

摘要

随着移动互联网的发展、手机的普及以及数据的增长，移动视觉搜索正在逐渐变革移动互联网的搜索体验，将视觉信息作为查询请求已经逐渐取代文字查询请求成为解放用户双手的重要交互模式。移动视觉搜索作为一个新兴的研究领域，催生了一些新的问题和挑战。虽然近些年的一些代表性工作已经开始着手解决移动视觉搜索中的一些相关问题，但面对当前更为复杂多变的移动互联网环境，移动视觉搜索技术仍在区分力、可伸缩性、复杂度等方面仍面临着问题与挑战：

第一，图像特征表达（即图像描述子）的区分力有限。由于移动端在内存、计算能力等方面相对欠缺，一般在移动端对图像进行特征表达时，会采用一些复杂度相对低的特征提取手段（比如采用小码本构建聚合描述子），但这也导致了图像特征表达的区分力有限。

第二，图像特征表达的可伸缩性有限。一般而言，视觉搜索采用的码本、特征都是定长的表达。对于移动客户端与服务端存在交互的、不对称的视觉搜索场景，需要更为灵活的且具备可伸缩性的图像特征表达。

第三，查询图像的压缩效率有限。除了通过客户端发送压缩描述子，另外一类移动视觉搜索框架通过传输压缩图像的形式完成查询意图的传输。然而当前的压缩图像方法并未针对视觉搜索设计，压缩率和压缩图像的区分力均有所不足。

为此，针对当前存在的问题与挑战，本文围绕面向移动视觉搜索的图像特征表达展开研究，提出了如下几个创新点：

首先，针对聚合描述子区分力有限的问题，本文提出了基于多级码本的层次化级联聚合描述子的结构。面向移动视觉搜索中的典型聚合描述子如 VLAD、Fisher Vector，码本规模固定，且单词数量多设定为较小的量级，描述子的区分力有限；本文提出的方法首先通过多级码本及层次量化加速局部描述子的量化。之后，针对每层码本分别构建聚合描述子，并将聚合描述子级联形成层次化级联聚合描述子结构。通过上层描述子缓解下层描述子的量化误差，以提高描述子区分力。本方法充分利用了层次量化的速度优势以及大码本量化粒度精细的优势，使得最终可以在较低的计算复杂度快速完成高区分力的聚合描述子提取。

其次，针对聚合描述子可伸缩性有限的问题，本文提出了基于多级码本的非对称聚合描述子，使得不同长度的聚合描述子之间可以匹配。针对移动视觉搜索场景中，为了解决计算能力较弱的客户端与计算能力较强的服务端之间存在非对称性的问题，本方法利用了多级码本之间的层级联系，使得搜索系统可以在不同的资源约束下生成不同

长度的聚合描述子：计算资源欠缺的设备可以在较低的复杂度下生成较短的聚合描述子，而计算资源较充足的设备则在较高的复杂度下生成更长的聚合描述子（同时获得更好的检索效果）。同时，利用码本之间的层级关系，不同长度的描述子均可以和服务器端的描述子完成匹配和检索。本方法通过利用多级码本的层次关系和非对称性，使得码本的尺寸、描述子长度均具备可伸缩性，能够满足更为多变的应用需求。

最后，针对查询图像的压缩效率有限的问题，本文提出了基于优化量化表和重置 DCT 量化系数阈值的查询图像压缩方法，使得在保持检索精度的前提下以极低的复杂度压缩查询图像的尺寸，从而减少上行带宽传输开销，进一步提高面向传输图像视觉搜索的用户体验。基于优化量化表和重置 DCT 量化系数阈值的查询图像压缩方法基于经典的 JPEG 图像压缩标准优化，具备复杂度低、保持标准压缩码流结构的特点。

本文提出的方法有效解决了现有移动视觉搜索存在的问题，进一步在实现了视觉搜索在复杂度、区分力、可伸缩性上的权衡。结合本文对提出创新点，本文还针对实际应用场景进行了原型系统的设计、开发与测试，进一步验证了本文提出创新点的有效性。

关键词：移动视觉搜索，聚合描述子，多级码本，图像压缩，JPEG