统计的范围为 $[0, 360^o]$，分为 10 个或者 8 个 bin，绘制好梯度方向直方图之后，使用高斯函数对直方图进行平滑，为了得到更精确的角度描述，可以用抛物线插值法对直方图插值，最后取直方图的峰值为特征点的主方向，若存在相当于主峰值 80% 能量的柱值时，则认为这个方向时特征点的辅助方向。

到这里对每个特征点我们得到了 (x, y, σ, θ) 即位置尺度和方向的信息。

1.5 生成特征描述

首先，以特征点为中心，在附近邻域内，将坐标轴旋转 θ 角度，即使坐标轴和特征点主方向重合，数学表示即

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \tag{11}$$

于是我们对旋转之后以像素为中心取 8×8 的窗口。如图5，每个小格子代表一个像素，里面的箭头代表梯度，将窗口分为四块，统计每块的梯度方向直方图，每个直方图都取 8 个 bin，之后对每个直方图归一化，然后把所有直方图表示的向量堆砌起来，作为 SIFT 描述子。

当然，除了以像素为中心取 8×8 的窗口之外，还可以取 16×16 的窗口，将窗口分为 4×4 块，然后把每块统计得到的梯度方向直方图归一化，最后拼成一个向量。

这个特征描述子和 HOG 特征非常相似，这里的块就类似于 HOG 特征的 Cell，窗口类似于 HOG 特征的 block。但是也有不同，HOG 特征的归一化是在 block 内归一化（即我们先把窗口中的所有梯度直方图拼成向量之后再做归一化），除此之外，HOG 特征的 block 会有交叠。

至此，我们得到了 SIFT 算子的四个信息 (x, y, σ, θ) 以及特征描述子。之后便可以利用这个特征描述子去匹配特征了。

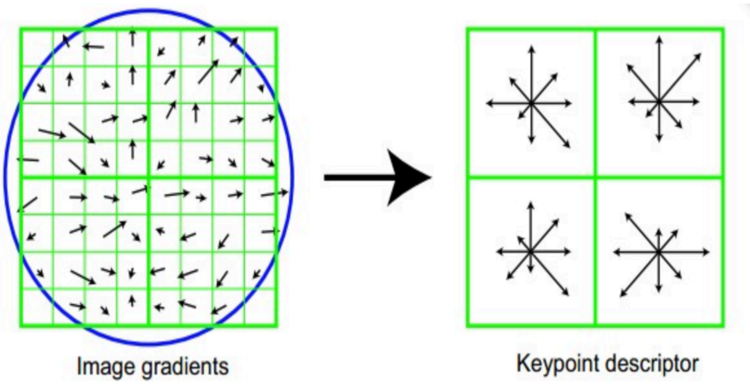

图 5: 特征描述子

1.6 总结

从上述步骤中我们可以看出为什么 SIFT 特征会对尺度、方向、光照不变。我们在不同的高斯尺度和像素尺度上都寻找了特征点，所以它对尺度是不变的。我们用梯度方向直方图求得了 SIFT 算子的方向，然后在求 SIFT 描述子的时候，将图像的坐标轴旋转至其主方向上，所以即使物体旋转了，我们获得的 SIFT 描述子仍然是不变的，除此之外剔除了边缘的特征点也有助于对方向不变的特性。最终我们计算特征描述子时，有归一化这一操作，这就保证了 SIFT 算子的光照不变性。

最后得到的特征描述子包含了很多信息，有助于我们在大量的特征中进行快速准确的匹配。又由于最后我们得到的描述子是一个向量，他也能够很自然的与其他特征向量进行联合。

SIFT 有很多优点，是一种非常稳定的特征点，可以非常放心的使用。