

# 移动视觉搜索技术与标准化进展

段凌宇 黄铁军 高 文

北京大学信息科学技术学院数字媒体研究所 北京 100871

**摘 要** 移动视觉搜索成为未来移动世界中有影响的基础技术之一。文章介绍了移动视觉搜索面临的技术挑战,探讨了包括紧凑视觉描述子、视觉检索流程、检索系统互操作性等关键技术。围绕紧凑视觉描述子概述了移动视觉搜索国际标准的工作进展,提出建设大规模视觉对象数据集的重要意义。

**关键词** 视觉搜索; 移动互联网; 紧凑视觉描述子; 移动智能终端

## 概述

根据国际电联(ITU)公布的数据,2010年全球移动互联网用户数为8.65亿人(与之对应,全球互联网用户总数为20.8亿人)。其中,中国有3.03亿人,美国为1.15亿人,日本计0.86亿人。中国互联网络信息中心(CNNIC)发布《第30次中国互联网络发展状况统计报告》显示,2012年上半年,通过手机接入互联网的网民数量达到3.88亿,超出台式电脑的3.80亿,成为我国网民的第一大上网终端。ITU预计,到2015年全球使用移动互联网的人数将超过桌面互联网。伴随移动互联网的高速发展,人们对移动性与多媒体信息的需求急剧上升。越来越多的人希望在移动中接入互联网,获取急需的信息。

随着移动互联网产业的迅速崛起,越来越多的IT厂商推出移动智能终端(如智能手机、平板电脑等)登陆移动互联网。移动智能终端在计算性能上有了飞速发展并保持持续增长。此外,智能终端携带了丰富的传感器,如摄像头、GPS、重力感应器、电子罗盘。通过移动终端以及运行其上的强大应用,人们可以在现实世界与信息世界之间建立关联,从而便捷地获取全面的多媒体信息及服务,比如基于位置的多样化服务。

毫无疑问,移动视觉搜索技术将成为支撑未来移动

互联网应用的基础技术之一。如图1所示,通过移动智能终端与视觉搜索技术的有效结合,人们可以快速便捷地采集现实世界的视觉对象,从移动互联网另一端获取感兴趣的关联信息。



图1 移动视觉搜索系统基本架构

若将移动视觉搜索与增强现实技术有机结合起来,以移动智能终端为平台的“所见即所知”式的增强现实应用将可能成为继搜索引擎之后的互联网新一代服务模式。例如手机拍摄到燕园博雅塔时,互联网上的相关文字内容、图片、视频、甚至三维几何模型将精确地叠加显示在移动终端屏幕上;若与移动位置服务结合,用户打开手机摄像头就可以查看周围商户及餐馆的信息;如用户开启增强现实客户端后,将手机摄像头对准景物,增强现实软件将对屏幕内的景物信息进行分析处理,随即提供关于该景物及周边地区的旅游信息等;利用摄像头对准书页,获得大家的书评;在火灾等重大事故发生现场,救援人员开启增强现实移动终端,获得建筑三维框架图的增强呈现。

视觉搜索研究领域已经有大量算法和技术方案可

基金项目: 国家“973”计划基金资助项目(2009CB320902); 国家自然科学基金(60902057)

供选择,人们通常经验性地从中挑选若干模块去构建系统。然而,多样化的不同设计架构下的视觉搜索系统,必然引发一个互操作性的问题。如果移动视觉搜索使用图像视觉特征而不是以图像作为交换信息,如何使得这些应用程序和图像数据库兼容于广泛的设备与平台?交互数据规范化需求日益突出。由于移动视觉搜索技术在手机平台拥有巨大的潜在价值,移动互联网终端厂商、芯片厂商以及各类行业巨头均对这项技术给予了重视,他们正积极推动该技术的发展。无疑,多样化的视觉搜索应用系统挑战互操作性,移动视觉搜索应用程序的标准化成为亟待解决的技术瓶颈。

本文除介绍移动视觉搜索技术与标准化进展外,还将探讨移动视觉搜索技术推广应用将面临的大规模视觉对象数据采集等瓶颈问题。

## 1 移动视觉搜索技术

移动视觉搜索是指利用移动终端获取真实世界中对象的图像或视频作为查询对象,通过移动互联网检索视觉对象关联信息的检索方式。

视觉搜索赢得了学术界和工业界的广泛关注和研发努力,然而实际应用中的视觉搜索系统存在查询性能不理想或者不稳定的问题,任何一次糟糕的搜索结果都将损害用户体验。围绕服务端,大量研究工作致力于如何有效提高视觉检索性能。另一重要方面,为迎接无线网络环境的带宽有限、带宽波动、时延和获取性等方面挑战,移动视觉搜索必须考虑视觉查询的紧凑表示。发送图像的紧凑视觉签名,通常要比传送整张图像更有优势,尤其是在带宽有限的无线网络环境下,大数据量传输造成的上行查询传输延迟,直接影响用户体验。围绕移动智能终端,针对视觉查询进行紧凑表示,在不损失基于原始图像的检索性能前提下,亟待突破特征提取速度慢、内存开销大、数据不紧凑的瓶颈;这也是多媒体信号处理、图像分析与检索等多个交叉研究领域共同关注的问题。

以下将从紧凑视觉描述子和视觉搜索流程两方面介绍移动视觉搜索技术研究。

### 1.1 紧凑视觉描述子

描述图片是计算机视觉领域的一个基本挑战。除优化视觉描述子以更好地应用于图像视频内容分析与理解外,研究还需要关注描述子压缩问题。如上所述,发送一幅图像的紧凑视觉签名,往往要比传送整张图像更有优势,尤其是在带宽有限的无线网络环境下。此外,压缩后的描述子可以避免视觉对象分类器训练过程中经常面临的高维特征带来的负面影响,尤其当面对大规模视觉对象分类和识别问题时,紧凑描述子对百万量级以上的图片或视频的可扩展存储、处理、索引发挥着至关重要的作用。

图像特征表示是视觉搜索的关键环节,局部特征描述子得到广泛的研究,其中尺度不变描述子<sup>[1]</sup>(Scale Invariant Feature Transform, SIFT)是被广泛应用的一种。图2给出基于SIFT局部描述子的图像对匹配实例,左边为查询图像,右边为参考图像,黄色圆圈代表自动检测出来的尺度不变描述子,连接两幅图像中局部描述子集合的红色线条代表通过最近邻准则检测出来的匹配局部描述子。如图2所示,分别从查询图像与参考图像提取局部特征,基于局部特征进行最近邻比对,然后针对匹配上的特征对集合进行几何一致性校验,最终判断两幅图像是不是匹配对。尽管CD封面图像的尺度发生很大改变,存在严重旋转,但基于SIFT描述子仍然完成精确的匹配。围绕移动视觉搜索应用,近期研究工作更多地关注描述子压缩问题,对描述子进行深度压缩的同时尽可能地维持甚至提升描述子本身的区分力。

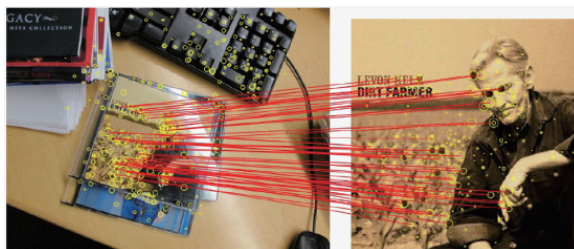


图2 基于局部描述子SIFT的图像对匹配实例

局部描述子压缩的相关研究工作可以划分为三类。第一类采取降维。代表性工作有PCA-SIFT<sup>[2]</sup>,提出在梯度块(Gradient Patches)上做主成分分析(PCA),只

保留方差较大的维度。也可采用线性判别分析(Linear Discriminant Analysis, LDA)<sup>[3]</sup>利用监督信息将描述子映射到低维空间,匹配的描述子之间保持近距离,而非匹配的描述子被分离开来。第二类采用二进制码表示图像。Torralba等人<sup>[4]</sup>利用机器学习技术将GIST描述子转化为紧凑的二进制码,这是一类比较流行的满足大规模图像搜索需求的描述子压缩方法。第三类以量化为基础,其中最具代表性的工作是Chandrasekhar等人提出的梯度直方图描述子压缩(Compressed Histogram of Gradient, CHoG)<sup>[5]</sup>,其核心思想是使用霍夫曼树来压缩量化后的梯度直方图表达,使得一个兴趣点的局部特征只要60比特来表达。

基于图像层面的视觉签名压缩方法有两类。第一类采取局部描述子集成。代表性工作有VLAD(Vector of locally Aggregated Descriptors)<sup>[6]</sup>。另一类采取词汇包直方图压缩。Ji等人<sup>[7]</sup>引入机器学习手段,提出利用上下文信息(如GPS、标签等)的单词选择方法,建立位置自适应的紧凑视觉词典,并将其应用于移动地标搜索,单张查询只需约100比特。

面向视觉搜索,紧凑描述子以及搜索流程标准化工作MPEG CDVS(Compact Descriptors for Visual Search)<sup>[8]</sup>已经确定建立在视觉特征的局部描述与全局描述之上的紧凑视觉描述,如图3所示,融合了局部描述子压缩的图像对匹配优势及图像全局签名的图像检索优势。图像对匹配也作用于图像检索的重排序阶段。MPEG CDVS的标准参考软件集成的全局图像签名采纳了北京大学提出的全局描述子SCFV(Scalable Compressed Fisher Vector)<sup>[9]</sup>,局部描述子压缩采用两套方案,即标量量化器与矢量量化器,矢量量化器采纳了北京大学提出的MSVQ(Multi-Stage Vector Quantizer)<sup>[10]</sup>。

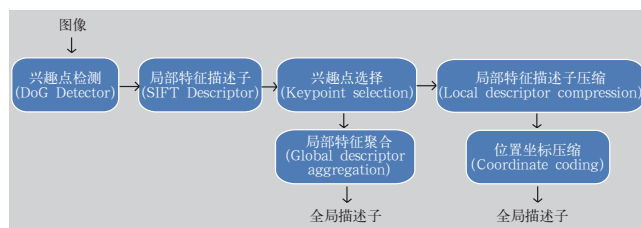


图3 紧凑视觉描述子提取流程

此外,图像紧凑视觉描述子必须保存局部特征的位置坐标信息。图像对匹配与检索,一方面依赖局部视觉描述子,另一方面依赖特征点的空间位置信息。正如图2所示,图像对匹配需要局部特征的位置信息完成几何一致性校验。面向移动视觉搜索,有必要针对兴趣点位置坐标编码。任意一张VGA图像(分辨率为640x480)的兴趣点位置坐标在不压缩的情况下平均占用约20比特,若图像有500个兴趣点,那么仅坐标位置信息就需要大约1250字节,几乎是MPEG CDVS设定的最低操作点(单幅查询大小)512字节<sup>[8]</sup>的三倍;因此,局部特征的位置坐标压缩也是紧凑视觉描述子的重要环节,提供降低查询传输延迟的技术手段。当前MPEG CDVS参考模型采用了斯坦福大学提供的有损压缩算法,单个兴趣点的坐标信息编码约需6比特<sup>[11]</sup>。

## 1.2 视觉搜索流程

视觉搜索性能还有很大提升空间,它在计算机视觉领域方面也有很大的技术挑战性。为了突破基于视觉特征实现对象精准搜索的瓶颈,需要广泛探索如何结合实际系统的应用环境以及特定需求,优化视觉搜索性能。举例来说,结合移动设备携带的丰富上下文信息辅助视觉搜索,提升性能与效率<sup>[7]</sup>;研究高效的基于几何一致性校验的重排序方法<sup>[11]</sup>;研究全局描述子与局部描述子的融合算法<sup>[9]</sup>;如何利用主题模型提升搜索性能等<sup>[12]</sup>。

此外,针对视觉查询的多样性,准确有效地预测查询的视觉搜索性能也是一项有意义的技术突破<sup>[13]</sup>。具体来说,提醒用户放弃或者改善视觉查询,可以较大程度尊重用户的搜索意图。另一方面,移动视觉搜索应用中,发送一次图片会导致查询延迟以及带宽开销,移动终端执行视觉检索性能预测可以避免不必要的无线带宽开销,降低资费,从而大幅提升用户体验。

基于视觉搜索领域的最新研究现状,确立一套规范的搜索流程具有重要的意义。一方面,有利于缩短移动视觉搜索最新研究成果与应用技术开发的距离;另一方面,有利于提升多样化视觉检索系统之间的互操作性。以下介绍视觉搜索的通用框架和基本流程。值得一提



的是, 下列图像对匹配流程和检索流程, 已经被MPEG CDVS<sup>[11]</sup>标准化工作组采纳。因为篇幅限制, 上述优化检索性能的研究工作无法一一展开, 部分研究工作将会在以下流程介绍中体现。

1) 视觉搜索架构。视觉搜索应用系统涉及两项基本任务: 图像对匹配(Pairwise Matching)与图像检索(Retrieval)。前者针对两张图像进行自动校验, 确定是否包含相同对象或场景。后者是在大规模图像数据集中搜索和发现包含与查询图像相同的对象或场景的所有图片。为完成上述任务, 搜索系统从查询图像以及参考图像中提取特征描述子。图4(a)、4(b)分别给出图像对匹配以及图像检索的一般性框架, 提高图像对匹配性能与效率是一项重要的研究工作。面向大规模图像数据库, 如何使得搜索架构可扩展也是一项重要工作。此外, 匹配与检索均建立在描述子之上, 考虑到紧凑描述子的使用, 如何基于压缩后的局部描述子紧致全局签名, 利用数据压缩带来的特性, 提高匹配与检索的计算效率, 也是有意义的课题。比如, 通过量化后的特征进行查表计算, 可以大大提升特征比对的效率; 因为高效的图像对匹配, 同等时间约束下, 可以完成更多图像的几何一致性校验, 从而大大提升检索性能。

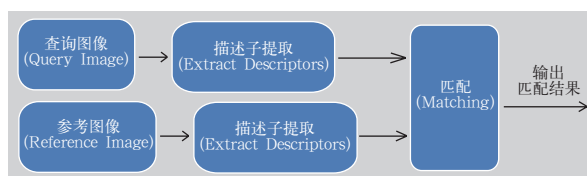


图4(a) 图像对匹配架构

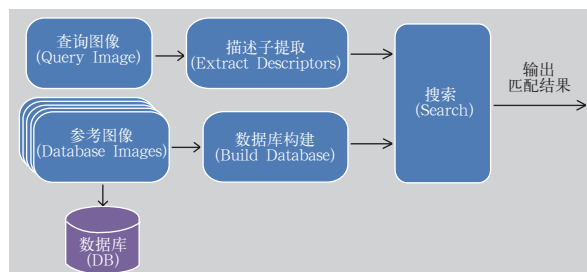


图4(b) 图像检索架构

图4 图像对匹配及图像检索架构

2) 图像对匹配流程。图像对匹配首先确定查询图片与参考图片是否描述相同物体或者场景。若匹配, 系

统将输出定位信息, 即匹配物体在图像中的位置。图5给出基于紧凑描述子的图像对匹配流程, 局部描述子与全局描述子共同参与图像对匹配计算。

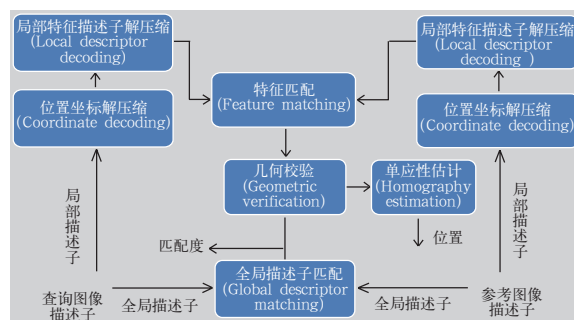


图5 图像对匹配流程

匹配流程可以划分为两个阶段, 即第一段的基于局部描述子匹配和第二阶段的基于全局描述子匹配。①局部描述子匹配阶段。首先确定查询图像与参考图像中的兴趣点匹配对(Key Point Matches or Correspondences); 然后通过几何一致性校验, 确定兴趣点匹配对的内群点(Inliers), 若内群点超过设定的阈值, 该图像对将被视为匹配对; 最后, 通过单应性(Homograph)估计, 针对检测出的匹配对, 计算匹配物体的位置。②全局描述子匹配阶段。通过全局描述子的相似性计算, 检测出第一阶段的基于局部描述子匹配无法界定的匹配对。第一阶段采用了严格的几何一致性校验, 图像匹配对的误检率非常低, 所以第二阶段主要任务是区分未检测出的匹配对与绝大多数的非匹配对。

不同于局部描述子, 全局描述子是建立在图像层面的视觉特征的高阶统计特性。面向视觉搜索, 高阶统计特性使得全局描述子更具区分力。

上述两阶段图像对匹配流程充分发挥了局部特征表达以及全局特征聚合的优势。前者采用几何约束的参数模型, 并提供定位功能; 后者采用数据驱动统计模型, 提供更加出色的图像层面的区分力。容易理解, 严格几何约束带来更加精准的匹配对检测率(降低误检率), 而统计模型可以获得更加鲁棒的图像表达能力, 提供更加丰富的特征用于图像识别与分类。

3) 图像检索流程。图像检索的目标是从大规模图

像数据集中快速准确地搜索和发现包含与查询图像相同的对象或场景的所有图片。若不考虑时间约束,可以将查询图像与图像库中的所有参考图像逐一比对,然后根据相似性匹配数值进行排序,返回检索到的图片序列。在大规模图像检索的实际应用中,检索效率是一个不容忽视的指标。从图6中看出图像检索流程分为两个阶段,即第一阶段的基于全局描述子的检索和第二阶段的基于局部描述子的图像重排序阶段。通过第一阶段的“粗筛选”与第二阶段的“细排序”,实现图像检索系统在检索效率与检索性能之间的平衡。

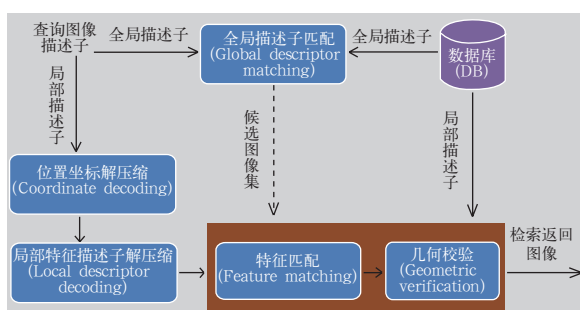


图6 图像检索流程

全局描述子检索阶段,查询图像的全局特征与图像数据集中预先计算好的参考图像全局特征进行逐一比对,通过相似性匹配值排序,返回一组候选图像(Top Matches)。全局描述子通常支持快速运算(比如,汉明距离计算)<sup>[9]</sup>,适合大规模图像数据集的基于高效线性查找的图像比对。对比传统的词汇包倒排索引模型,全局描述子检索可以通过返回较少的候选图像集合,获得较高的召回率(Recall Rate)。此外,基于全局特征直接比对的线性查找方式可以避免索引建立过程,更好地支持参考图像数据集的在线更新。

图像重排序阶段,检索第一阶段返回的候选图像接受几何一致性校验。根据几何校验数值,候选图像序列重新排序,满足几何一致约束的图像将被调整至前列。因为基于全局描述子的检索大大缩小了候选匹配图像集合,查询图像与参考图像的几何一致性校验只在少量图像对进行,图像重排序阶段的时间复杂度大大降低。

## 2 视觉搜索技术标准化

如上文所述,多样化的视觉搜索应用系统挑战互操作性,移动视觉搜索应用程序的标准化成为亟待解决的技术瓶颈。国际标准化工作开展得如火如荼,致力解决多样化应用之间以及应用与后台数据库之间的互操作问题。突飞猛进的移动搜索、增强现实研究与应用及日趋庞大的移动互联网用户,使多家科技巨头正大力推动移动增强现实产业进入快速发展阶段,手机芯片厂商对这项技术给予了高度重视。移动搜索及增强现实技术的国际标准化工作将极大地促进硬件厂商移动互联网应用芯片的开发与使用。

国际标准化组织(ISO)和国际电工委员会(IEC)下属的动态图像专家组(Moving Picture Experts Group)正在开展的紧凑视觉描述子(Compact Descriptors for Visual Search)<sup>[8]</sup>的标准制定将使芯片等硬件厂商直接受益。符合标准规范的硬件保证视觉搜索应用程序及数据库的互操作性;保证与标准一致的高性能实现;简化视觉搜索应用中描述子提取和图像匹配的设计;保证移动终端的硬件能够支持描述子提取和图像匹配过程;降低无线网络传输视觉搜索相关信息带来的网络负载。为开发完整的视觉搜索应用,这一标准有可能与其他现有标准进行结合,如MPEG QUERY FORMAT、HTTP、XML、JPEG以及JP Search。

关于紧凑视觉描述子标准化的必要性,在2009年斯坦福大学主办的第一次移动视觉搜索研讨会上被首次提出,2011北京大学主办的第二次移动搜索研讨会再次就相关的关键技术展开讨论。在一系列来自工业界与学术界的技术讨论和需求驱动下,MPEG美国代表团正式向标准委员会提交提案,将移动视觉搜索应用程序的标准化工作作为一个潜在兴趣点进行探索。最终,一个对未来移动视觉搜索技术标准的探索活动拉开序幕,并在随后的Ad-hoc会议上产出了一系列文档,对标准所涉及的应用程序、用例、目标、范围和要求进行了描述。表1列出MPEG CDVS标准化工作的时间

表和进程。北京大学数字视频编解码技术国家工程实验室(NELVT)、美国斯坦福大学(Stanford University)、英国萨里大学(Surrey University)、意法半导体公司(STMicroelectronics)、华为慕尼黑(Huawei Munich)、美国高通公司(Qualcomm)、诺基亚(Nokia)、意大利电信集团(Telecom Italia)、日本电气公司(NEC)、韩国三星电子(Samsung)等多家单位参与了MPEG CDVS视觉搜索国际标准制定,围绕标准测试模型、核心试验,有重点地突破技术瓶颈,为标准制定提供全面的技术选择依据。

表1 MPEG CDVS标准化工作进程

MPEG会议	日期	关键点
97 <sup>th</sup>	2011,7	Call for Proposals
98 <sup>th</sup>	2011,12	Proposals Evaluation,Test Model Determined
99 <sup>th</sup>	2012,2	Six Core Experiments Set Up
100 <sup>th</sup>	2012,4	Evaluation of Proposals
101 <sup>st</sup>	2012,7	First Working Draft
103 <sup>rd</sup>	2013,1	Committee Draft
105 <sup>th</sup>	2013,7	Draft of International Standard
107 <sup>th</sup>	2014,2	Final Draft of International Standard

MPEG CDVS标准化进程中,总共设立了八项核心技术试验,分别是CE1全局特征、CE2局部描述子压缩、CE3特征点位置压缩、CE4兴趣点检测、CE5局部描述子、CE6检索流程、CE7特征选择、CE8融合局部与全局特征的图像对匹配流程。CE1~CE5、CE7直接

影响紧凑视觉描述子的码流(Normative Part),CE6、CE8提供描述子的优化使用方法。

2012年2月确立核心技术试验以来,经历三轮会议周期的技术提案竞争已被专家组确立为符合标准需求的核心技术,并计划在MPEG 103次会议进入标准的Committee Draft阶段。结合第二节介绍的紧凑描述子与检索流程,表2列出部分核心实验的技术采纳情况。但尚未列出CE4兴趣点检测技术,该项核心试验仍在紧张进行。开展CE4实验的主要目的是为了规避标准可能涉及的SIFT检测子的专利问题。另外,CE5局部描述子技术仍然沿用SIFT描述子。

整体而言,MPEG CDVS标准化工作严格按照预设的时间点稳步推进。经过专家组一年多的努力,视觉搜索核心技术在性能、效率、复杂度等方面均得到显著提升。在视觉搜索领域的多个公开基准数据集上,CDVS技术均以显著优势超越当前文献报告的最好结果。

### 3 视觉对象数据库建设

视觉搜索服务依赖于视觉对象的高精度匹配和识别。基于高度结构化的对象数据集,人们可以快速便捷地将现实世界的视觉对象关联至信息世界的视觉对象,从而准确获得移动互联网另一端的关联信息,提供“所见即所知”的新一代搜索服务。

表2 MPEG CDVS标准化工作采纳的核心技术

核心实验	技术名称及提案文档编号	技术特点	技术贡献方
CE1: 全局特征	Scalable Compressed Fisher Vector(SCFV) [m26726]	特征紧凑性高、大小可伸缩、区分力强、特征提取复杂度低、基于汉明距离的图像匹配计算复杂度低	北京大学
CE2: 局部描述子压缩	Transform+Scalar Quantizer [m25929], Multi-stage Vector Quantizer(MSVQ) [M26727]	内存开销小、大小可伸缩、提供标量量化与矢量量化两套压缩方案	北京大学、视觉原子(Visual Atoms)、希思卫(Sisvel Technology)、萨里大学(Surrey University)、意大利电信
CE3: 特点点位置压缩	Context based location coding [m25883]	位置坐标有损压缩,损失的坐标精度对图像匹配以及定位带来的影响可忽略	斯坦福大学、华为慕尼黑
CE6: 检索流程	Retrieval using SCFV [m26726], Retrieval using bag-of-features [N12367,m24783]	面向大规模图像数据集,提供高度可扩展的检索方案;无须建立图像索引,计算复杂度低,适合低端硬件平台	北京大学、意大利电信
CE7: 特征选择	Bayesian learning based feature selection [N12367]	过滤噪声特征,大大降低了图像质量以及特征噪声对紧凑视觉描述子的负面影响	意大利电信
CE8: 图像对匹配流程	DISTance RATio coherence(DISTRAT) [N12367], Scalable Compressed Fisher Vector based global matching [m26726], Weighted Hamming Distance [m26726,m24757]	快速几何一致性校验算法、融合全局签名与局部特征的高性能图像匹配算法、基于机器学习的优化汉明距离,提升了匹配准确率	北京大学、意大利电信、斯坦福大学

如何有效地采集、分析、管理现实世界的大规模视觉对象数据集是一个挑战性课题。比如,如何从互联网上发现相关图片,并从中过滤有价值的图片;针对某个垂直领域,如何高效地建立具有精确标注的对象数据集,支持高质量的信息关联;面向海量图像数据处理,如何建立高性能数据存储平台,支持图像数据的高效存储、检索、更新。

近期,北京大学与新加坡南洋理工大学启动了为期五年的高校国际合作项目“博云搜索”(Rapid-Rich Object Search Lab),该项目将致力于解决移动视觉搜索应用面临的对象数据集挑战,主要目标是建立亚洲最大的结构化对象库,开发快速丰富的移动视觉搜索应用。项目执行期间,以实验中心为平台,支持关键技术的基础性研究开发,并将广泛联合中国与新加坡的搜索引擎及互联网企业、通信运营商,促进移动搜索技术的商品化和产业化。

#### 4 总结与展望

随着移动互联网、智能移动终端以及云计算的迅速崛起,移动视觉搜索成为未来移动世界中有影响的基础技术之一。人们通过移动视觉搜索技术,可以快速便捷地采集现实世界的视觉对象,从移动互联网获取感兴趣的关联信息。本文介绍了移动视觉搜索面临的技术挑战,概述了移动视觉搜索国际标准的工作进展,提出建设大规模视觉对象数据集的重要意义。

同时,数据分享困难日益成为移动搜索及增强现实业务实现产业化的主要瓶颈。研发此类应用技术与系统,必须借助产业与大学紧密合作,联合互联网领域的软件技术提供商和平台运营商,发挥其构建海量关联信息以及规范化知识体系的数据与平台优势。

中国在互联网领域的企业影响力和技术水平正在飞速前进。世界正在目睹一大批中国互联网企业的腾飞。继Web2.0时代互联网应用逐步满足了人们对于互动性、即时性、个性化需求的同时,中国的移动互联网行业正在呼唤崭新的杀手级应用模式,提升下一代搜索引擎核心技术的国际竞争力。视觉搜索以及移动增强现实应用

便在其中,还原互联网的真实感,让互联网跟真实世界的关系更紧密。在移动互联网产业大背景下,如何借助Web3.0的东风,在虚拟现实与数字媒体前沿领域寻求基础性突破,亟需学术界、产业界同仁的共同努力。

#### 参考文献

- [1] Lowe D G. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004, 60(2): 91-110
- [2] Ke Y, Sukthankar R. PCA-SIFT: A more distinctive representation for local image descriptors[C]// IEEE Conf. Computer Vision & Pattern Recognition, 2004
- [3] Brown M, Hua G, Winder S A J. Discriminative learning of local image descriptors[J]. IEEE Trans. Pattern Analysis and Machine Intelligence, 2011, 33(1): 43-57
- [4] Torralba A, Fergus R, Weiss Y. Small codes and pattern recognition[C]// IEEE Conf. Computer Vision & Pattern Recognition, 2008
- [5] Chandrasekhar V, Takacs G, Chen D, et al. CHoG: Compressed histogram of gradients[C]// IEEE Conf. Computer Vision & Pattern Recognition, 2009
- [6] Jegou H, Douze M, Schmid C, et al. Aggregating local descriptors into a compact image representation[C]// IEEE Conf. Computer Vision & Pattern Recognition, 2010
- [7] Ji R, Duan L Y, Chen J, et al. Location discriminative vocabulary coding for mobile landmark search[J]. International Journal of Computer Vision, 2012, 96(3): 290-314
- [8] CDVS. Call for Proposals for Compact Descriptors for Visual Search[S]. ISO/IEC JTC1/SC29/WG11/N12201, Torino, IT, July 2011
- [9] Lin J, Duan L Y, Chen J, et al. Peking Univ. Response to CE1: A scalable low-memory global descriptor[S]. M26726, ISO/IEC JTC1/SC29/WG11, Shanghai, China, Oct. 2012
- [10] Chen J, Duan L Y, Lin J, et al. Peking Univ. Response to CE 2: Local descriptor compression[S]. M26727, ISO/IEC



- JTC1/SC29/WG11, Shanghai, China, Oct. 2012
- [11] CDVS. Test Model 4: Compact descriptor for visual search[S]. W12929, ISO/IEC JTC1/SC29/WG11, Shanghai, China, Oct. 2012
- [12] Ji R, Duan L Y, Chen J, et al. Towards compact topical descriptors[C]//IEEE Conf. Computer Vision & Pattern Recognition, 2012
- [13] Li B, Duan L Y, Chen Y, et al. Predicting the effectiveness of queries for visual search[C]//IEEE Conf. Acoustic, Speech, and Signal Processing, 2012

---

### 作者简历

---



段凌宇

副教授，主要研究方向为多媒体分析与检索、移动多媒体计算、计算机视觉。



黄铁军

教授，主要研究方向为数字媒体编码与理解、数字图书馆、模式识别与智能交互。



高文

教授，中国工程院院士。主要研究方向为多媒体数据压缩、图像处理、计算机视觉、多模式接口和人工智能。

### Technical Research and Standardization in Mobile Visual Search

Duan Lingyu

Huang Tiejun

Gao Wen

School of Electronic Engineering and Computer Science, Peking University, Beijing 100871, China

**Abstract** Mobile visual search has become one of the influential supporting technologies for the future mobile world. In this article, we introduce a few technical challenges in mobile visual search, discuss some key technologies including compact visual descriptors, visual search pipeline, interoperability of visual search applications and databases. The standardization progress in compact descriptors for visual search is discussed, and the significance of building up the large-scale structured visual object dataset is proposed as well.

**Keywords** Visual Search; Mobile Internet; Compact Visual Descriptors; Mobile Smart Terminals

---