利用颜色信息进行图像检索的视觉词汇袋方法

**摘要 —由于图像数据库的快速增长基于内容的图像检索引起研究人员的关注。基于内容的图像检索的许多方法已经提出，但尚未完全获得令人满意的结果。本文提出了一种基于内容的新方法，即使用视觉单词袋（BoW）进行图像检索建议。在这种方法中，两个单独的视觉词汇是根据尺度不变特征变换（SIFT）生成和色相描述符。传统的SIFT BoW模型不使用图像描述中的任何颜色信息。我们的方法中加入色相的描述符，它添加了补充信息以BoW表示图像并增加检索的整体性能。BoW模型中的量化是必不可少的一步，但是它会导致量化损失，对最终结果有很大影响。克服这一不利条件效果，在我们提出的方法中使用了软分配，其中每个特征向量都可以分配给多个最近的视觉对象基于一些加权函数的单词。实验结果基准数据集上的数据表明，该方法具有相对于其他方法而言，性能优越。**

# 一、 导言

图像检索是近十年来引起研究者关注的一个活跃的研究领域。在互联网日益普及的今天，图像的搜索和快速检索变得非常重要。

在图像检索领域引入了多种方法。这些方法分为两类：基于文本和基于内容的图像检索方法。基于文本的图像检索方法是20世纪70年代提出的一种利用合适的关键词对图像进行标注的检索方法。在测试阶段，检索与用户查询具有相似关键字的图像。基于文本的图像检索方法有两个主要缺点。首先，注释任务是耗时的，而且对于庞大的数据库是不可行的。其次，为图像标注指定关键字是主观的。此外，关键字不能完全解释图像。为了克服这些困难，基于内容的图像检索在20世纪90年代被提出，其思想是使用图像内容而不是像关键字这样的元数据。

在基于内容的图像检索中，对于数据库中的每个图像，都使用颜色、纹理和形状等线索。这一类别最重要的优点之一是它独立于人类的相互作用。

Hiremath和Pujarin将图像及其补码划分为大小相等的不重叠图像块。每个图像的最终描述符是通过图像块和相应的补码块的条件共现颜色通道来实现的。这个描述符可以考虑纹理信息。为了考虑图像的形状信息，还使用了边缘图像的不变矩。利用梯度矢量流（GVF）实现边缘图像。图像被分成大小相等的块，然后计算每个块的LBP直方图。在测试阶段，查询映像以相同的方式分区。在下一步中，查询图像中的每个块独立地与训练图像的块匹配。相似性度量计算为相应块之间的距离之和。辛格等人。 为了考虑空间信息，将图像分成三个水平不重叠的块。每个块由三个颜色矩和Gabor滤波器的响应来描述。计算了4个尺度和6个方向的Gabor滤波器。

图像检索的许多方法都是基于视觉词汇袋（BoW）。BoW的思想来源于文本检索领域。在词袋模型中，首先对特征进行提取和描述。然后将这些特征量化为若干个簇。聚类中心称为视觉词。最后，每个图像被表示为一个视觉词汇的集合。使用SIFT提取低级特征，然后用SIFT和LBP描述这些特征。接下来，使用kmeans集群来获取码本。最后，将每个图像的特征向量映射到视觉词上，得到每个图像的直方图。

本文提出了一种基于BoW模型的新方法。本文的主要贡献是结合SIFT和色调描述子，将颜色信息融入到图像的描述子特征中。选择SIFT作为描述符是因为它对缩放、旋转、变换和小失真具有不变性。同时，还通过色调获得了光照几何和镜面反射的不变性。此外，为了减少BoW模型中的量化误差，我们提出了一种软分配方法。这种方法的一般示意图如图1所示。

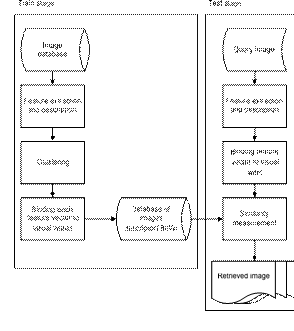


图 1 基于BoW模型的检索算法流程图

1. 建议的方法

在本节中，提出了可行的方法。该方法首先提取低层特征（如关键点）。本文采用SIFT进行关键点的检测，而任何兴趣点的检测都可以使用。然后，在每个关键点周围考虑一个局部补丁并加以描述。为了描述这些面片，使用了SIFT和色调描述符。在下一步中，使用k-均值算法对获得的描述符进行量化以生成码本。需要注意的是，SIFT码本和色调描述符是分开生成的。最后，将每个图像表示为量化特征向量的频率，称为可视词。本文采用软赋值方法，而不是将每个特征硬赋值给一个可视词。在软分配中，每个特征作为最终表示中视觉词的加权组合贡献，可以减少最近邻分配中存在的量化误差。

下面分别在A小节和B小节给出了特征提取和描述，C小节讨论了量化和表示方法，最后在D小节给出了相似度度量的说明。

A、 特征提取

如前所述，该方法的第一步是特征提取。有很多方法可以提取低级特征。这里，使用SIFT是因为它对缩放、旋转、过渡和小失真具有不变性。提取的低层特征称为关键点。

B、 特征描述

在此步骤中，描述提取的关键点。这里使用了两个描述符：SIFT和色调描述符。每个描述符的计算方法如下。

SIFT：首先，在每个关键点周围考虑16 16补丁。然后将这些局部补丁划分为四个块。每个块的梯度分布被连接起来以生成最终的描述符。最后，用128维向量描述每个面片。

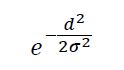
色调：这里，考虑了SIFT关键点周围的一个20 20补丁。然后将该RGB块转换为HSV颜色空间。在HSV颜色空间中，色调描述子对光照几何和镜面反射都是不变性的，但在灰轴附近是不稳定的。为了克服这个挑战，魏杰等人对色调转换应用分析。结果表明，色调的确定性与饱和度成反比。因此，通过使用饱和度对每个色调样本进行加权，色调直方图将具有鲁棒性。因此，本文采用了饱和度加权的色调直方图。

C、 量化与表示

图像的描述空间被量化以生成码本。为此，使用K-means算法。结果表明，量化过程中存在损耗，这对最终的结果有很大的影响。为了考虑量化带来的损失，本文采用了软分配方法。

软分配与硬分配不同。硬赋值是将每个特征向量赋给可视词最常用的方法。在这种方法中，特征向量可以只分配给一个可视词。该方法在量化方面存在损失，有两个缺点。首先，它不能考虑描述空间中特征之间的距离。例如，尽管在描述符空间中很接近，但是可以将两个相对接近的特征向量指派给两个不同的可视词。其次，如果三个特征向量分配给一个可视词，则无法指定哪个特征向量最接近。

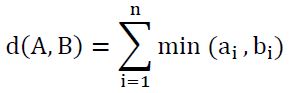
在软赋值中，每个特征向量都可以根据其与视觉词之间的距离的权重，被赋给r个最近的视觉词。该权重根据（1）计算：

（1）

其中d表示特征描述符到可视词的距离。σ2参数表示有多少可视单词可以获得实质性权重。本文将其设为6.2，在实际应用中具有良好的性能[10]。在这里，每个作业中有三个最接近的可视单词（r=3）。然而，已经表明，如果最近的可视单词的数目被设置为大于4的值，则不提供附加信息。软分配的主要优点之一是减少了量化误差。

D、 相似性度量

在此步骤中，计算查询图像和训练图像之间的相似度。迄今为止，已有许多相似性度量方法被提出。它们包括Minkowski距离、余弦距离、直方图交集等。本文采用直方图交集，其计算公式如下：

(2)

其中，A和B表示标准化图像的直方图，n并表示可视字的数目。根据（2），检索到的图像从最相似到最不相似进行排序。

如前所述，在提出的方法中，利用SIFT和色调描述符建立了双弓模型。因此，计算每个模型的相似性度量

分别用和表示的独立的。最后，根据这些相似性的加权和检索图像，计算如下：

3(3)

三、 实验结果

在本节中，我们将我们的方法应用于Wang数据集，Wang数据集是Corel数据集的一个子集，由10个类别的1000个图像组成。这些类别包括非洲、海滩、建筑、公共汽车、恐龙、大象、花、马、山和食物。每个类别都有100幅256 384或384大小的图像

256. 数据库图像的一些示例如图2所示。

对于训练集中的每个图像，提取SIFT特征。然后，用SIFT和色调描述符描述这些特征。为了获得码本，采用k-均值聚类算法，该算法需要聚类个数。实验结果表明，当色调和筛选描述符都设置为200时，可以获得最佳性能。然后根据码本对每一幅图像建立BoW模型。

在测试步骤中，根据（3）的相似性值对检索到的图像进行排序。实验表明，当SIFT和色调描述子（，）的权重分别设置为0.8和0.2时，系统的性能最佳。

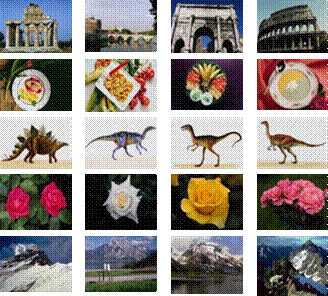
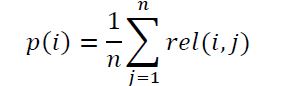
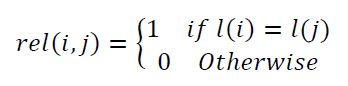


图 2 wang数据集示例

系统的性能是通过精确评分来衡量的，该评分反映了系统仅返回相关图像的能力。每个查询图像（i）的精度计算如下：

（4）

其中j表示训练图像，计算如下：

（5）

其中和表示图像i和j所属类别的标签。

对于数据集的每个类别，我们使用100个图像作为查询，并考虑20个排名靠前的结果进行性能评估。最终的结果是基于每个类别的平均精度。将该方法与其他图像检索方法进行了比较。

各类别的平均精度和各类别的总平均精度见表1。如图所示，与其他方法相比，该方法具有更好的性能。此外，所提出的方法在10个类别中有7个类别有良好的性能。这些情况用粗体显示在表一中。四个查询的结果如图3所示。

在图3的示例中，对于每个查询，表示前三个检索到的图像。

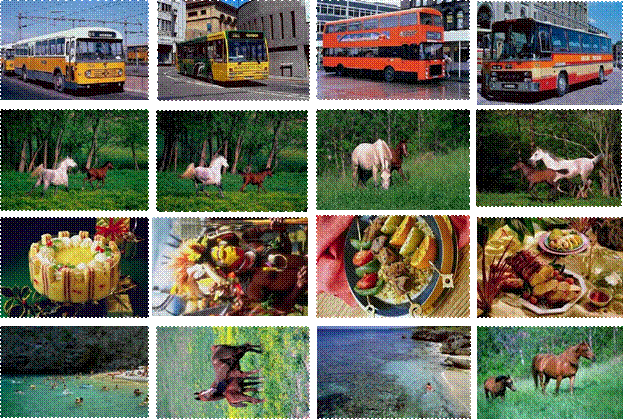


图 3得到了该方法的结果。A）查询图像，B），C），D）检索结果

图4示出了用于查询和检索图3的行2和3的图像的BoW模型。如图3所示，第3行中的第一个检索图像与查询图像不在同一类别中。产生这种错误的原因是检索图像的SIFT和色调的BoW模型与查询图像的BoW模型非常相似。尽管检索图像的BoW模型在视觉上看起来可能与查询图像的不同，但是（3）条件的结果表明相似性。

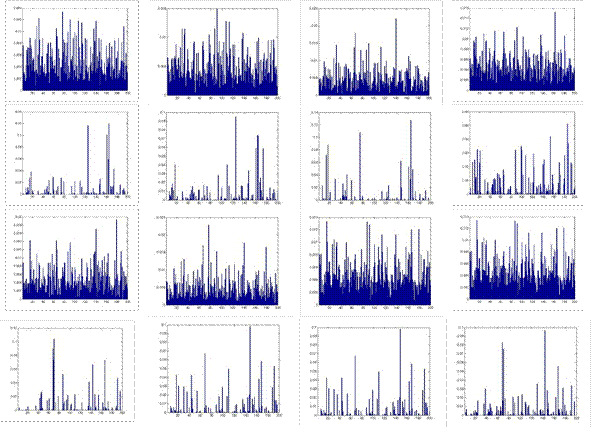


图 4 图3第2排和第3排的词袋模型

软分配对最终结果的影响如表二所示。可以看出，通过应用软分配，可以达到17.8%的改进。这是因为软分配试图减少量化中的损失。为了证明色调描述子对检索结果的有效性，我们分别研究了SIFT和色调描述子系统的性能。所得结果如表3所示，证实了这两个描述符的组合。

表1 表2 表 3

四、 结论

本文提出了一种新的图像检索方法，该方法基于一个包含颜色信息和软分配的视觉词包模型。在这种方法中，我们使用SIFT和色调描述符来描述图像。SIFT描述子对缩放、旋转、变换和小失真具有不变性。色调描述符对于光照几何和镜面反射都是不变的。因此，对于数据库中的每个图像，提取SIFT关键点，并用SIFT和色调描述子描述其周围的局部区域。然后利用k-均值聚类对图像的描述空间进行量化，生成两个独立的码本。码本生成中的硬分配会导致量化损失，从而降低检索性能。为了克服这一不利影响，该方法采用软赋值代替硬赋值。实验结果表明，本文提出的图像检索方法与其他同类方法相比具有更好的性能。