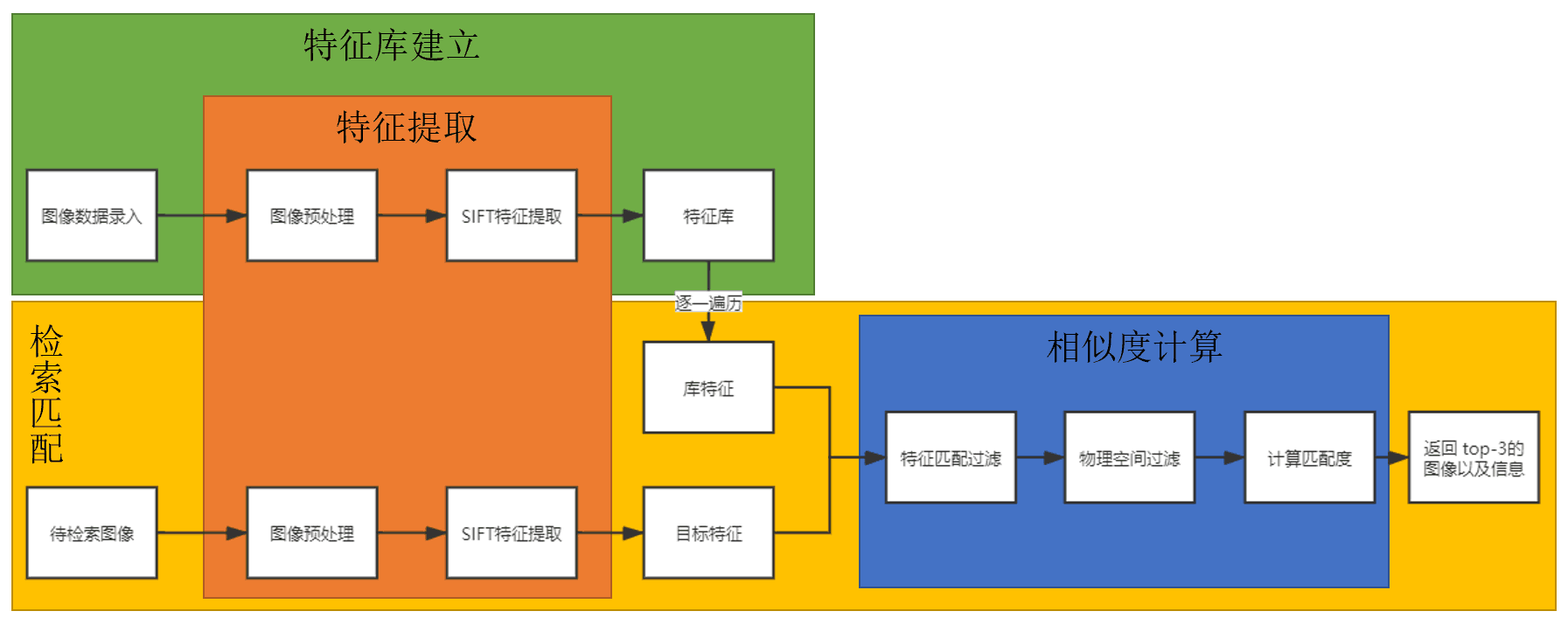
# 匹配算法

## 算法流程

匹配算法的核心主要由特征提取与匹配检索两部分构成，流程包括特征库的建立和目标检索。



## 特征提取

主要包括图像预处理与SIFT特征提取两部分。特征提取的目标是提取原始数据中具有描述性与鲁棒性的特征，一方面提高检索的速度效率，另一方面保证同一个人在不同角度和设备下拍摄的牙片不影响结果。

### 图像预处理

由于库数据的基本参数存在差异，因此需要预处理。图片具有不同的尺寸，部分的右侧有信息栏，这些差异会对后续的SIFT特征提取与物理空间过滤产生影响，所以我们首先截去图像左侧150pix，保证去除信息栏，然后统一缩放到2800\*1400，保证图像大体对齐。

### SIFT特征提取

SIFT，即尺度不变特征变换，这种描述子具有尺度不变性，可在图像中有效检测出关键点局部特征描述子。算法流程步骤如下：

尺度空间极值检测：搜索所有尺度上的图像位置。通过高斯微分函数来识别潜在的对于尺度和旋转不变的兴趣点。

关键点定位：在每个候选的位置上，通过一个拟合精细的模型来确定位置和尺度。关键点的选择依据于它们的稳定程度。

方向确定：基于图像局部的梯度方向，分配给每个关键点位置一个或多个方向。所有后面的对图像数据的操作都相对于关键点的方向、尺度和位置进行变换，从而提供对于这些变换的不变性。

关键点描述：在每个关键点周围的邻域内，在选定的尺度上测量图像局部的梯度。这些梯度被变换成一种表示，这种表示允许比较大的局部形状的变形和光照变化。

描述子筛选：在全部关键点中，依据响应强度筛选出前200个描述子，将描述子与对应的特征点坐标保存到数据库，仅保留200个描述子是为了降低匹配耗时，提高后续的检索效率。

## 匹配检索

主要包括特征匹配过滤、融合物理信息、相似度计算三个部分。匹配检索的目标是依据计算出的待检索特征在特征库中找出最为相似的个体，并报告相应的相似度，以便于进行身份确认。匹配时我们从特征库中逐一取出特征与目标特征进行计算相似度，由于每次计算间相互独立，所以可以并行化加速，最终报告top3的结果。

### 特征匹配过滤

首先在库特征与目标特征的200个描述子向量间进行最邻近匹配。由于描述向量维数较高，为了加速匹配这一过程，我们采用了FLANN算法，即快速近似最近邻搜索算法，算法只要通过建立多棵随机k-d树，从具有最高方差的N维中随机选取若干维度，用来做划分，加快最邻近的搜索速度。

由于特征点的响应强度有差异，为了提高鲁棒性，对已经找到的匹配对进行强度筛选。采用的方案是top-1匹配的描述向量间的距离应当小于0.7\*top-2匹配的描述向量间的距离，满足条件的匹配对强度较高，准确性更好。

### 融合物理信息

由于SIFT特征只是局部描述子，所以没有有效利用全局信息。为了有效利用全局信息，在已经匹配上的点对之间计算物理坐标之间的欧氏距离。通过下述的直方图可见，在正例中匹配点的物理坐标之间的欧氏距离分布在很小的阈值下，而负例中匹配点的物理坐标之间的欧氏距离分布则较为分散，所以我们设定阈值进行过滤。

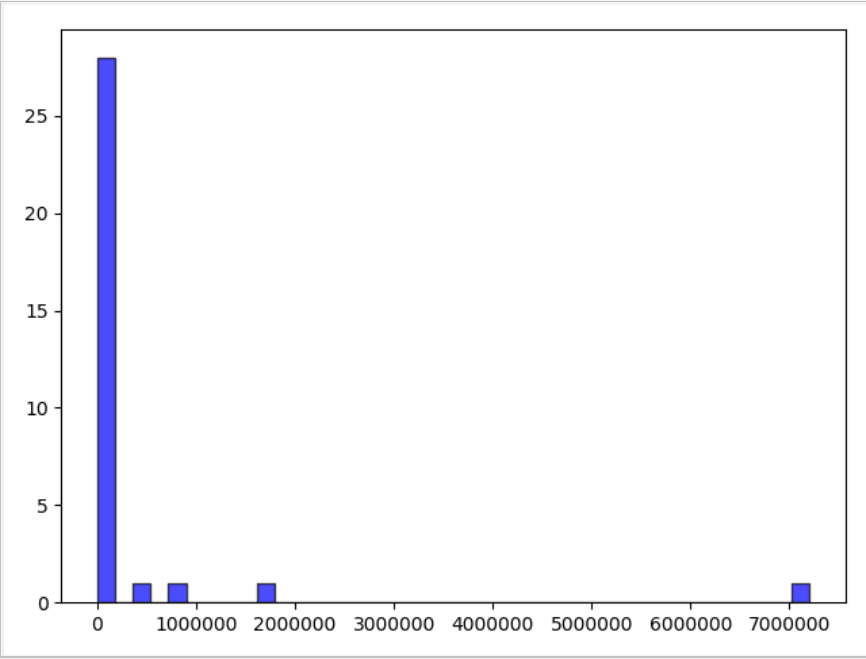
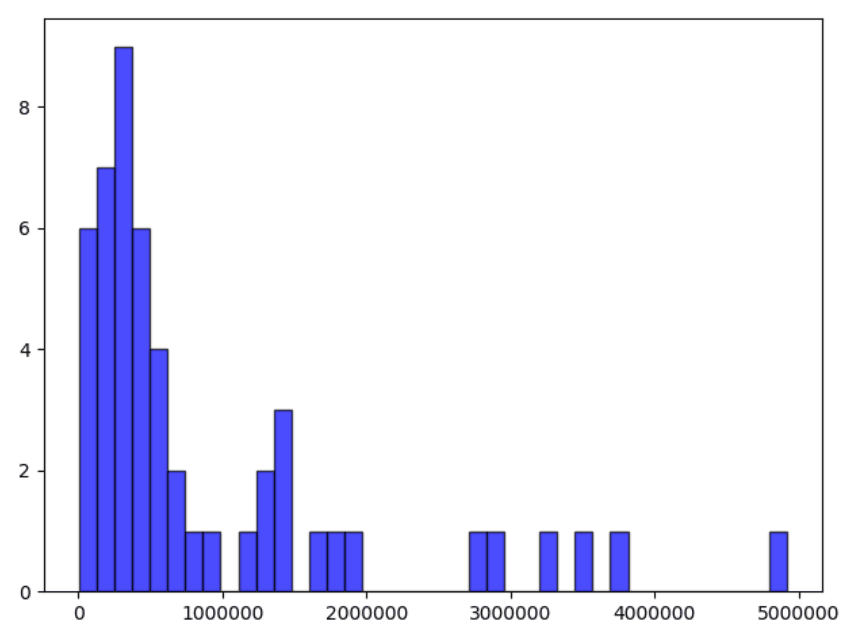
 

Figure 正例的匹配点间距离分布直方图 Figure 2负例的匹配点间距离分布直方图

### 相似度计算

最终以余下的匹配特征点的对数作为相似度。由于仅保留了200个描述向量，所以相似度的范围为0-200，在计算出全部的相似度后，我们报告相似度最高的top3

由于不能保证检索的图片在特征库中均存在对应，考虑到正例的相似度大体在20左右，所以对于相似度小于10的个体，依然报告但标注存疑。