**联合优化以实现有效和高效的图像搜索**

**（这是研究图像搜索的）**

**一．引言**

**图像搜索方向需要注意的问题（1）搜索效率（2）由于基于视觉词（论文3提到过）匹配的图像搜索带来的准确性问题：**

图像搜索旨在从大型图像数据库中高效有效地检索所需图像。 然而，由于照明，比例尺，视点，方向等方面的外观变化，识别相关图像仍然是一项艰巨的任务。[1] – [5]，特别是对于大型图像数据库。TRECVID评估已通过实验验证了这一点。 自动/手动/交互式搜索[6]。 此外，随着图像数据库大小的快速增长，图像搜索效率也正成为一个关键问题。最近，词袋模型（BoW）由于其简单性和有效性而在多媒体搜索领域变得很流行[7]，[8]，[9]。 基本思想是将图像表示为无序视觉词的集合，并将所有视觉词组织到倒排表中，以进行有效的图像搜索。 早期的工作主要集中在通过使用不同的矢量量化技术[1]，[9] – [13]来最佳逼近本地图像描述符来构建视觉词汇。

由于视觉词是局部描述符的近似表示，因此基于视觉词匹配的图像搜索的准确性将不可避免地由于量化误差而降低。 为了减轻它，最新的扩展[14] [15]引入了附加的嵌入代码以补偿量化误差。

**本文为了解决什么问题而出现：**

虽然已经提出了许多基于BoW的方法和嵌入方法，但是人们付出了较少的努力来定量地了解它们的工作机制。例如，量化误差如何影响搜索性能？ 减少这种情况的主要因素是什么

量化误差？ 嵌入方法如何提高搜索性能？ 了解它们的本质对于开发更好的图像搜索系统至关重要。 为此，我们系统地研究了BoW模型和嵌入方法的一些关键因素如何影响搜索效果和效率，并总结了一些观察意见，以指导系统设计。

**本文提出PKMLSE的新图像搜索办法：**根据这些观察，我们提出了一种称为PKMLSE的新图像搜索方法，该方法通过结合分区的k均值聚类方案（PKM）和线性分段嵌入（LSE）方法来最佳地平衡图像搜索系统的有效性和效率。

**本文的主要贡献可归纳如下：**

1. 我们定量地发现了BoW模型和嵌入方法的工作机制，并总结了来自综合实验的一些观察结果。 尽管没有为这些观察结果提供严格的证据，但它们确实可以很好地指导有效性和效率方面更好的图像搜索引擎的设计。 此外，这些观察结果并非特定于所提出的PKMLSE方案，它们通常可用于优化现有工作。 据我们所知，观察结果从未发表过。
2. 我们提出了一种新的方案，可以有效地构建非常大的视觉词汇并分配视觉单词。 代替[15]中报道的生成视觉单词的辅助信息，在这里PKM方法被用来构建视觉词汇。 更重要的是，提出的PKM词汇构造方法极大地满足了所提出的观察中最优规则的要求。 如实验部分所示，PKM方案提供了从本地描述符到视觉单词的极其有效的映射过程，并且具有非常小的存储成本用法而不会降低搜索效率
3. 我们提出了一种判别性嵌入方法，称为线性分段嵌入。 尽管此方法非常简单，但是生成的嵌入代码提供了出色的能力，可以区分真假匹配。 结合提出的PKM方案，嵌入方法通过根据提出的观察共同优化参数设置，显着提高了搜索准确性，并降低了计算成本。

**二．相关工作**

**构建视觉词的发展路程：**BoW模型首先用于自然语言处理和信息检索领域，其中文档被表示为无序单词袋。在多媒体搜索领域，采用BoW模型将图像表示为无序视觉词的集合[1]，其中一个关键问题是如何构建视觉词汇[16]。本质上，视觉词汇的构建是对训练集中的局部图像描述符进行矢量量化。广泛使用的构造方案是执行精确的k均值聚类（FKM），并将每个聚类质心视为一个视觉单词[9]。在单词分配阶段，通过搜索最接近的聚类质心，将新的本地描述符映射到视觉单词。

尽管此方法非常有效，但其时间和空间复杂度非常高，使其对于处理大量训练样本甚至不切实际。为了提高计算效率，已经设计了许多方法[1]，[10]，[11]，[17]，[18]来有效地构造视觉词汇。近似k均值聚类方案[11]通过用近似最近邻方法[19]代替精确计算来加快词汇构造。分层k均值聚类方案（HKM）[1]通过递归执行k均值来降低时间复杂度集群数量少的集群。其他有效的词汇构建方案包括[10]，[17]。但是，所有这些方法都通过在原始特征空间中对训练样本进行量化来构造视觉词汇，在构建大型视觉词汇时需要非常大的训练集，从而导致训练阶段的计算成本高和占用大量内存。另外，由于每个视觉词被单独存储为高维向量，因此在查询阶段用于保持较大视觉词汇的存储器使用也很重要。

**本文，为了解决这些问题，我们通过在低维度上分别构建一系列小的视觉副词来构建视觉词汇子空间，并通过笛卡尔将这些子词汇组合起来构造视觉词汇**。这样，时间和空间成本仅花费在构建和存储这些子词汇上，这是微不足道的。在我们的最新工作中，详细讨论了PKM的一些良好功能，例如无偏属性[33]。

**构建后的与数据库交互过程?：**【给定一个视觉词汇，通常用于大规模图像搜索的范例是在用无序视觉词集合表示这些图像后，用一个反向文件来组织数据库图像。特别地，每个视觉词对应于一个列表或单元格。给定一个新的数据库描述符，通过使用最近的邻居搜索将其映射到某个单词，然后将包含描述符图像ID的新条目插入到相应的列表中。一旦我们以离线方式将数据库图像的所有本地描述符索引到反向表中，下一步就是在给定查询描述符时查询反向表。查询过程类似于索引过程。同样，查询描述符首先映射到视觉单词，然后将对应列表中的所有项目作为匹配项返回。】（和论文3有点联系 数据库索引）

**基于视觉词匹配的搜索精度上的发展路程：**由于视觉词是局部图像描述符的近似表示，因此基于视觉词匹配的搜索精度将不可避免地由于量化误差而降低。为了提高标准BoW框架的有效性，已在引入更多辅助信息方面做出了一些努力。在[14]，[15]和[20]中，

（1）将嵌入方法引入基于标准BoW的方案中。基本思想是嵌入物体进入低维嵌入空间中的点，其中点之间的距离近似于物体之间的距离[21]，[22]。以这种方式，在嵌入空间中用视觉单词和紧凑的描述来表示局部图像描述符。例如，[15]中的基于乘积量化的嵌入（PQE）方法将本地图像描述符映射到64位嵌入代码中。

(2)除了引入嵌入代码外，还可以通过探索空间信息来提高图像搜索性能。为此，Zhang等。 [23]在构建视觉词汇时直接考虑了局部图像描述符之间的空间上下文信息，而其他人则通过金字塔匹配[24]-[27]引入了空间信息。

（3）然而，它们中的大多数是基于一个基本假设，即本地图像描述符仅映射到一个视觉词（即M = 1），而其中一些通过采用多重分配显示出性能提升[28]，[29]。 换句话说，没有充分考虑多重分配。 另外，视觉词汇量K和分配量M之间的关系也没有被系统地研究。 注意，将本地描述符映射到M个视觉单词是要找出视觉词汇中的M个最近邻居。 **在本文中，我们旨在发现BoW模型和嵌入方法的工作机制，以及提出一些指导更好的图像搜索系统设计的规则。 正如我们将在以下各节中讨论的那样，了解它们的本质可以有益于优化系统的设计，并在有效性和效率方面显着提高性能。**

**三．本论文【就是纯提出设计规则吧，然后写出了自己为了追求精准度和完整度，在实验中做了哪些改进】**

在基于BoW的图像搜索框架中，搜索过程是查询图像和数据库图像中匹配的视觉单词的投票过程。理想情况下，如果视觉单词可以完美地代表本地图像描述符，则投票分数可以准确反映出图像的相似性。但是，由于量化误差，几乎不可能获得理想的结果，因为匹配单元中只有部分项是给定查询描述符的真实匹配。因此，图像搜索性能在很大程度上取决于量化误差。

如现有工作[7]所示，影响量化的两个关键因素是（1）视觉词汇量K（2）词汇构造方法。实际上，当K足够大时，即使随机量化也可以获得良好的近似。也就是说，我们可以通过扩大词汇量来采用非最佳方法来构建具有低量化误差的视觉词汇。补充量化误差的另一种方法是通过使用嵌入方法在粗匹配的单元格中细化这些项。仅将其与查询描述符的嵌入距离不超过阈值的项目作为匹配项返回。这样，可以消除一些错误的匹配，并且可以相应地提高准确性。对于单个分配情况，当一个单元格中的错误匹配远远大于真实匹配时，精炼过程可以删除大量错误匹配并保留足够的真实匹配。但是，当K大时，由于量化，匹配单元中的真实匹配比率很高误差相对较小。在这种情况下，使用嵌入方法不仅会修剪错误的匹配项，还会删除更多的真实匹配项，这将违反描述符匹配项。换句话说，当M设为1时，现有的嵌入方法仅适用于某些中小型视觉词汇。

为了验证上述结论，我们进行了几次实验-【见实验】从匹配的准确性和完整性的角度出发。在我们的实验中，我们使用玩具数据集进行测试，其中10KSIFT描述符用作查询描述符，而另一个1000K SIFT描述符被视为数据库描述符。对于原始的BoW模型，所有数据库描述符都是通过预先构造直接映射到倒置表视觉词汇，并针对以下内容进行搜索过程每个查询描述符通过返回数据库描述符落入匹配的单元格中。对于嵌入方法，请添加-国家/地区代码与单元格中的项目相关联，只有嵌入距离查询达到阈值的项目旧的将作为匹配项返回。在这里，汉明嵌入（20）中的（HE）方法与最优参数一起使用设置在原始纸张中报告。在我们的实验中每个查询描述符的100个最近邻居被视为地面真相，通过蛮力搜索获得数据库描述符。为了评估准确性和完整性，描述符匹配的强度，我们提出了两种测量方法，即加权平均精度（AP）和加权平均召回率（AR），定义为。。。。。。。。巴拉巴拉

**四．实验**

**五．总结**

在本文中，我们系统地研究了BoW模型的基本工作机制和嵌入方法，并在基于描述符匹配的实验基础上总结了一些观察结果。根据这些观察，我**们提出了一种新的图像搜索方案，该方案在有效性和效率方面进行了共同优化**。我们的综合实验表明，这些发现有助于设计新的图像搜索方案。此外，我们还对非常大的图像数据集测试了该方案，以证明其可扩展性。实验结果表明，所提出的方法优于最新方法。然而，尽管所提出的图像搜索方案取得了出色的性能，但先前工作中存在的内存使用问题仍然没有得到解决。反向表的大小与本地描述符的数量和嵌入代码的长度密切相关，因此无法将其装入内存。将来，我们将开发一些更紧凑的图像表示来解决这个问题。