SIFT/尺度不变特征转换

## Conception

SIFT(Scale-invariant feature transform)是一种检测图像的局部特征的算法，能够在空间中计算出极值点，并提取出其位置、尺度、旋转不变量。

## Method

* 首先从一组参考图像中提取对象的SIFT关键点，并将其存储在数据库中。
* 通过将新图像中的每个特征分别与此数据库进行比较，并根据其特征向量的欧几里得距离查找候选匹配特征，从而在新图像中识别出一个对象。
* 从整套匹配中，识别出与对象及其在新图像中的位置，比例和方向一致的关键点子集，以滤除离异匹配。
* 通过使用广义霍夫变换的有效哈希表实现，可以快速实现consistent clusters的确定。
* 然后，将对3个或更多特征一致的一个对象及其姿势的每个cluster进行进一步的详细模型验证，并随后舍弃异常值。
* 最后，根据拟合的准确性和可能的错误匹配的数量，计算出一组特定的特征指示存在对象的概率。
* 通过所有这些测试的对象匹配可以高可信度地标识为正确。

## Overview

|  |  |
| --- | --- |
| 问题 | 解决技术 |
| 关键的定位/规模/旋转问题 | DOG/规模空间金字塔/方向分配 |
| 几何扭曲变形 | 局部图像的方向平面的模糊/重采样 |
| 索引和匹配 | 最近相邻(nearset neighbor)最佳优先搜索(Best Bin First search) |
| 集群识别 | 霍夫投票 |
| 模型验证/偏离值检测 | 线性最小二乘法 |
| 假设检验 | 贝叶斯公式 |

### Main Stage

#### 尺度不变特征检测

* David Lowe的方法将图像特征转化为大型特征向量集，其中每个特征向量都与图像平移，缩放和旋转无关，部分特征向量与照明变化部分无关，并且对于局部几何扭曲有着很好的鲁棒性。（类比于视觉皮层神经元，这些神经元编码基本形式，颜色和运动以用于灵长类动物视觉中的物体检测）。
* 关键位置定义为在比例空间中应用的DOG结果的最大值和最小值一系列平滑和重新采样的图像。沿边缘的低对比度候选点和边缘响应点将被舍弃。主导方向（Dominant orientations）分配给局部关键点。
* 这些步骤确保了关键点对于匹配和识别而言更加稳定。然后，通过考虑关键位置半径附近的像素，局部图像方向平面的模糊化和重采样，可以获得对局部仿射失真具有鲁棒性的SIFT描述子。

#### 特征匹配和索引

* 索引包括存储SIFT密钥（key）和从新图像中识别匹配的密钥。Lowe用了从k-d树改进的最佳优先搜索（[best-fin-first search](https://en.wikipedia.org/wiki/Best_bin_first)），仅使用有限的计算量就可以识别高可能性的最近邻居（nearest neighbor）。BFF算法使用了对k-d树改进的搜索次序，以便按查询位置的最近距离来搜索特征空间中的[bin](https://en.wikipedia.org/wiki/Best_bin_first)（我想的是指的是一个块/箱体之类的）。为有效判断搜索次序，其需要使用堆结构的优先队列。
* 通过从训练图像得到的关键点数据库里识别最近邻居，可以找到每个关键点的最佳候选匹配。最近邻居定义为从每个给定的SIFT描述子的向量里欧氏距离最小的关键点。匹配正确的可能性通过取最近邻的距离与次近邻的距离之比确定。
* Lowe的[paper](http://ceessnoek.info/courses/computervisionbylearning/2014/lowe-ijcv2004.pdf)里拒绝了所有距离比大于0.8的匹配，这可以消除90%的错误匹配而只舍弃了低于5%的正确匹配。进一步提高BFF搜索的效率可以在检查了头200个最近邻候选点后停止。对于具有100,000个关键点的数据库，这可以使精确的最近邻居搜索的速度提高大约2个数量级，但正确匹配的数量损失不到5%。

#### 通过霍夫投票进行聚类识别

* 霍夫变换用于对可靠的模型假设进行聚类，搜索符合特定模型姿态的关键点。霍夫变换通过使用每个特征对与该特征一致的所有对象姿态进行投票来识别具有一致解释的特征集群。当特征集群被发现投票给相同的一个物体姿态，这种解释正确的概率就会比其他的单个特征更高。
* 在哈希表中创建一个条目，根据预计的匹配预测模型的位置，方向和比例。搜索哈希表，识别一个bin中至少有3个条目的所有集群，然后将这些bin按大小的降序排序。
* 每个SIFT关键点指明了2D空间下的位置，比例和方向，并且数据库里每个匹配的关键点会记录它们与被找到的关键点的相对参数。这4个参数蕴含的相似变换只能对6个自由度的3D姿态空间作近似，并且也对非刚体变形没有效果。
* 因此，lowe 定位使用30度的宽bin尺寸，比例使用2倍，位置使用最大投影训练图像尺寸（使用预测的比例）的0.25倍。比例较大的SIFT关键样本的权重是比例较小的SIFT的两倍。 这表示更大的比例更有能力滤出在小比例检查下的最可能邻居。
* 为了防止bin assignment里的边缘效应，每个关键点匹配给每个维度最近的两个bin投票，给出对每个假设的16个条目的总和，并且进一步拓宽姿态范围。

#### 利用线性最小二乘法的模型验证

* 对每个识别出的聚类进行验证。在该过程中，对将模型与图像相关的仿射变换的参数计算线性最小二乘解。 对模型点[ ]到图像点[ ]可以写为如下形式

其中模型变化为[ ]，仿射旋转，比例和拉伸由参数，，和表示。为了求解变 换参数，可以重写以上公式，以将未知数表示到一个列向量中。

这个公式显示了单个匹配项，但是可以添加任意数量的其他匹配项，每个匹配项对第一个和最 后一个矩阵添加另外两行，且为了求解至少需要3个匹配项。

* 可以将这个线性系统写为

其中是一个已知的的矩阵（通常），是一个未知的维参数向量，是一个已 知的维测量向量。

因此，最小化向量是这个**normal equation**的解  
 通过

线性方程组的解以矩阵形式给出，称为的伪逆矩阵。

这样可以最大程度地减少从投影模型位置到对应图像位置的距离的平方和。

#### 离异值检测

* 给定参数解，现在可以通过检查每个图像特征与模型之间的一致性来删除异常值。给定线性最小二乘解， 每个匹配都必须在霍夫变换bin中用于参数的误差范围的一半以内。当离异值被舍弃后，这个线性最小二乘解会用剩下的点重新求解，这个过程是不断重复的。如果在舍弃离异值后剩下的点少于3个，匹配就会被拒绝。另外，自上而下的匹配阶段用于添加与投影模型位置相符的任何其他匹配，由于相似度变换近似或其他错误，这些匹配可能已从霍夫变换bin中丢失。
* 最终决定是否接受或拒绝模型假设基于具体的概率模型 。这个方法首先给定模型的预计尺寸，区域内的特征数量以及拟合的精度，计算预计的模型姿态的错误匹配数量。然后由贝叶斯概率分析，根据找到的匹配特征的实际数量，给出物体存在的概率。只有一个解释子的最终正确概率大于0.98时模型才会被接受。Lowe的基于物体识别SIFT方法给出了除了在宽光照变化和非刚性转换的其他情况的一个优秀的结果。