

尺度不变特征变换匹配算法

Scale Invariant Feature Transform (SIFT)

- SIFT提出的目的和意义



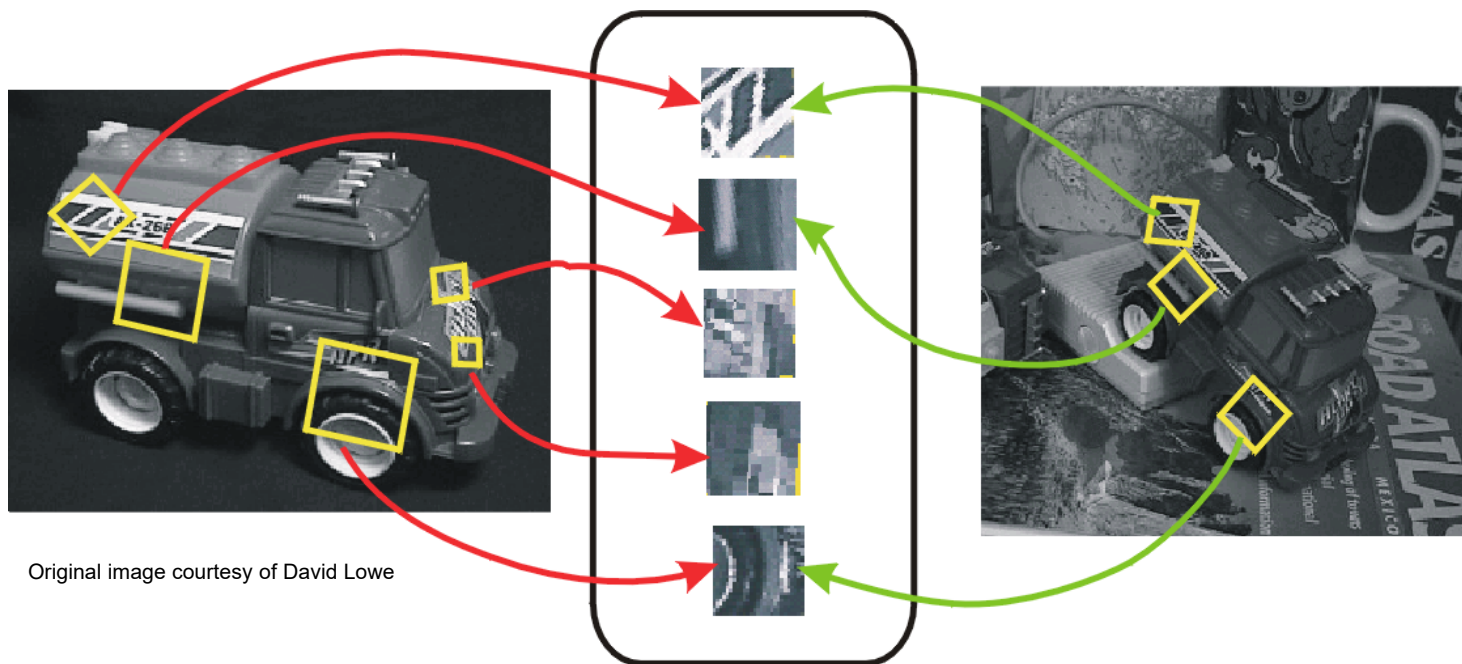
David G. Lowe
Computer Science Department
2366 Main Mall
University of British Columbia
Vancouver, B.C., V6T 1Z4,
Canada
E-mail: lowe@cs.ubc.ca

- 1999年British Columbia大学大卫·劳伊 (David G. Lowe) 教授总结了现有的基于不变量技术的特征检测方法，并正式提出了一种基于尺度空间的、对图像缩放、旋转甚至仿射变换保持不变性的图像局部特征描述算子—SIFT（尺度不变特征变换），这种算法在2004年被加以完善。

SIFT简介

SIFT

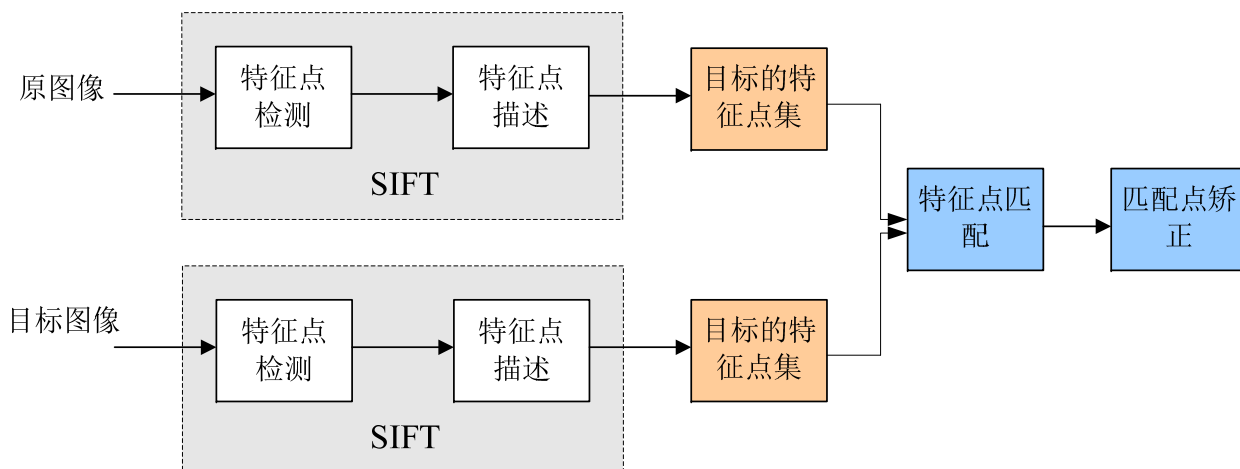
Scale Invariant Feature Transform



- 将一幅图像映射（变换）为一个局部特征向量集；特征向量具有平移、缩放、旋转不变性，同时对光照变化、仿射及投影变换也有一定不变性。

- SIFT算法实现步骤简述

SIFT算法的实质可以归为在不同尺度空间上查找特征点（关键点）的问题。



SIFT算法实现物体识别主要有三大工序，1、提取关键点；2、对关键点附加详细的信息（局部特征）也就是所谓的描述器；3、通过两方特征点（附带特征向量的关键点）的两两比较找出相互匹配的若干对特征点，也就建立了景物间的对应关系。

1. 哪些点是SIFT中要查找的关键点（特征点）？

这些点是一些十分突出的点不会因光照条件的改变而消失，比如角点、边缘点、暗区域的亮点以及亮区域的暗点，既然两幅图像中有相同的景物，那么使用某种方法分别提取各自的稳定点，这些点之间会有相互对应的匹配点。

所谓关键点，就是在不同尺度空间的图像下检测出的具有方向信息的局部极值点。

根据归纳，我们可以看出特征点具有的三个特征：

尺度

方向

大小

2. 什么是尺度空间（scale space）？

我们要精确表示的物体都是通过一定的尺度来反映的。现实世界的物体也总是通过不同尺度的观察而得到不同的变化。

尺度空间理论最早在1962年提出，其主要思想是通过对原始图像进行尺度变换，获得图像多尺度下的尺度空间表示序列，对这些序列进行尺度空间主轮廓的提取，并以该主轮廓作为一种特征向量，实现边缘、角点检测 and 不同分辨率上的特征提取等。

尺度空间中各尺度图像的模糊程度逐渐变大，能够模拟人在距离目标由近到远时目标在视网膜上的形成过程。

尺度越大图像越模糊。

研究表明，高斯核是唯一可以产生多尺度空间的核，一个图像的尺度空间 $L(x, y, \sigma)$ ，定义为原始图像 $I(x, y)$ 与一个可变尺度的2维高斯函数 $G(x, y, \sigma)$ 卷积运算。

高斯函数

$$G(x_i, y_i, \sigma) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(x-x_i)^2 + (y-y_i)^2}{2\sigma^2}\right)$$

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

尺度是自然存在的，不是人为创造的！高斯卷积只是表现尺度空间的一种形式...

3. 高斯模糊

高斯模糊是在图像处理软件中广泛使用的处理效果，通常用它来减小图像噪声以及降低细节层次。这种模糊技术生成的图像的视觉效果是好像经过一个半透明的屏幕观察图像。



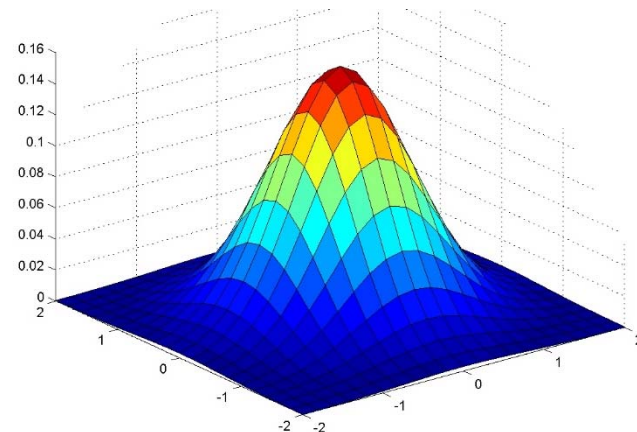
关键点检测的相关概念

SIFT

Scale Invariant Feature Transform

$$G(r) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{r^2}{2\sigma^2}\right)$$

r 为模糊半径, $r = \sqrt{x^2 + y^2}$

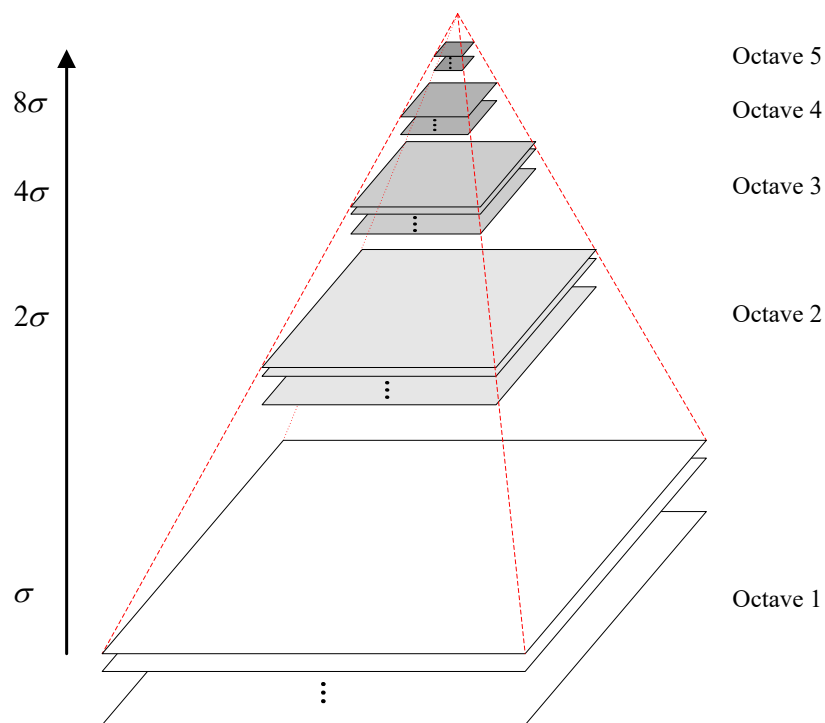


在减小图像尺寸の場合经常使用高斯模糊。在进行欠采样的时，通常在采样之前对图像进行低通滤波处理。这样就可以保证在采样图像中不会出现虚假的高频信息。

4. 高斯金字塔

- 高斯金字塔的构建过程可分为两步：
 - 对图像做高斯平滑；
 - 对图像做降采样。

为了让尺度体现其连续性，在简单下采样的基础上加上了高斯滤波。一幅图像可以产生几组（octave）图像，一组图像包括几层（interval）图像。



关键点检测的相关概念

SIFT

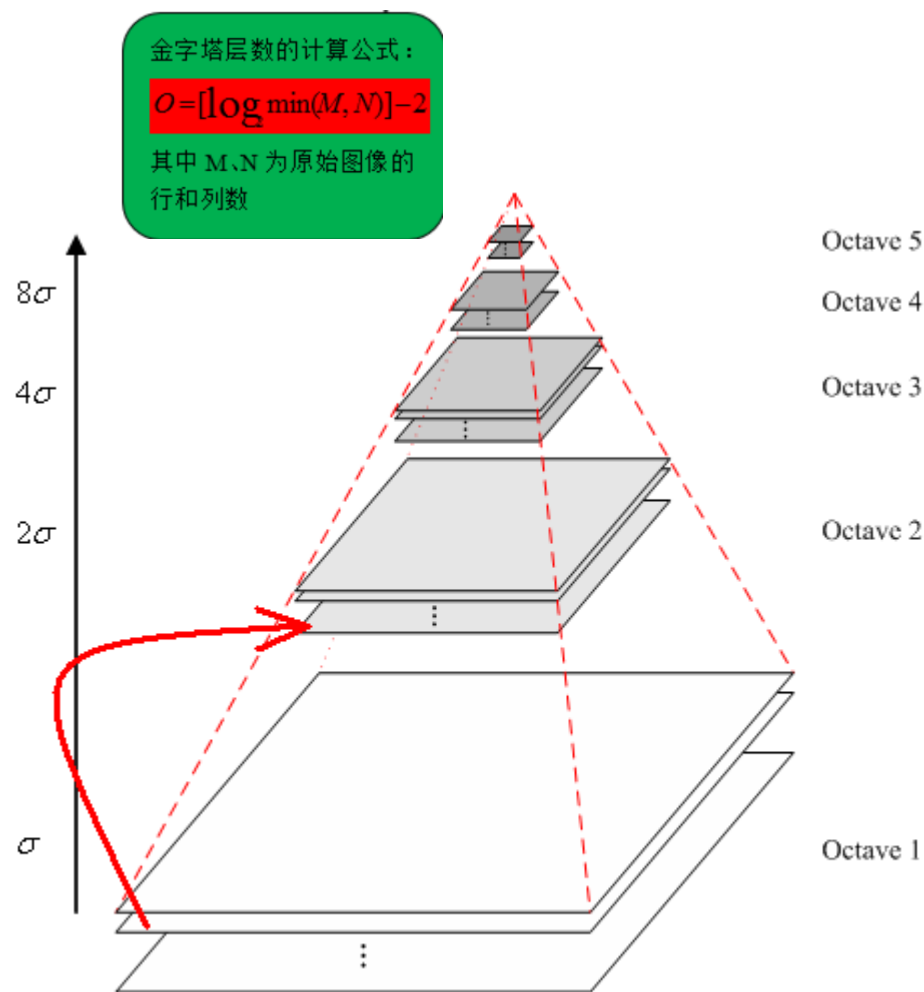
Scale Invariant Feature Transform

- 上一组图像的底层是由前一组图像底层图像进行因子为**2**的隔点采样生成的。这样可以保持尺度的连续性。

每一组 (Octave) 有 6 层 (Interval) 尺寸相同但模糊系数 (σ) 不同的采样图像组成，其变化公式：

$$\sigma(o, r) = \sigma_0 2^{o + \frac{r}{s}} \quad o \in [0, \dots, O-1], r \in [0, \dots, s+2]$$

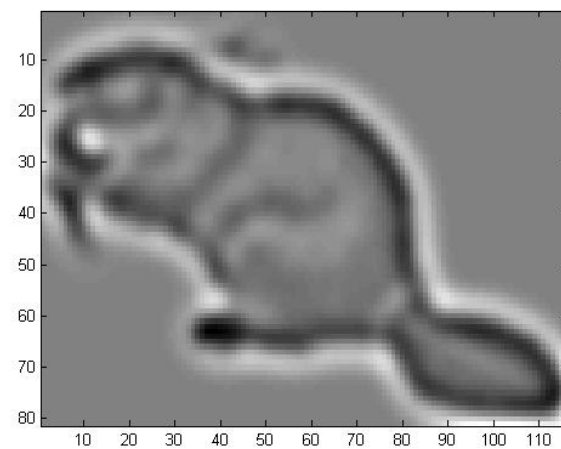
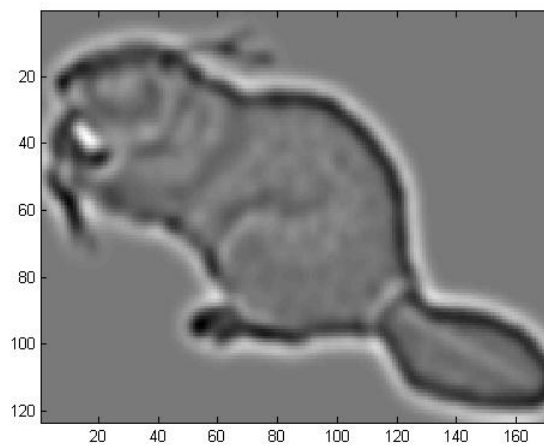
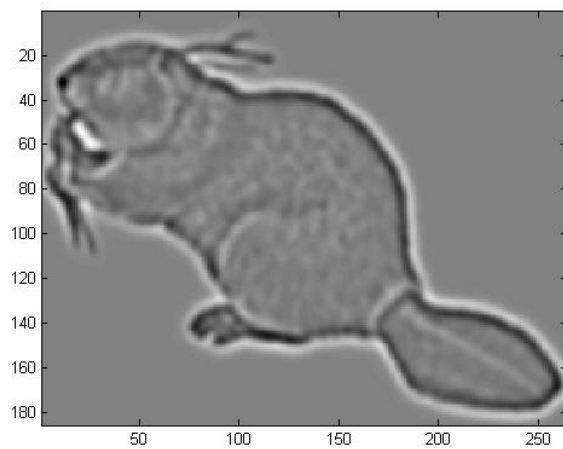
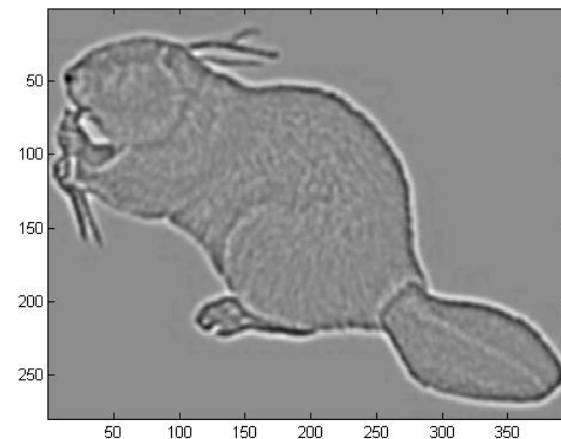
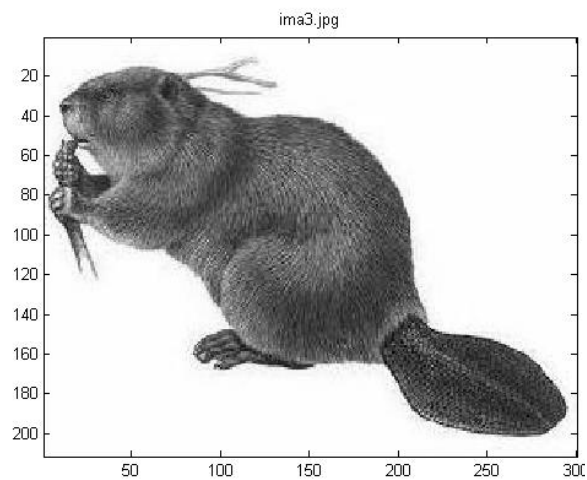
其中 o 为组索引序号； r 为层索引序号； s 为高斯差分金字塔每组层数, $s=3$ ； O 为金字塔组数； σ_0 为高斯模糊初始值，Lowe 建议 $\sigma_0 = 1.6$ ，考虑相机已对图像进行 $\sigma=0.5$ 的模糊，故实际的 $\sigma_0 = \sqrt{1.6*1.6-0.5*0.5} = 1.52$



关键点检测——DOG

SIFT

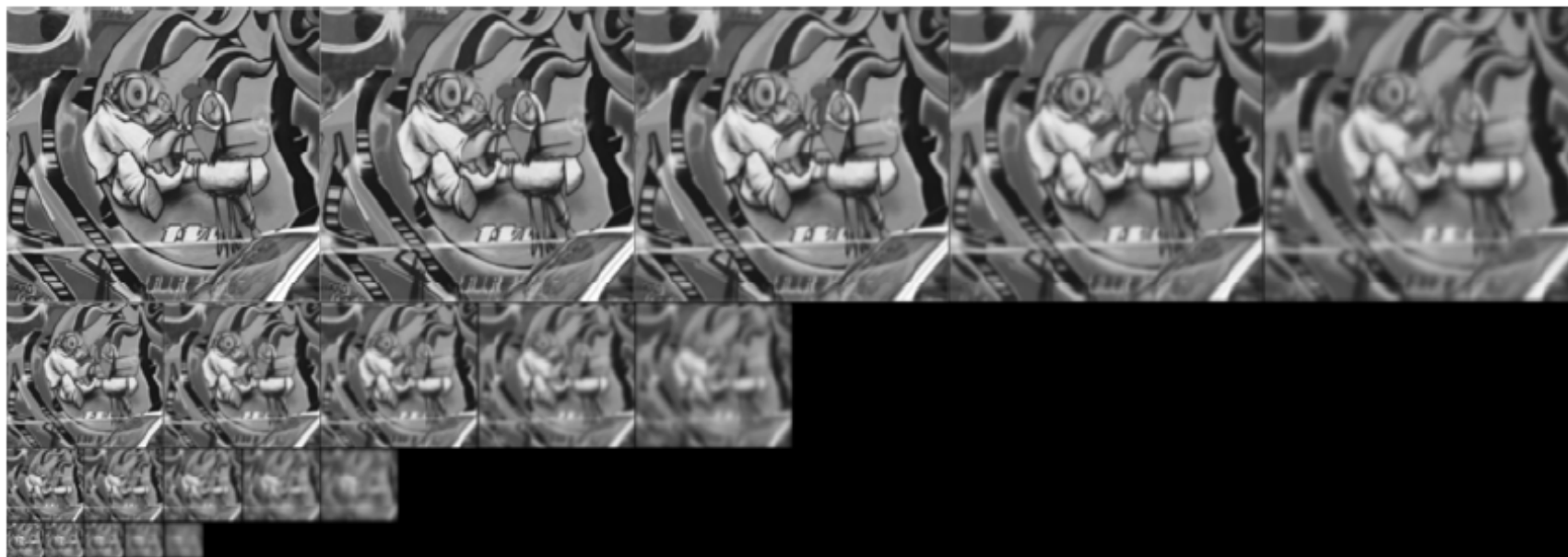
Scale Invariant Feature Transform



高斯尺度空间

SIFT

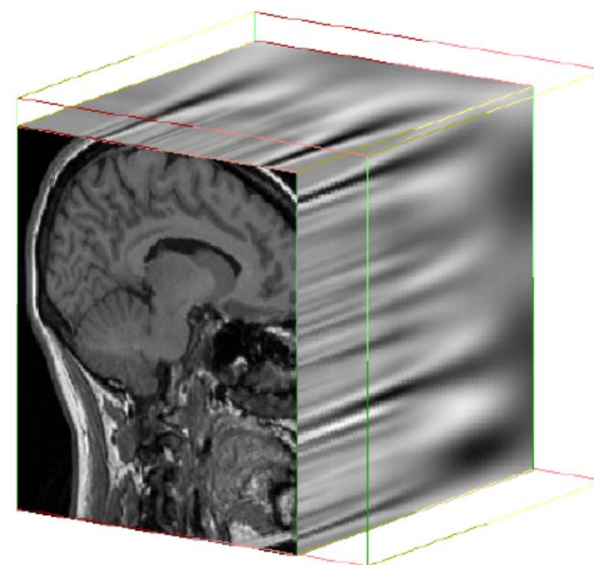
Scale Invariant Feature Transform



DoG (Difference of Gaussian) 函数

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

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

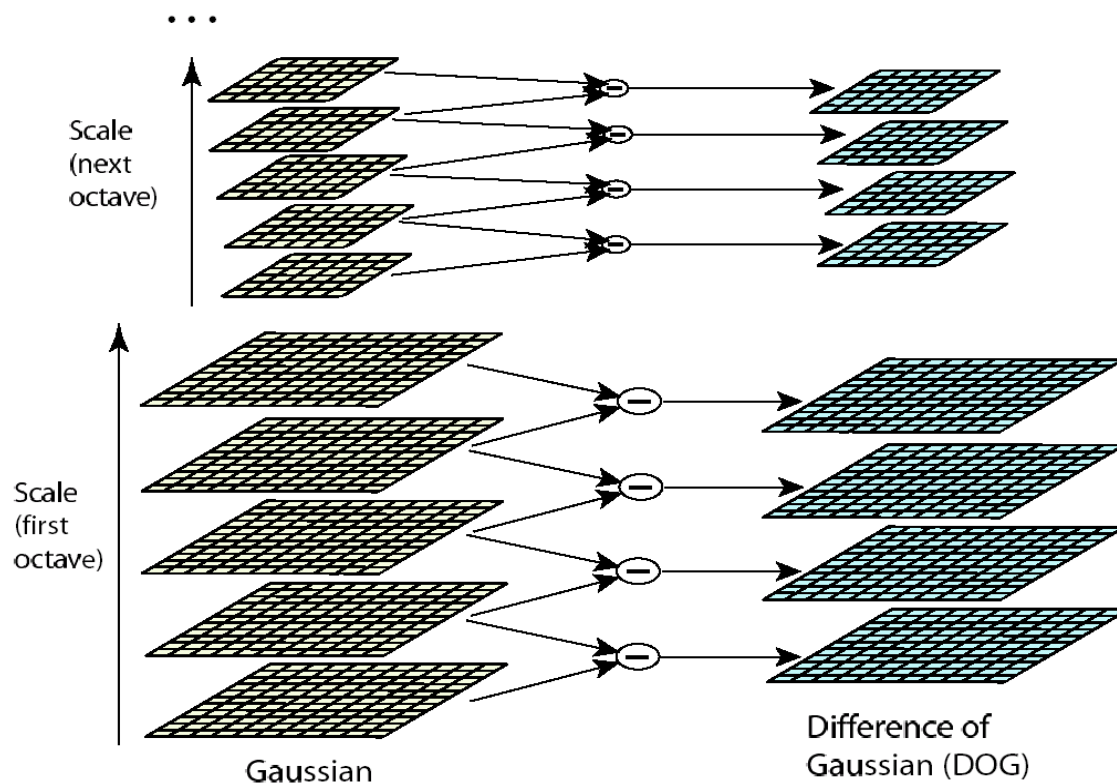


DoG在计算上只需相邻尺度高斯平滑后图像相减，因此简化了计算！

DoG高斯差分金字塔

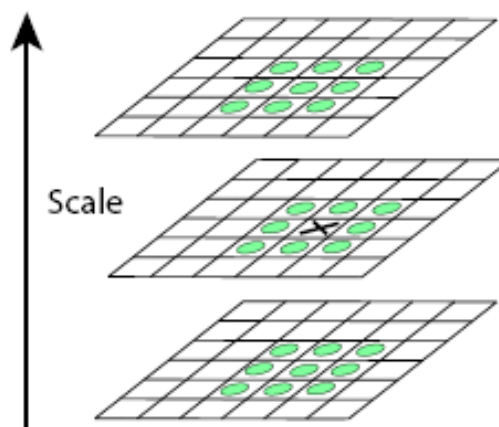
对应DOG算子，我们要构建DOG金字塔

我们可以通过高斯差分图像看出图像上的像素值变化情况。（如果没有变化，也就没有特征。特征必须是变化尽可能多的点。）DOG图像描绘的是目标的轮廓。



- DoG的局部极值点

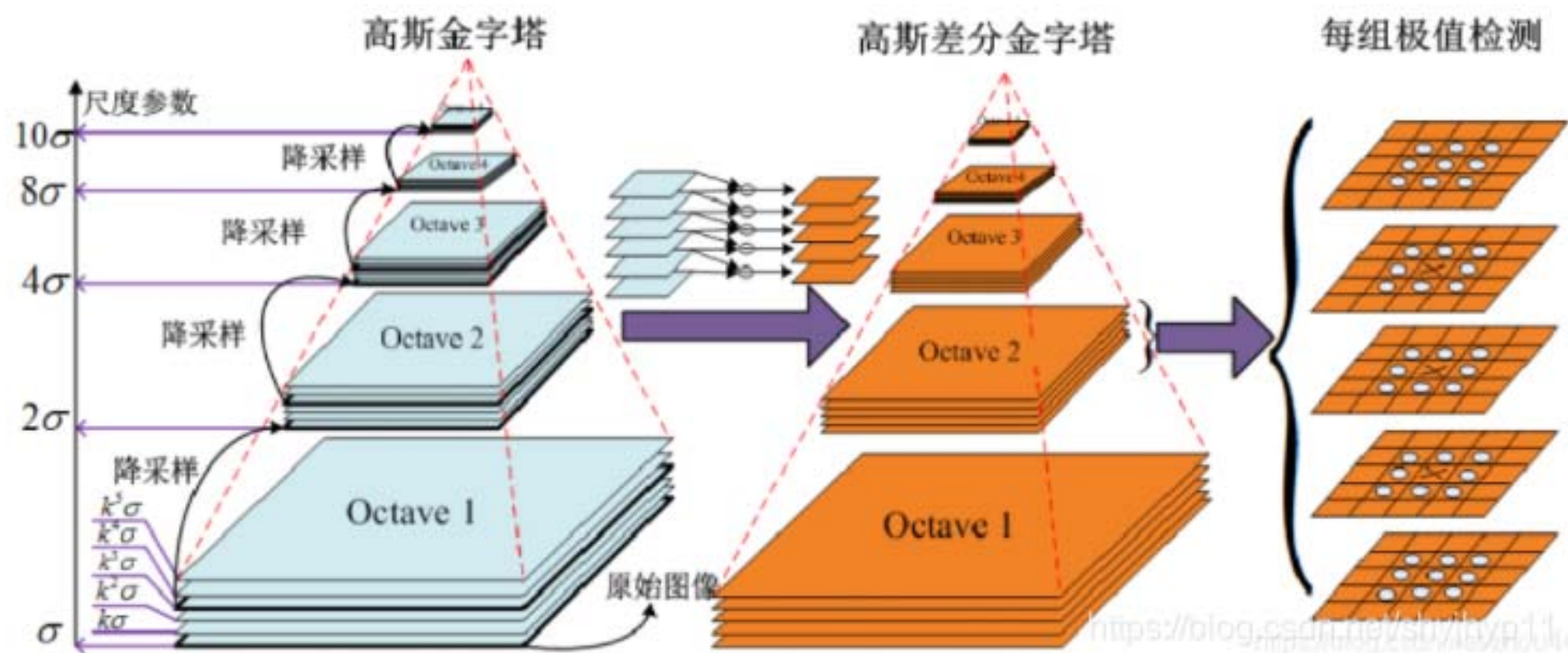
关键点是由DOG空间的局部极值点组成的。为了寻找DoG函数的极值点，每一个像素点要和它所有的相邻点比较，看其是否比它的图像域和尺度域的相邻点大或者小。



中间的检测点和它同尺度的8个相邻点和上下相邻尺度对应的 9×2 个点共26个点比较，以确保在尺度空间和二维图像空间都检测到极值点。

SIFT

Scale Invariant Feature Transform



- 关键点精确定位

由于DoG值对噪声和边缘较敏感, 因此, 在上面DoG尺度空间中检测到局部极值点还要经过进一步的检验才能精确定位为特征点。

为了提高关键点的稳定性, 需要对尺度空间DoG函数进行曲线拟合。
利用DoG函数在尺度空间的Taylor展开式:

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

其极值点 $\hat{X} = (x, y, \sigma)^T$

在计算过程中，分别对图像的行、列及尺度三个量进行了修正，其修正结果如下：

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

$$\text{求解得 } \hat{X} = -\frac{\partial D^T}{\partial X} \left(\frac{\partial^2 D}{\partial X^2} \right)^{-1}$$

\hat{X} 为修正值

在Lowe的程序中，对坐标进行了五次修正。

$$\text{将修正后的结果代入式 } D(X) = D + \frac{\partial D^T}{\partial X} X + \frac{1}{2} X^T \frac{\partial^2 D}{\partial X^2} X$$

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

上式去除那些对比度较低的不稳定极值点。Lowe的试验显示，

所有

2023/6/16

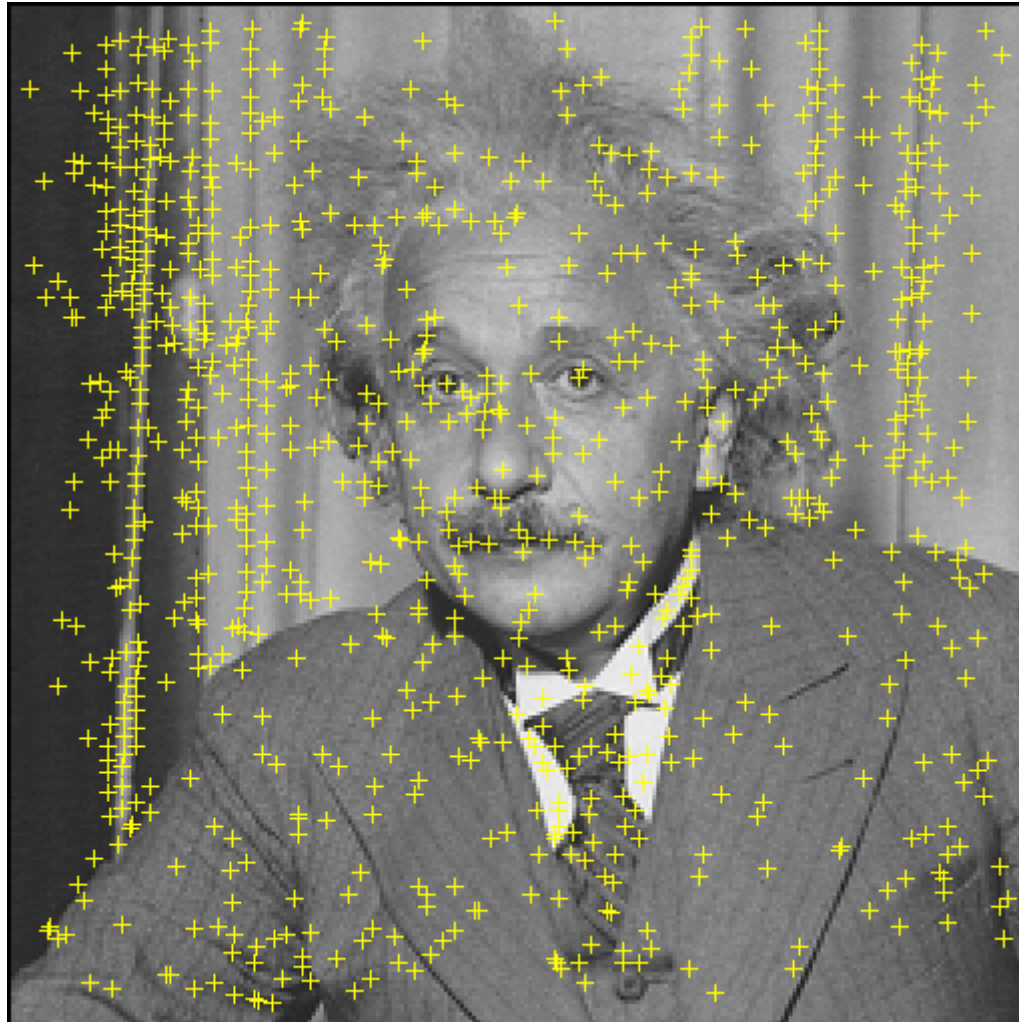
[0, 1]) 。

取值小于0.04的极值点均可抛弃（像素灰度值范围

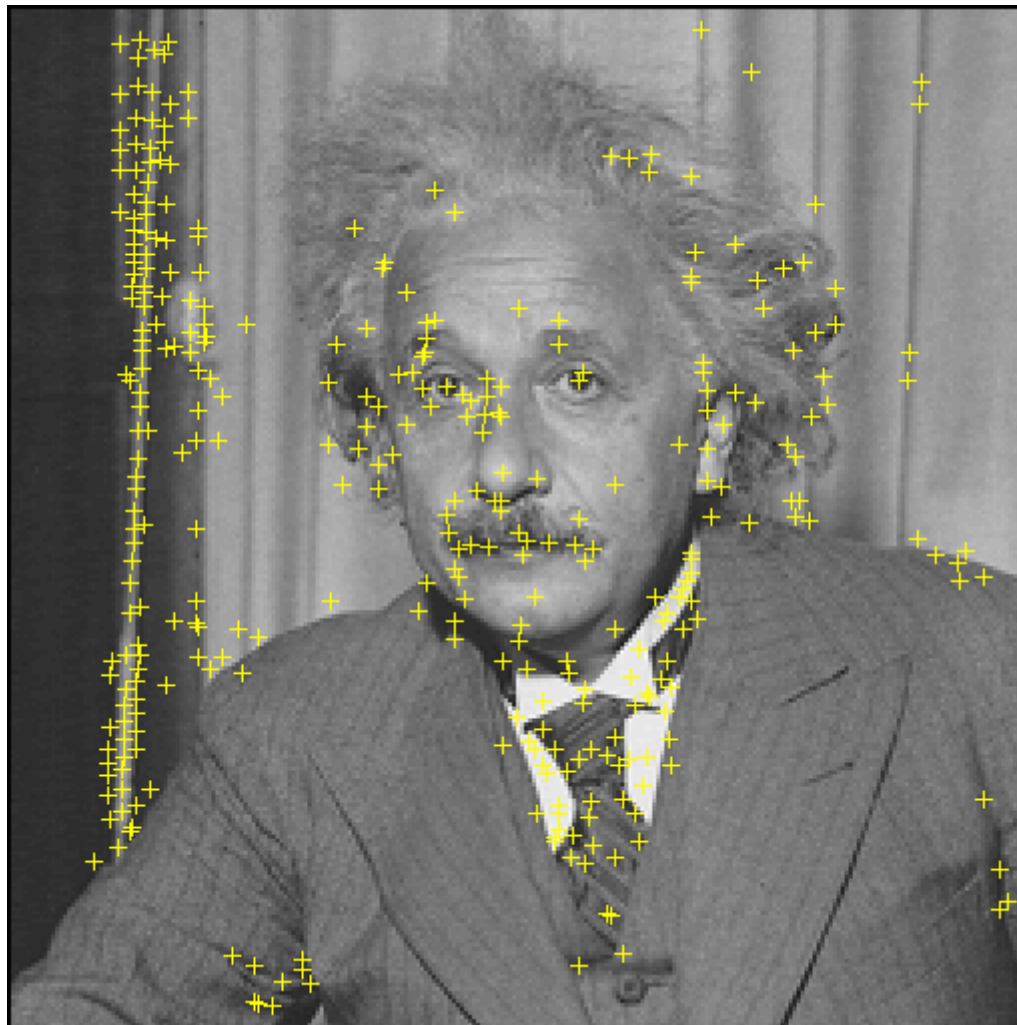
SIFT

Scale Invariant Feature Transform

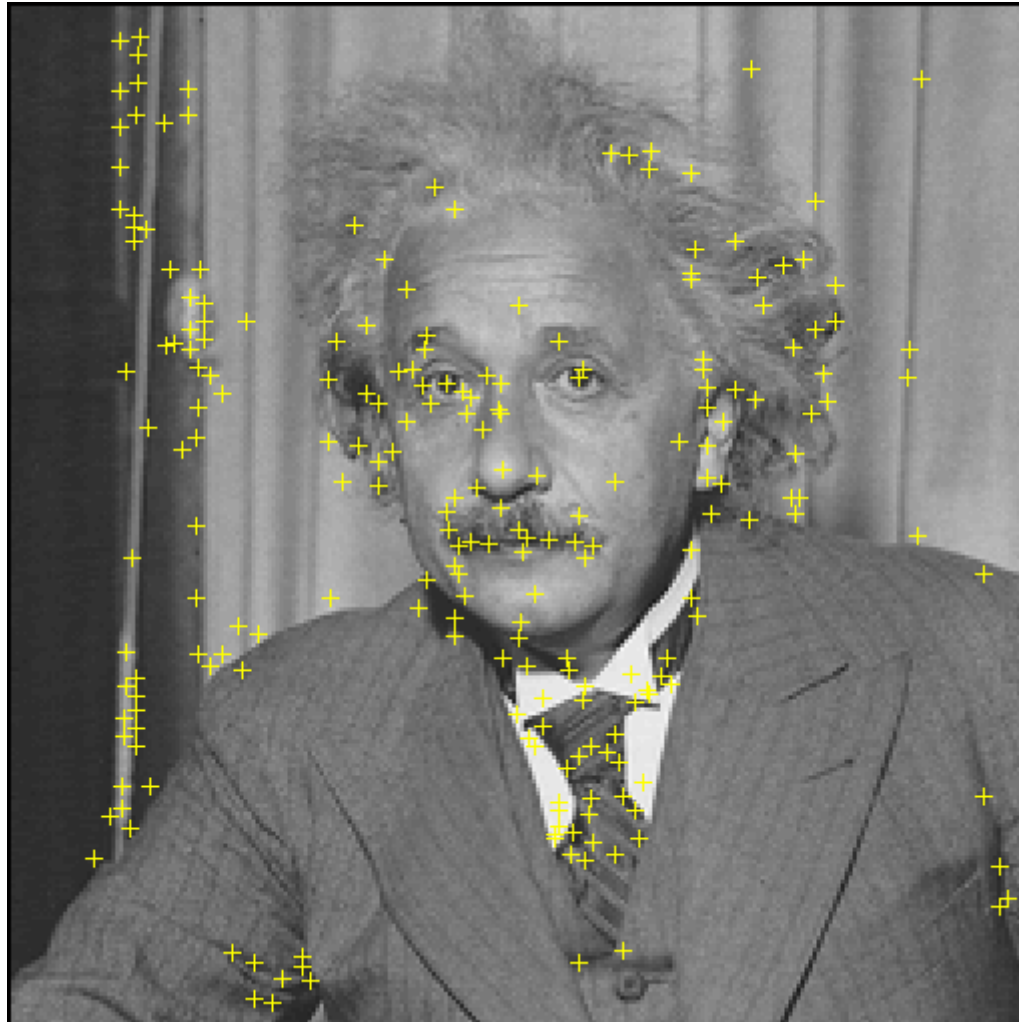
没有进行任何处理的关键点



去除了低对比度之后的点



去除边界相应的点



通过尺度不变性求极值点，可以使其具有缩放不变的性质，利用关键点邻域像素的梯度方向分布特性，我们可以为每个关键点指定方向参数方向，**从而使描述子对图像旋转具有不变性。**

我们通过求每个极值点的梯度来为极值点赋予方向。

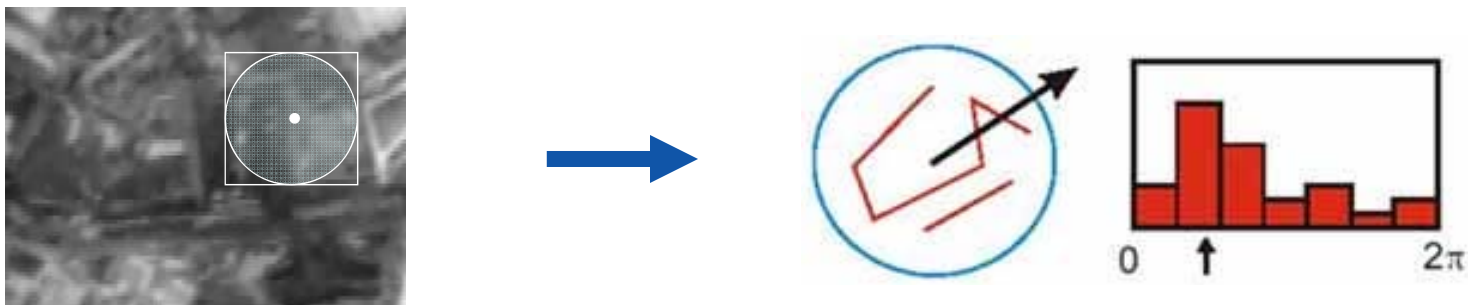
- 像素点的梯度表示

$$\text{grad}I(x, y) = \left(\frac{\partial I}{\partial x}, \frac{\partial I}{\partial y} \right)$$

梯度幅值:
$$m(x, y) = \sqrt{\left(L(x+1, y) - L(x-1, y) \right)^2 + \left(L(x, y+1) - L(x, y-1) \right)^2}$$

梯度方向:
$$\theta(x, y) = \tan^{-1} \left[\frac{L(x, y+1) - L(x, y-1)}{L(x+1, y) - L(x-1, y)} \right]$$

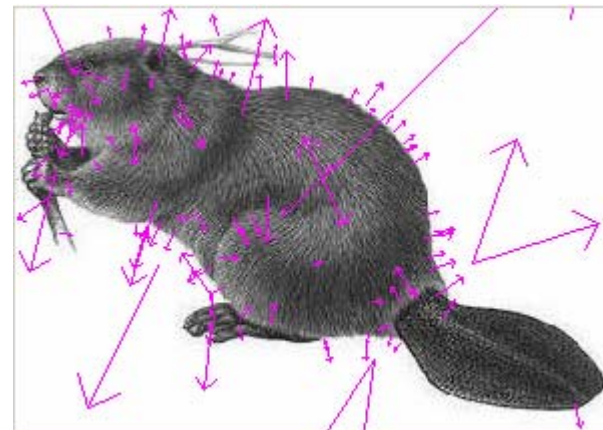
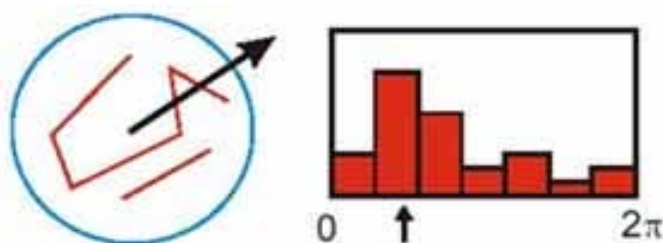
- 方向直方图的生成



确定关键点方向采用梯度直方图统计法，统计以关键点为原点，一定区域内的图像像素点对关键点方向生成所作的贡献。

1. 直方图以每10度方向为一个柱，共36个柱，柱所代表的方向为像素点梯度方向，柱的长短代表了梯度幅值。
2. 根据Lowe的建议，直方图统计半径采用 $3 * 1.5 * \sigma$ 。

关键点的主方向与辅方向



- 关键点主方向：极值点周围区域梯度直方图的主峰值，也是特征点方向
- 关键点辅方向：在梯度方向直方图中，当存在另一个相当于主峰值80%能量的峰值时，则将这个方向认为是该关键点的辅方向。

这可以增强匹配的鲁棒性，Lowe的论文指出大概有15%关键点具有多方向，但这些点对匹配的稳定性至为关键。

● 方向分配实现步骤

1. 确定计算关键点直方图的高斯函数权重函数参数 ；
2. 生成含有36柱的方向直方图，梯度直方图范围0~360度，其中每10度一个柱。由半径为图像区域生成；
3. 对方向直方图进行两次平滑；
4. 求取关键点方向（可能是多个方向）；
5. 对方向直方图的Taylor展开式进行二次曲线拟合，精确关键点方向；

图像的关键点已检测完毕，每个关键点有三个信息：位置、尺度、方向；同时也就使关键点具备平移、缩放、和旋转不变性。

[Back](#)

- 描述的目的

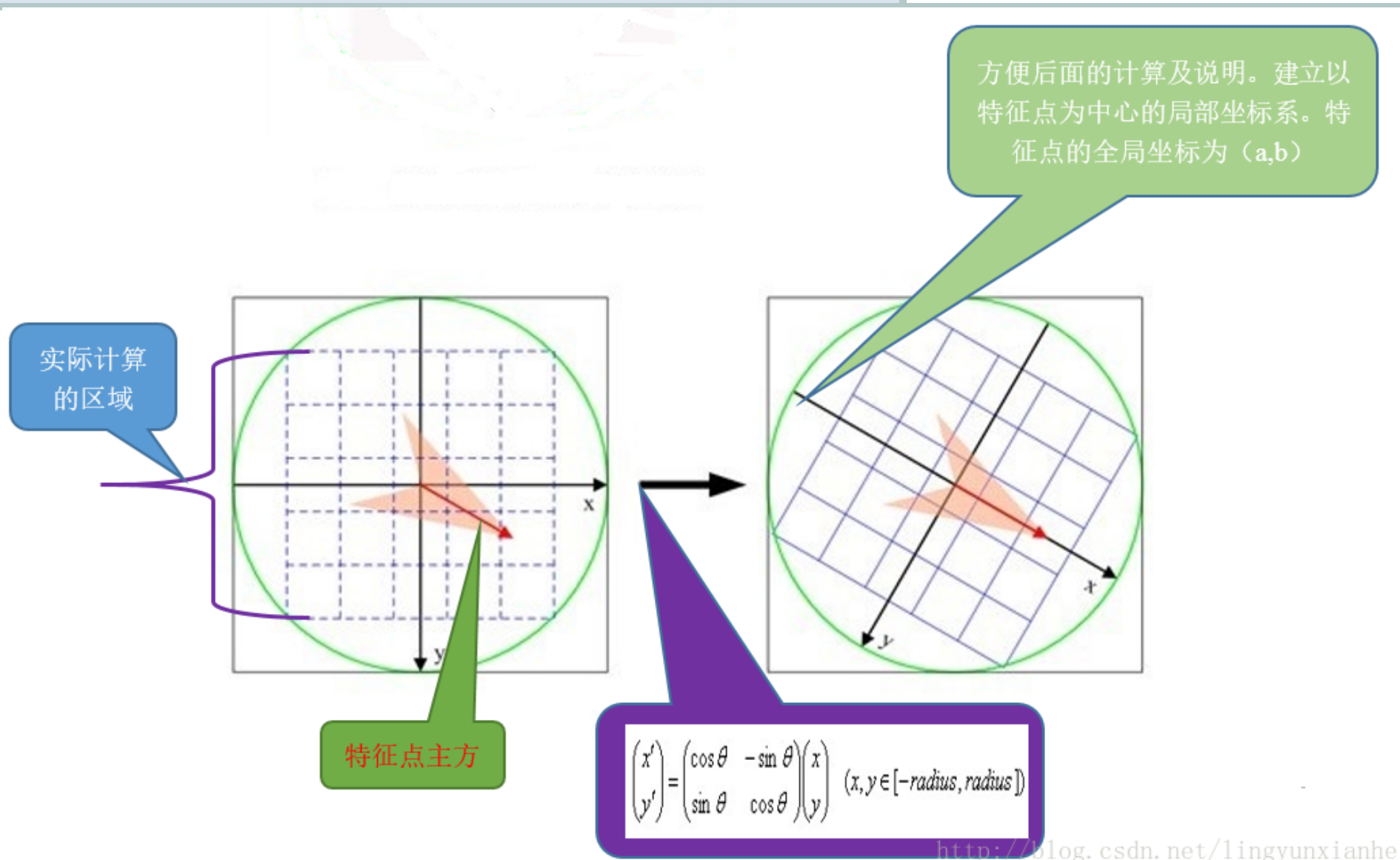
描述的目的是在关键点计算后，用一组向量将这个关键点描述出来，这个描述子不但包括关键点，也包括关键点周围对其有贡献的像素点。用来作为目标匹配的依据，也可使关键点具有更多的不变特性，如光照变化、3D视点变化等。

- 描述的思路

通过对关键点周围图像区域分块，计算块内梯度直方图，生成具有独特性的向量，这个向量是该区域图像信息的一种抽象，具有唯一性。

SIFT

Scale Invariant Feature Transform



关键点描述

SIFT

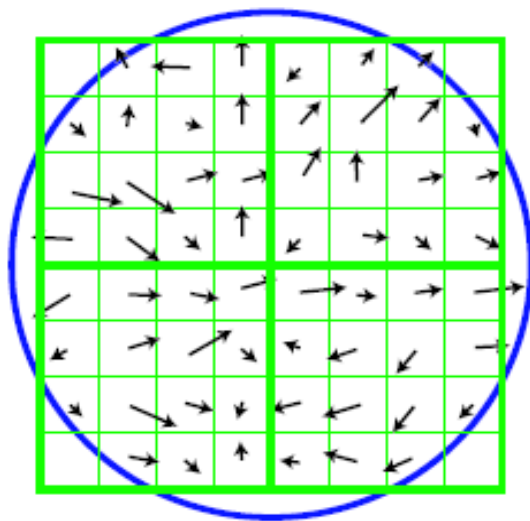
Scale Invariant Feature Transform

8×8 单元。每一个小格代表了特征点邻域所在的尺度空间的一个像素，箭头方向代表了像素梯度方向，箭头长度代表该像素的幅值。

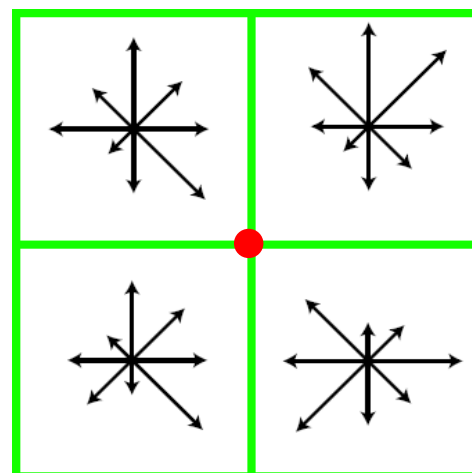
在 2×2 的窗口内计算8个方向的梯度方向直方图。

一个描述子由 $2 \times 2 \times 8$ 维向量表征，也即是 2×2 个8方向的方向直方图组成。

总长度 $4 \times 32 = 128$

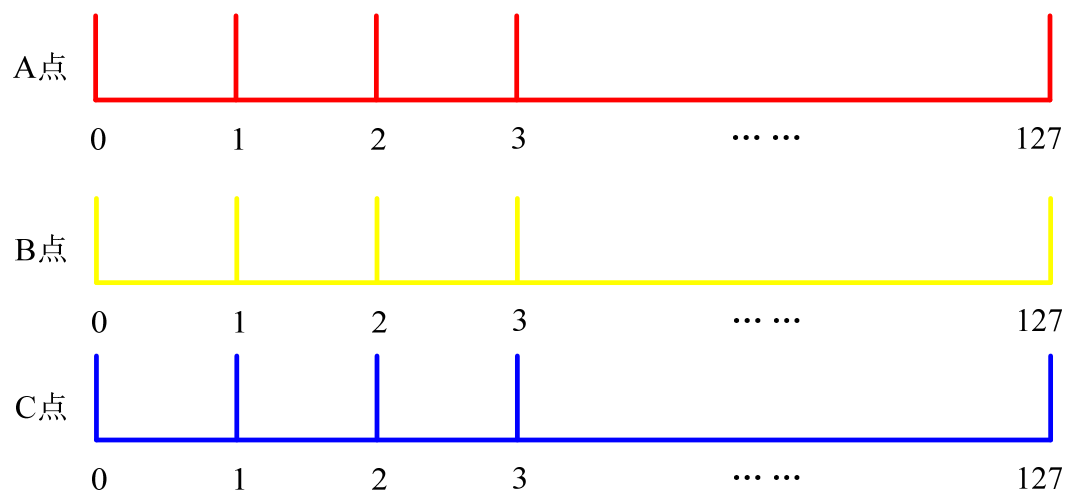


关键点周围区域图像梯度

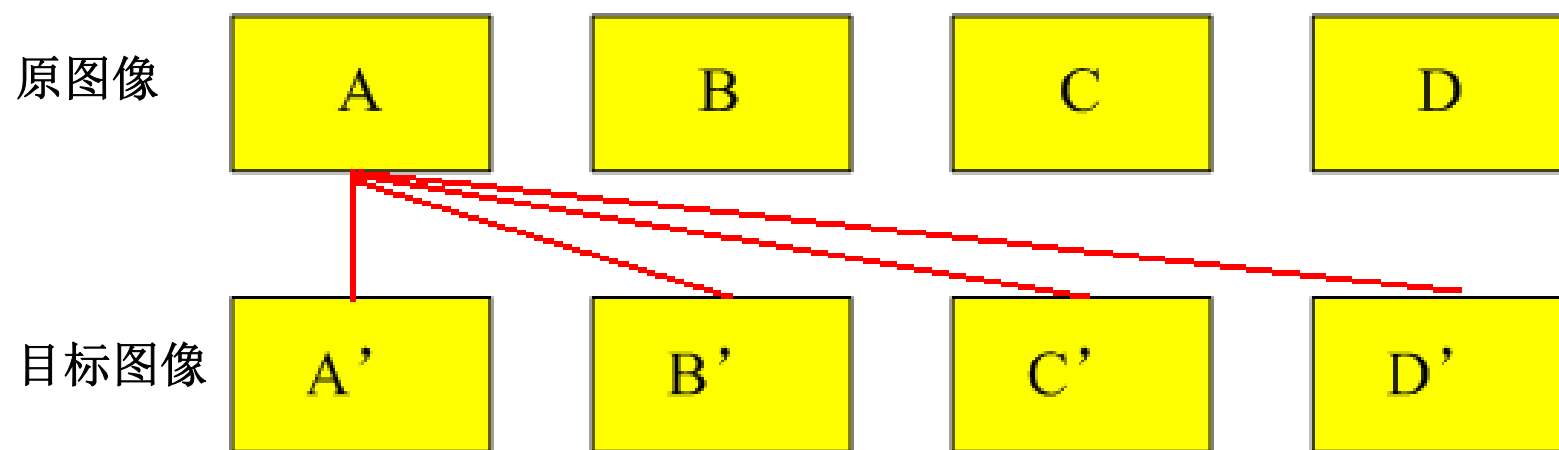


关键点描述子

分别对模板图（参考图，reference image）和实时图（观测图，observation image）建立关键点描述子集合。目标的识别是通过两点集内关键点描述子的比对来完成。具有128维的关键点描述子的相似性度量采用欧式距离。



穷举匹配



- 模板图中关键点描述子: $R_i = (r_{i1}, r_{i2}, \dots, r_{i128})$
- 实时图中关键点描述子: $S_i = (s_{i1}, s_{i2}, \dots, s_{i128})$
- 任意两描述子相似性度量: $d(R_i, S_i) = \sqrt{\sum_{j=1}^{128} (r_{ij} - s_{ij})^2}$

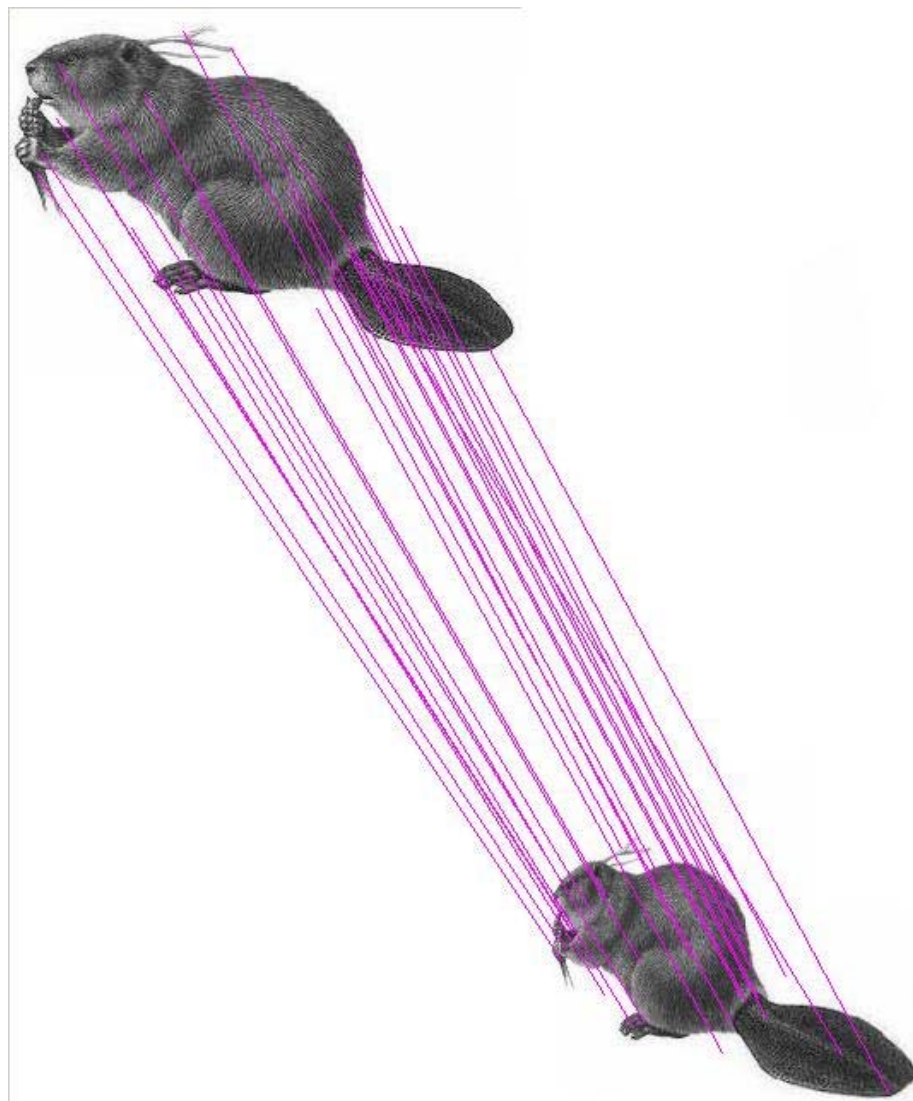
要得到配对的关键点描述子, $d(R_i, S_j)$ 需满足:

$$\frac{\text{实时图中距离 } R_i \text{ 最近的点 } S_j}{\text{实时图中距离 } R_i \text{ 的次最近点 } S_p} < Threshold$$

关键点匹配

SIFT

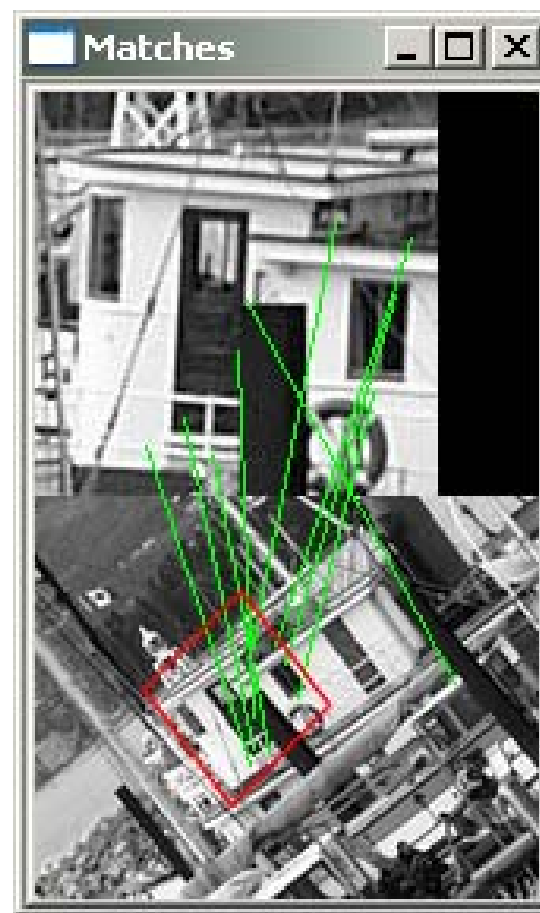
Scale Invariant Feature Transform



[Back](#)

- 关键点匹配并不能标志着算法的结束，因为在匹配的过程中存在着大量的错配点。

图中交叉的绿线为错配点



RANSAC简介

RANSAC (Random Sample Consensus, 随机抽样一致)
是一种鲁棒性的参数估计方法。

RANSAC实质上就是一个反复测试、不断迭代的过程。

RANSAC的基本思想:

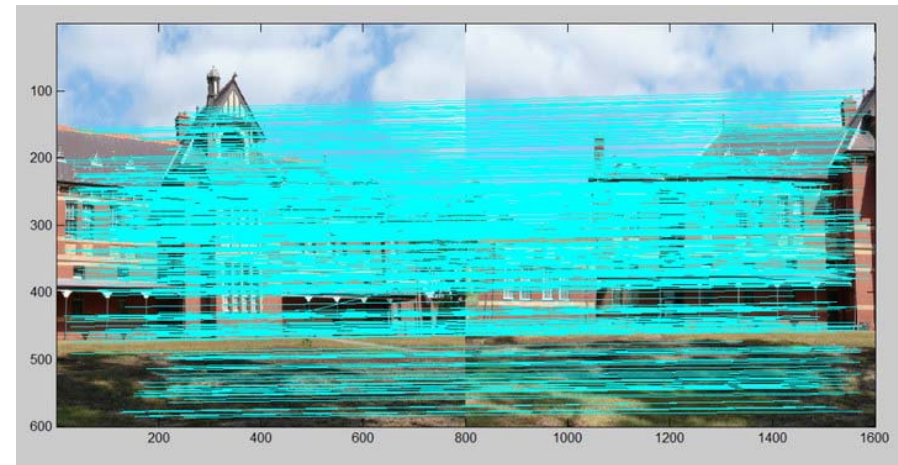
首先根据具体问题设计出某个目标函数，然后通过反复提取最小点集估计该函数中参数的初始值，利用这些初始值把所有的数据分为“内点” (inlier) 和“外点 “ (outlier)，最后用所有的内点重新计算和估计函数的参数。

SIFT算法的应用

SIFT

Scale Invariant Feature Transform

- 图像拼接

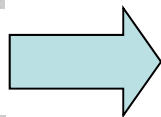


SIFT算法的应用

SIFT

Scale Invariant Feature Transform

- 图像拼接



SIFT算法的应用

SIFT

Scale Invariant Feature Transform

- 来自网友的创意
——周正龙的老虎

图1周正龙的华南虎照片与年画上的华南虎照片12点匹配

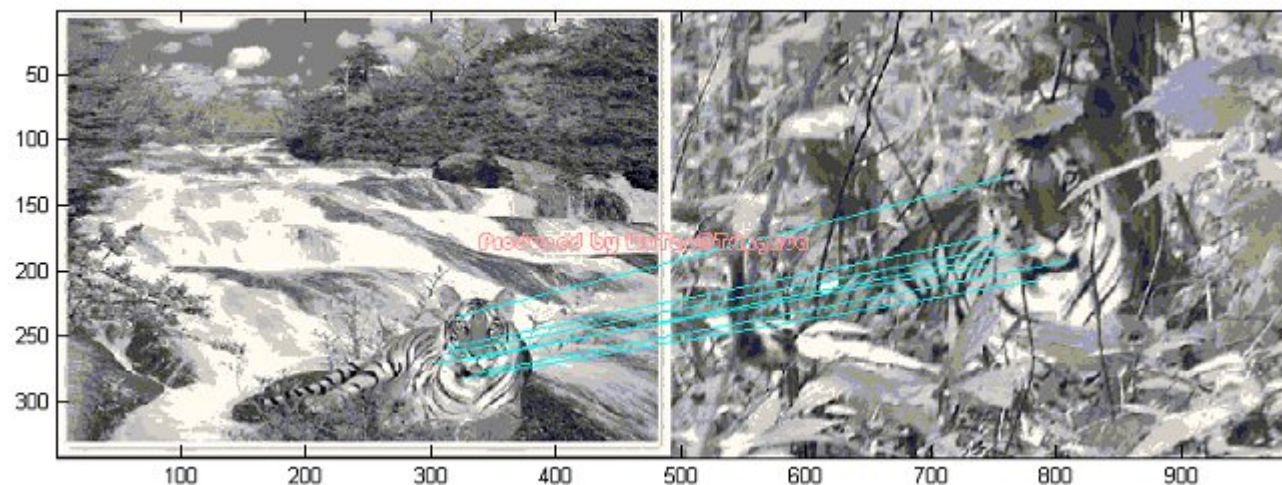
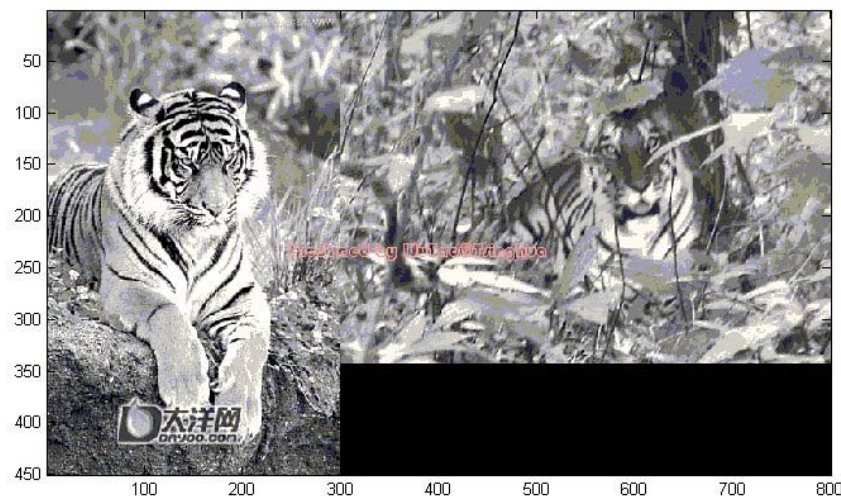
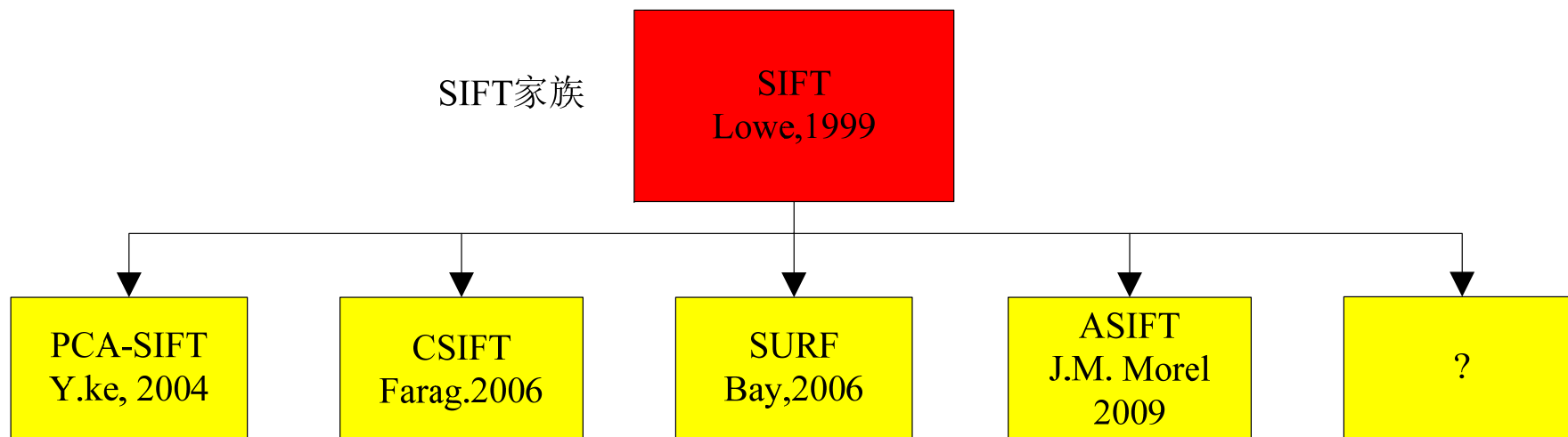


图2周正龙的华南虎照片与真实的华南虎照片0点匹配



[Back](#)

SIFT在图像的不变特征提取方面拥有无与伦比的优势，但其并不是完美的，仍然存在着实时性不高、有时特征点较少、对边缘模糊的目标无法准确提取特征点等缺陷。自从1999年，SIFT算法问世以来，人们从未停止对它的优化和改进。



谢 谢!