词袋生成和使用

一、词袋概念和作用

词袋模型(Bag of Words, BoW)是视觉SLAM(同步定位与建图)系统中用于**高效场景识别和闭环检测**的核心技术。

在视觉SLAM中，词袋模型将图像视为"视觉单词"的无序集合，这些单词是通过**对图像局部特征进行聚类**得到的。与文本处理类似，它忽略特征点的空间排列，仅统计各类视觉单词的出现频率，形成图像的紧凑表示。

核心思想

词袋模型的核心思想包含三个关键步骤：

1. 特征提取：从图像中检测并描述局部特征点(如SIFT、ORB等)
2. 视觉词典构建：通过聚类算法(如K-means)将特征描述子空间离散化为视觉单词集合
3. 图像表示：将图像特征映射到词典空间，生成词频向量(直方图)

这种表示方法具有视角不变性的特点——即使观察角度变化导致特征点顺序改变，只要内容相似，生成的词频向量也会相似

词袋模型在SLAM中的作用

在视觉SLAM系统中，词袋模型主要承担以下关键功能：

1. **闭环检测**

闭环检测是词袋模型在SLAM中最重要的应用。当机器人或移动设备重返之前访问过的地点时，系统需要识别这种"回环"以消除累积误差。

词袋模型通过以下流程实现闭环检测：

- 将当前帧与关键帧数据库中的历史帧进行快速相似度比对
- 当相似度超过阈值时，触发闭环候选
- 结合几何验证(如RANSAC)确认有效闭环

相比暴力匹配，词袋模型能大幅提升闭环检测效率。

例如ORB-SLAM3中，通过词汇树结构可在毫秒级完成大规模数据库检索。

2. **重定位**

当SLAM系统因剧烈运动或遮挡导致跟踪丢失时，词袋模型可帮助系统快速重定位。它通过比对当前视图与地图中的关键帧，找到最可能的位置假设。

3. **特征匹配加速**

词袋模型构建的直接索引可加速帧间特征匹配：

- 仅需比较属于同一视觉单词或中间节点的特征
- 避免全特征暴力匹配的高计算成本
- ORB-SLAM中可减少90%以上的匹配计算量

4. **关键帧选择**

通过分析词袋向量的相似性，系统可智能判断是否添加新关键帧，避免冗余存储，优化地图紧凑性

二、词袋模型的分类

根据不同的标准，SLAM中使用的词袋模型可分为以下几类：

1. 按特征类型分类

(1) 传统手工特征词袋

- SIFT词袋：使用128维SIFT描述子，具有尺度、旋转不变性，但计算成本高
- ORB词袋：采用256位二进制ORB描述子(BRIEF)，效率高但不变性较弱
- BRIEF词袋：纯二进制描述子，匹配速度极快(汉明距离)，适合实时系统

(2) 深度学习特征词袋

- 卷积特征词袋：使用CNN中间层特征(如Pool5)，语义性强但维度高(需PCA降维)
- 混合特征词袋：结合手工特征与深度特征，平衡效率与判别力

2. 按词典结构分类

(1) 扁平式词袋

- 使用单层K-means聚类生成视觉单词
- 结构简单但规模受限，适合小规模场景

(2) 分层词袋(词汇树)

- 通过分层K-means(HKM)构建树状结构

- 叶子节点为视觉单词，非叶子节点为中间概念
- 支持对数时间检索，适合大规模应用
- ORB-SLAM使用的词汇树包含超过100万个节点

(3) 二进制词袋(BoBW)

- 专为二进制描述子(如BRIEF)设计
- 使用K-medians聚类和汉明距离
- 内存占用小，匹配速度比SURF快10倍

3. 按训练方式分类

(1) 无监督词袋

- 仅使用K-means等无监督聚类
- 不利用类别标签，通用性强

(2) 有监督词袋

- 结合类别信息优化词典
- 提升特定场景下的判别力

4. 按应用领域分类

(1) 视觉词袋

- 处理2D图像特征
- 适用于单目/双目SLAM

(2) 3D点云词袋(如BoW3D)

- 处理LiDAR点云特征(如LinK3D)
- 直接输出6-DoF闭环位姿
- 在KITTI数据集上仅需50ms完成闭环

三、词袋模型的构建流程详解

理解词袋模型的完整构建流程有助于深入掌握其工作原理。下面以ORB-SLAM中的实现为例，分步骤详细说明

1. 离线词典训练阶段

(1) 特征提取

- 收集大量训练图像(通常来自类似场景)
- 每幅图像提取ORB特征点(FAST角点+BRIEF描述子)
- ORB特征为256位二进制向量，适合高效匹配

(2) 分层聚类

- 设定树的分支数K和深度L(如K=10，L=6→100万叶子节点)
- 第一层聚类：对所有描述子执行K-means，得到K个簇中心
- 递归聚类：对每个簇重复上述过程，直至达到深度L
- 最终形成词汇树，叶子节点即为视觉单词

(3) 权重计算

- 为每个单词计算IDF权重：IDF = log(N/N_i)
  - N：总图像数
  - N_i：包含该单词的图像数
- 出现频率越高的单词，区分度越低，权重越小

2. 在线应用阶段

(1) 图像表示

- 对新图像提取ORB特征
- 每个描述子从词汇树根节点开始，逐层选择汉明距离最小的分支
- 到达叶子节点后，对应单词的词频(TF)增加
- 最终生成TF-IDF加权词向量

(2) 索引构建

- 直接索引：记录特征点与词汇树中间节点的归属关系，加速帧间匹配
- 逆向索引：记录每个单词出现在哪些关键帧中，加速数据库检索

(3) 相似度计算

常用方法包括：

- L1范数：s(v1,v2) = Σ|v1_i - v2_i|
- L2范数：欧氏距离

- 余弦相似度：$v1 \cdot v2/(|v1||v2|)$
- 卡方距离：$\Sigma(v1\_i-v2\_i)^2/(v1\_i+v2\_i)$

四、实际案例

现在有一个室内空间定位的需求，要求在使用图片重建好的地图上，使用新图片快速定位，得到新图片的位姿。

目前使用superpoint进行特征提取，那么为了加速图像的匹配，避免每次都需要暴力搜索全部数据库，可以使用词袋模型来预先筛选出相似的图片帧，能够大大减少检索和匹配的时间。

使用superpoint特征点训练和使用词袋的流程如下：

- 获得地图数据库中图片id以及描述子
- 对所有描述子进行聚类，得到若干个单词向量，保存单词向量文件
- 计算出每一个图片的word（通过直方图统计单词出现的频次，然后归一化得到图片的word）
- 使用faiss检测word索引，方便快速查找