计算机研究与发展
Journal of Computer Research and Development
DOI:10.7544/issn1000-1239.2017.20170001
54(6):1225-1237,2017

高通量图像视频计算

唐金辉[1] 李泽超[1] 刘少礼[2] 秦 磊[2]
[1](南京理工大学计算机科学与工程学院 南京 210094)
[2](中国科学院计算技术研究所 北京 100190)
(jinhuitang@njust.edu.cn)

High-Throughput Image and Video Computing

Tang Jinhui[1], Li Zechao[1], Liu Shaoli[2], and Qin Lei[2]
[1](*School of Computer Science and Engineering*, *Nanjing University of Science and Technology*, *Nanjing* 210094)
[2](*Institute of Computing Technology*, *Chinese Academy of Sciences*, *Beijing* 100190)

Abstract In recent years, image and video data grows and spreads rapidly in the Internet. The data not only has huge amount, but also has the characteristics of high concurrency, high dimension and high throughput, which brings huge challenges into the real-time analysis and processing of them. To promote the image and video data processing efficiency of big data platforms, it is necessary and important to study the task of high-throughput image and video computing, and propose a series of high-throughput image and video computing theories and methods by considering the new hardware structures. Towards this end, this work first overviews previous high-throughput image and video computing theories and methods in details, and then discusses the disadvantages of the existing high-throughput image and video computing methods. Furthermore, this work analyzes three research directions of the high-throughput image and video computing task in future: the high-throughput image and video computing theories, the high-throughput image and video analysis methods, and the high-throughput video coding methods. Finally, this work introduces three key scientific problems of high-throughput image and video computing. The solutions of these problems will provide key technical support for the applications of content monitoring of Internet images and videos, the large-scale video surveillance, and the image and video search.

Key words image analysis; video analysis; high throughput; video coding; visual computing

摘 要 互联网上的图像和视频数据正在飞速地产生和传播.这些数据不仅规模庞大,还具有高并发、高维度、大流量的显著特性,导致了目前对它们的实时分析和处理面临着巨大的挑战.这就需要开展高通量图像视频计算方面的研究,需要结合新型硬件结构,利用其体系结构优势,提出一系列实用的高通量图像视频计算理论与方法,提升数据中心的图像视频数据处理效率.为此,在详细地分析了现有的高通量图像视频计算相关方法与技术的基础上,探讨了现有高通量图像视频计算方法研究的不足;进一步

收稿日期:2017-01-03;**修回日期**:2017-03-07
基金项目:国家"九七三"重点基础研究发展计划基金项目(2014CB347600);国家自然科学基金项目(61402228);国家自然科学基金优秀青年科学基金项目(61522203)
This work was supported by the National Basic Research Program of China (973 Program) (2014CB347600), the National Natural Science Foundation of China (61402228), and the National Natural Science Foundation of China for Excellent Young Scientists (61522203).
通信作者:李泽超(zechao.li@njust.edu.cn)

地,分析了高通量图像视频计算的3个未来研究方向:高通量图像视频计算理论、高通量图像视频分析方法及高通量视频编码方法.最后,总结了高通量图像视频计算需要解决的3个关键科学问题.这些问题的解决将为互联网图像视频内容监管、大规模视频监控、图像视频搜索等重要应用提供关键技术支持.

关键词 图像分析;视频分析;高通量;视频编码;视频计算

中图法分类号 TP391

近年来,图像和视频数据正在以前所未有的速度不断地产生和传播,已成为这个时代真正的大数据.它的两大特点是大容量和高并发.高并发意味着单位时间内产生的请求或任务的数量大.无论对国家公共安全还是日益增长的互联网经济来说,如何对这些具有高并发性的海量图像视频数据进行实时高效的分析和处理,已成为一个亟待解决的重要问题.

高通量图像视频计算就是高效地处理大容量和高并发的图像视频数据.目前已有相关工作开始关注高通量图像视频计算,比如图像视频并行计算方法、图像视频的多级计算模型等.然而,这些模型与方法大多是关注如何提高单个图像视频分析和处理的精度及速度,而较少关注高并发环境下的系统吞吐能力和高通量化研究.因此,为了满足日益增长的高通量图像视频计算需求,这就需要结合新型硬件结构,利用其体系结构优势,提出一系列实用的高通量图像视频计算理论与方法,提升数据中心的图像视频数据处理效率.为此,本文重点关注大容量和高并发图像视频数据的低延迟、高精度计算方法,在详细分析了现有的高通量图像视频计算相关方法与技术的基础上,进一步讨论了现有高通量图像视频计算方法研究的不足,最后分析了高通量图像视频计算的未来研究方向及需要解决的科学问题.

本文的主要贡献有3点:

1) 详细阐述了高通量图像视频计算的相关研究现状;

2) 分析了现有高通量图像视频计算方法的不足;

3) 提出了高通量图像视频计算的未来研究方向及需要解决的科学问题.

1 背景和意义

在网络空间中,海量的网络用户时刻在创造大量的图像视频数据,例如YouTube视频分享网站的每分钟上传视频长度约为60 h,每日用户观看量超过30亿次;另一方面,24 h不断更新的监控视频数据也是海量视频的一个重要来源,例如北京奥运期间就安装了30万台摄像头,而英国伦敦2012年奥运会则安装了50多万台;此外,随着移动智能终端的拍摄与分享功能的不断增强,移动图像视频搜索等新型应用也面临着数量惊人的数据.这些图像视频数据不仅规模庞大,更重要的是还具有高并发、高维度、大流量的显著特性.比如,在互联网视频内容监管中,流量通常高达每秒几GB甚至几十GB.不幸的是,目前的计算机系统以及图像视频分析技术,都无法应对大容量和高并发带来的挑战.

首先,从计算机系统的角度来说,传统的高性能计算机追求单个并行应用的性能,主要用于科学计算.而高通量计算机是适用于互联网新兴应用负载特征、在强时间约束下处理高吞吐量请求的一种高性能计算机,采用以低成本、高扩展和集中的硬件、软件系统栈处理高并发负载的数据中心计算机系统.高通量计算在结构特征、资源管理、调度策略等方面非常类似于城市交通管理,因为两者的共同特征是单位时间内尽可能多的处理请求,并保证服务质量.传统的高性能科学计算的特点是计算密集型,追求的目标是高速度,即"算得快";而新型高通量计算的特点是请求密集型,追求的目标是高通量,即单位时间内"算得多".传统商业化的高通量计算机应用大多面向文本信息的处理,而对图像视频等多媒体信息的处理大多仍采用离线处理、在线分发的方式进行.随着多媒体内容服务成为互联网最主要的服务之一,面向多媒体信息处理的高通量计算机应用越来越多,包括视频转码服务、视频会议服务、视频版权鉴定服务、内容检索服务等.但目前这些面向高并发的多媒体服务仍然运行在数据中心的通用硬件上,给数据中心带来了巨大的压力,降低了数据中心的效率,使得其实时性、服务质量和规模都受到了严重的影响.

其次,从图像视频计算技术的角度来说,现有的研究大多关注如何提高单个图像视频分析和处理的精度及速度,而较少关注高并发环境下的系统吞吐能力.所谓系统的吞吐量(throughput)指的是一套

计算机系统单位时间内可处理的服务请求数.近几年,为了提高图像视频分析算法的效率,研究者针对GPU(graphics processing unit)或众核等并行计算平台的硬件特性,挖掘算法本身的可并行性.但目前基于GPU加速的各种算法大多是把已有算法简单地在GPU平台上实现,需要在不同存储器之间多次拷贝数据,因此效率的提升非常有限.现有的图像视频计算技术与方法无论在速度还是精度上都无法满足前面所提到的各种高通量应用的需求.

因此,为了满足日益增长的高通量图像视频计算需求,本文对现有的高通量图像视频计算方法进行调研与分析,探讨现有高通量图像视频计算方法的不足,并结合新型硬件结构,利用其体系结构优势,提出高通量图像视频计算的未来研究方向以及需要解决的科学问题.

2 国内外研究概况

面向高通量的图像视频计算主要涉及计算数据及任务的相关性分析、并行理论、图像视频特征提取、聚类和学习算法、图像视频语义计算、视频编码的并行模式选择和并行去块滤波等研究内容.接下来,将分别从3个相关方面对现有方法进行详细介绍和分析.

2.1 高通量图像视频计算理论

高通量图像视频计算任务往往是基于一系列基本学习子任务的一个较大任务.而这些子任务之间往往是使用相同的数据或者关联数据的.如果能充分地利用子任务间的相关性和数据间的相关性,就能极大地提高计算的并行度和计算效率.

1) 相关性分析.在数据挖掘、统计学习和概率论等领域中已有很多研究成果和经典算法[1-2],比如典范相关分析(CCA)[3]等.这里的相关性指的是2个随机变量或2组数据在统计上的依赖关系.这些方法主要是为了挖掘和发现数据分布上的相关性.在多媒体领域中,有研究者提出挖掘语义相关性(semantic correlation)[4]以帮助提高检索的性能.语义相关性往往是基于多种媒体的共现频率定义的.在机器学习领域中,利用数据和任务的相关性提高多个学习任务的性能也已有很多研究成果.比如,在分类、回归和聚类等任务中使用迁移学习(transfer learning),利用在一个域上学习获得的模型或知识辅助另一个域上的学习任务[5].在迁移学习中,根据源域和目标域是否相同,以及源任务和目标任务是否相同,可以分为3类:归纳迁移学习(inductive transfer learning)、直推式迁移学习(transductive transfer learning)和无监督迁移学习(unsupervised transfer learning).①归纳迁移学习中,源任务和目标任务是不一样但相关的.针对该问题,研究者们提出了多任务学习框架(multi-task learning),如多任务特征学习[6-7]和正则化的多任务学习[8]等.②直推式迁移学习中的源域和目标域是不同但相关的,该领域主要采用域适应(domain adaptation)[9]等方法.③无监督迁移学习主要研究聚类、降维和密度估计等问题[10].

高通量图像视频计算的目的是实现图像视频的高通量计算,而高通量以提高计算效率为核心.此外,高通量图像视频计算任务往往是基于一系列基本学习子任务的一个较大任务,以及数据是限定于图像和视频这种高维媒体数据.从目前的研究现状可以看出,利用数据和任务的相关性提高基本学习任务的性能是主要目的.目前针对于提高图像视频计算效率的相关性分析研究并不多见.

2) 图像视频计算并行理论.由于通用处理器的性能有限,目前业界图像视频计算已经转向以GPU为主的硬件平台.GPU本质上是一种众核处理器[11],包含由大量处理简单任务的核心构成的阵列,本身作为图像处理的加速单元处理3D渲染、光源处理、立方体材质贴图等复杂的图像任务,从硬件本身提供强大的计算能力支持.

GPU提供了多个层面的并行性.以NVIDIA的Fermi为例,它拥有3层分级架构:4个图形处理图团簇(graphics processing clusters)、16个流阵列多处理器(streaming multiprocessors, SM)、512个硬件线程.每个团簇包括4个SM,每个SM包括32个硬件线程.在计算过程中,GPU计算最小的单位是线程,多个线程会被打包在一个warp内执行.由于GPU的并行性粒度限制得非常严格,软件如果无法拆分成32的整数倍个线程,就会出现硬件线程的浪费.

除GPU外,FPGA也被广泛应用到媒体计算中.FPGA一般由基于RAM的查找表(LUT)、DSP逻辑、SRAM块经可静态配置的二维多级网格链接而成.在FPGA上可以实现大量的运算器件.普通程序员缺乏硬件背景,很难直接组织这些器件,往往需要依赖OpenCL等相对高层次的语言来使用FPGA.OpenCL对硬件的抽象也是提供了3个层面:computer device级、compute unit级以及processing

element 级.这些级别和真正的软件应用之间也存在距离.

近年来,利用并行计算处理器进行图像视频计算方法优化,提升媒体计算效率逐渐成为新的研究热点.例如,在视频转码方面,Ko 等人根据转码所需要的缓存量来估计云转码系统所需要的机器数量,并设计了一个模拟器来计算合适的缓存数与机器数[12].Wu 等人根据每个用户的具体情况确定服务质量,使用虚拟机实现多用户的视频会议[13].Zhang 等人从降低功耗的角度给出了一种云端分配转码任务的算法,在队列延迟和处理功耗之间进行平衡[14].Jokhio 等人研究了云转码中离线转码存储转发和实时转码之间的能耗成本关系,并研究了平衡计算资源与存储资源成本的调度策略[15].然而,单纯在任务管理级别进行优化研究是不够的,要从根本上提高效率必须要结合具体的硬件.

由于受到实际硬件条件的限制,目前结合各种新型处理器进行图像视频处理优化的研究工作有限.新型的面向高通量计算机硬件的应用软件优化研究目前多集中在网页应用和数据挖掘等方面,对编解码应用的体系结构并行优化尚显不足.在编解码方面,Cho 等人利用 Cell 处理器特殊的 SPR 结构进行了 H.264 解码加速优化研究[16].Meenderinck 等人分析了从宏块到 GOP 级的所有级别的单路解码并行方法,并提出了 3D-Wave 的方法[17].以 Tilera 处理器为例进行解码并行化方面的研究工作也被提出了[18-19].在视频内容检索方面,高通量计算机并行优化的研究工作主要集中于移动视频检索、版本检测等方面.Diao 等人研究了在单个 GPU 上同时进行特征提取和检索的方法,并给出了在多 GPU 上进行扩展的模型[20].Fang 等人实现了一种并行视频内容检索算法[21],实验显示达到了 CUDA 实现的 SURF 算法性能的 46 倍.Liu 等人用 Map-Reduce 模型在 GPU 上实现了一套并行视频检索系统,与串行程序相比速度提升了 20 倍[22].由此可见,视频检索的硬件并行优化对提升系统整体性能作用显著.

综上,利用并行计算硬件资源进行图像视频处理算法优化已成为一个重要的研究方向,也取得了一定的成果.然而,目前的研究主要集中在如何利用现有并行计算硬件的体系机构特点提升图像视频计算效率.由于现有并行计算硬件的体系机构并不是针对媒体计算进行专门设计和优化,效率提升的空间有限.因此从根本上提高效率,必须有一套理论来刻画图像视频高通量计算的特点,在此理论指导下,从图像视频计算模型优化和并行计算硬件支撑 2 个方向共同努力,以实现图像视频的高通量计算.

2.2 高通量图像视频分析方法

1) 图像与视频的特征表示作为计算机视觉和模式识别领域的一个基本而重要的问题一直被广泛关注.高通量图像视频分析离不开高效的图像视频特征提取.这就需要开展高通量的图像视频特征提取手段.因此,下面对图像视频特征进行简要的分析.研究者已经提出了很多图像视频特征,这些特征大体可以分为人工设计的特征和基于数据学习的特征两大类.

人工设计的图像视频特征是针对图片分类、目标识别、视频检索、行为分析等应用,根据图像视频的颜色、纹理、亮度、边缘等属性,依靠专家的领域知识人工构造的特征描述方法.这些方法可以分为基于空间频域的特征和基于统计分布的特征.基于空间频域的特征主要利用频域变化方法提取局部上的空间频域特征.如 Gabor 特征采用 Gabor 小波变换实现频域特征的表示.Gabor 小波能从不同尺度和方向有效表示图像的局部特征,是一种被广泛应用的图像特征[23].相关工作[24-26]从不同角度对 Gabor 特征进行了扩展.基于统计分布的特征表示方法主要通过对像素的亮度或是梯度变化进行统计并计算相应的直方图特征.这种方法可以获得具有平移、旋转和尺度等不变性的特征.由 Lowe 提出的 SIFT (scale-invariant feature transform)是其中最具代表性的工作[27].SIFT 具备很好的平移、旋转、放缩等不变性,在图像匹配、目标识别和目标检测等方面得到了广泛应用.在 SIFT 特征的启发下,研究者提出了很多基于统计分布的特征表示方法,如 SURF (speeded up robust features)[28],HOG(histogram of oriented gradients)[29],LBP(local binary patterns)[30],BRIEF(binary robust independent elementary features)[31],FREAK(fast retina keypoint)[32]和 BoW(bag of words)[33-34].人工设计的特征在特定的应用问题上取得了不错的效果,但是这种特征依赖于专家的领域先验知识,而领域先验知识很多时候和真实场景中的复杂图像视频信号并不相符.因此,需要通过学习的方法从数据中学习非可控条件下的图像视频特征.

基于数据学习的特征通过学习方法从大量图像视频数据中挖掘数据内在的表示方式,最近,以卷积神经网络(CNNs)为代表的深度学习特征取得了很大成功.深度学习本质上是一种多层神经网络,通过

多层网络来从大量数据中学习不同层的抽象表示,它以比较自然的方式体现了从底层特征到高层特征的逐级抽象[35].深度学习最初应用在数据降维、手写数字识别等问题中,近年来在更广泛的领域中展现出了其有效性,例如在大规模图像分类、人脸识别、物体检测、动作识别等领域中.自从卷积神经网络(CNNs)在大规模图像分类任务上取得了突破后,研究者们对CNNs进行了不断的改进,新提出的CNNs的准确度得到不断提高.如在2012年ImageNet大尺度视觉识别竞赛(ILSVRC)中Krizhevsky等人提出的7层AlexNet[36]取得了最好的性能,其top5分类错误率是16.4%.在2014年ILSVRC竞赛中,谷歌公司提出的19层GoogLeNet的top5分类错误率是6.7%[37].然而,这些网络的计算成本(尤其是更准确的,但较大的模型)也在显著增加.

综上所述,目前已经有了一些图像视频特征被成功应用于各个领域中.但是随着海量图像视频时代的到来,对于高通量的图像视频处理需求,例如云服务器每天需要处理上亿的图片,目前这些特征的提取方法的计算量还是比较庞大,难以满足高通量的图像视频大数据处理的实时性要求,已成为处理高通量图像视频数据的主要瓶颈之一[38].研究人员亟需提出有效的手段来提高图像视频特征提取的效率.

2)抗噪性聚类是多媒体视觉特征提取和高维数据信息建模的有效手段.常见方法是以相似度邻接矩阵为基础的一类方法[39-40].通过在邻接图结构上查找密集子图的方式,所得到的聚类分析结果比传统的k均值聚类[41-42]和谱聚类[43-45]等方法具有更好的抗噪性能.密集子图搜索方法已被深入研究[46-48].Motzkin等人[49]证明了在无权重图上查找密集子图可以等价为一个在单纯形上的二次优化问题.这种思想进一步被扩展到处理有权重的图上,也被称为优势集方法(dominant set method).优势集方法通过复制动态(replicator dynamics, RD)方法[50]求解对应的标准二次优化问题.Rota Bulo等人[40]的研究表明,给定n个图节点和完全的图邻接矩阵,每一次RD迭代求解的时间复杂度是$O(n^2)$.这极大地阻碍了其被用于处理大规模数据.所以,Bulo等人提出了一种感染免疫动态模型(infection immunization dynamics,IID)来求解该标准二次优化问题,使得每一步优化的时间和空间复杂度降到了$O(n)$.然而,由于每次IID迭代需要维护一个完全的邻接矩阵,其总体优化过程的时间和空间复杂度仍然是$O(n^2)$.

由于大多数的密集子图都存在于一个邻接图的局部区域,所以在整个图结构上运行RD是不经济的[38,51].所以,Liu等人[39]提出一种基于搜索和扩展的方法(shrinking and expansion, SEA).这种方法将所有的RD循环限制在一个小的局部区域上进行,从而有效地避免不必要的时间和空间开销.在这种情况下,SEA的时间和空间复杂度与图边的数量是呈线性相关的.所以,SEA的可扩展性容易受到一个邻接图的稀疏程度的影响.邻接扩散(affinity propagation, AP)[52]是另外一种典型的具有抗噪性的方法,并且被广泛用于多媒体和视觉信息处理.它的另外一个优点在于其无需预先制定聚类的数量.这种方法通过在图边上进行信息传递的方式去搜索聚类模式.然而,当有巨大数量的节点和边时,这种方式十分耗时.均值漂移(mean shift, MS)[53]与基于邻接矩阵的方法有显著不同,区别在于其直接在特征空间进行聚类模式搜索.然而,均值漂移容易受到搜索带宽设定和特征维度等因素的影响.之前提到的基于邻接图的方法,在当邻接矩阵已经计算好的情况下能够获得非常高的检测质量.然而,这类方法由于计算邻接矩阵的需要,在大数据上的时间和空间复杂度都在$O(n^2)$级别.同时,一般的抗噪性聚类方法并不具备并行化的技术解决方案和系统实现.最重要的是,对于这类方法的高通量化研究以及在多媒体和视觉计算方面的系统实现,在国内外都是空白.

3)基于多任务多特征学习的视觉语义高通量计算模型.在多个特征表示上构建图像分析模型,一个简单的方案就是将多个特征拼接成一个长特征向量.另外一种方案是在单个特征上进行模型学习,最后融合多个统计模型的判别能力[54].在半监督学习领域,典型的有效利用多特征表示的半监督学习方法是联合训练[55]和多视角学习[56-57].

2004年,Lanckriet[58]和Bach[59]分别在不同的文章中提出和介绍了多核学习方法.然而,早期的多核学习方法的优化求解非常麻烦,因为其目标函数是一个具有二次约束的二次优化问题,必须用复杂的QCQP方法或者半定规划(SDP)方法加以解决,尽管Sonnenburg等人提出了可以用序列最小优化(SMO)求解[60],但复杂的附加条件仍然限制了多核学习方法的实用性.为了克服这一问题,Sonnenburg等人[60]在2006年提出一种基于cutting plane的大规模优化方法,并在工具包Shogun中实现,引发了相关学者的极大关注.进一步,Rakotomamonjy等人

在2008年提出了SimpleMKL[61],将多核学习问题用一个2步骤的选择优化机制(alternative optimization)去解决:步骤1,在给定核权值的情况下,优化一个等价的支持向量机二次优化问题;步骤2,在给定支持向量模型参数的情况下,自动更新核权值.此2步骤操作不断交替进行直到收敛.该方法尽管不能保证获得全局最优解,但仍保证了模型训练的低复杂度和模型的鲁棒性,从而使得多核学习方法逐渐流行起来,并被广泛使用于相关领域的研究.

多核学习的一个重要问题,是如何对核权重系数进行认识以及建模.早期研究的目的是对最佳核进行选择,故采用稀疏性约束(L1-norm).Bach等人提出对核权重采用复杂的结构稀疏性正则化约束[62].随后,Cortes等人提出了一种L2范式的多核学习[63],这个问题被形式化成一个一般性的Lp范式约束的多核学习模型[64-65],而该模型的求解也是用了与SimpleMKL类似的求解过程.Vishwanathan等人发现多核学习问题完全可以用SMO直接进行求解[66].

针对多核学习自身的特性,学者们从不同方面进行了研究.例如,Gonen等人[67]将多核学习的全局核权重扩展成局部核权重形式,这一思想被Yang等人借鉴并提出一种组敏感的多核学习模型[68].Suzuki等人进一步对多核学习的可扩展性进行了研究,并提出一种SpicyMKL[69]方法,利用近似梯度法(proximal gradient)对多核学习模型进行优化,通过并行化和加速可以处理具有上千个核的多核学习问题.Cortes等人利用Radermacher复杂度理论对多核学习的理论界进行了推导[70],这对多核学习的发展具有一定的指导意义.

在实际应用中常常需要处理包含成百上千个结构化语义类别的数据,这些语义类别可以组织成如WordNet[71]这种语义本体的结构形式.这种特性使得视觉特征在其空间中的分布非常杂乱.然而,如果对层次化语义信息进行分析,发现同一语义类别子集下面的图片往往有很多共同的视觉特性[72-73],而来自不同子集下面的图片则很容易被区分.这些先验知识可以促进语义概念之间的信息共享结构的构建,从而增强模型在实际应用中的图片分类能力.为了利用这些信息,近几年一些距离度量学习的方法被提出.例如Parameswaran等人[74]将最大边界近邻方法(large margin nearest neighbor method)[75]扩展成多任务距离度量学习;Hwang等人[76]提出学习一种距离度量树的方法来应对层次化的物体结构.

然而,已有研究主要集中在模型本身和优化求解2方面,对语义学习、度量学习和多任务学习等模型的高通量化研究在国内外尚属空白.

4) 图像视频的高通量多级计算模型.近年来,由于计算设备和机器学习技术的飞速发展,深度学习模型已逐渐成为视觉信息处理的一种基准方法.深度学习和人工智能早期的神经网络有着千丝万缕的联系.Hinton等人在2006年提出一种基于层叠式受限波尔兹曼机的深度模型[77],并提出了一种简单有效的模型优化方法,使深度学习(多层神经网络)避免了由于局部解造成的模型退化问题,受到了学术界的关注.该思想随后被广泛尝试,尤其是被用于数据表达学习方面[78-79].卷积神经网络(CNN)[80]也是一种深度学习模型,最早被用来处理特定类型的图像和语音信号.在融合了Hinton深度学习[77]的若干特点之后,CNN被首次尝试在一般性物体识别任务上[81],并在大规模基准视觉分类测试集上获得了比最好的非深度模型超过10%的性能提升.卷积神经网络通过对原始图像的多层多级滤波器卷积、池化和规整操作,将视觉信息进行解相关和重聚合,经过逐级映射形成具有语义显著性的高层特征,而在多级卷积层的后端和输出层之间的全连通层(或稀疏连通层)则起到了对卷积特征加以选择和融合的作用.这套多级学习机制有效地克服了从视觉信息的像素表示到语义输出之间的巨大语义鸿沟.卷积神经网络的成功应用极大地激发了学术界的关注,使得近年来深度学习方法被应用在计算机科学的各个方向,尤其是机器学习、模式识别、计算机视觉[81]、语音信号处理[82]、自然语言处理[83]等领域.在系统实现方面,深度学习的模型训练最初使用分布式Hadoop进行学习[81],之后被使用到高性能计算卡(GPU)或计算卡阵列上,训练速度获得了百倍甚至千倍的提升.然而,对深度和多级学习方法的高通量化研究尚待进一步探索.

2.3 高通量视频编码方法

由于视频编码过程极其复杂,各编码环节间和环节内存在广泛的数据依赖,因此高通量的视频编码研究包含对各编码环节内部的数据并行处理和编码环节间的并行化.

作为一个新兴的研究热点问题,基于GPU平台的HEVC编码拥有重要的学术价值和巨大的应用前景,吸引了来自于学术界和工业界不同领域的研究人员在这一问题上开展研究,著名的研究机构有Microsoft,Intel,MIT,CUHK,CityU,ICT,PKU

等.图1所示为HEVC编码器的结构图[84],其中,模式选择决定了运动估计等主要模块的效率,环路去块滤波在编码环节中占用了大量计算资源和带宽资源.由于HEVC标准2013年刚制定,目前针对HEVC的并行方法主要是面向以前的视频编码标准,大部分方法并行度不高,不能充分利用GPU这么多的计算单元.同时这些并行方法不能直接适用于HEVC标准,容易导致编码效率的损失.如何在保证编码效率的情况下,提高HEVC并行编码的并行度,实现高通量视频编码,已成为亟待解决的问题.

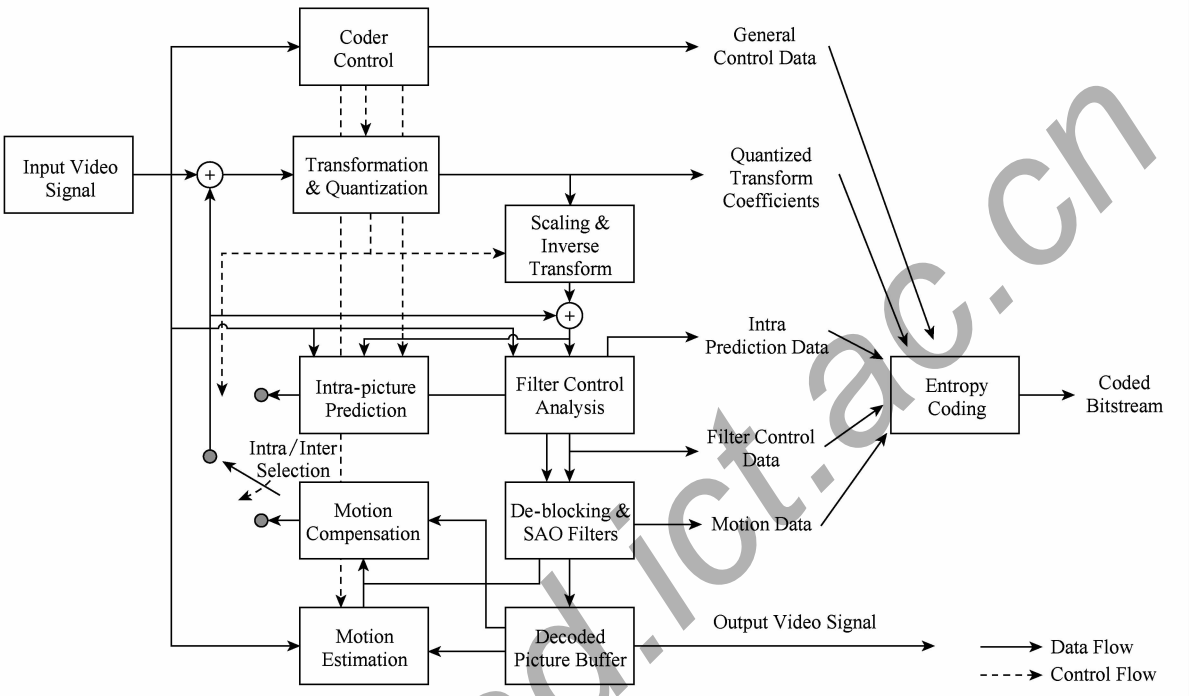


Fig. 1 The coder of HEVC

图1 HEVC编码器

1)并行模式选择.模式选择是对HEVC编码计算过程中的计算量和访存量影响最大的问题,贯穿于运动补偿等各主要模块,高效的并行模式选择方法是实现高通量视频编码的关键所在.

由于采用了灵活的编码结构,HEVC的编码模式搜索空间非常大.为了加快模式选择(mode decision, MD),HEVC标准本身采纳了多种便于并行编解码的特性,比如wavefront parallel processing (WPP)[85],Tiles[86]和MER[87].WPP对编码质量的影响很小,但是能提供的并行度不高.Chi等人基于WPP提出了一种增强算法[88],能在多帧之间并行,提高了并行度.使用Tiles将图像划分为很多独立编码的块能获得较高并行度,但是会对编码质量造成很大影响.WPP和Tiles都是图像区域之间的并行技术,MER则是定义了一个局部区域,在此区域内所有PU可以并行运动估计(motion estimation, ME).但是由于MER只适用于运动估计模块,模式选择的其他部分依然无法直接并行,整体并行度受到限制.

根据并行范围不同,目前的帧间MD方法可以分为全局并行方法[89-91]和局部并行方法[87,92].传统的全局并行方法在一帧图像内并行,全局并行方法从所有的候选匹配代价中选择匹配代价最小的块分割模式,所有4×4匹配代价可以并行处理得到,并行度高、实现简单,但是如果直接用于HEVC编码标准,会严重影响HEVC的编码效率.局部并行方法局限在一个块内部并行,适用于HEVC标准,保证了编码效率,但是并行度偏低,不能充分利用众核处理器平台的众多处理单元.

对于帧内MD并行,Zhao等人提出了一种使用前向无环图描述CTU之间依赖关系的方法[93],实现了CTU之间的并行处理,本质仍属于WPP,平均获得了5倍的加速比.Yan等人则使用分类器决策出最佳CTU的大小,通过较小的CTU大小能获得较高的加速比,平均达到10倍[94].Jiang等人提出了一种在TU四叉树划分时4个子节点并行帧内预测算法[95],但由于帧内预测仍依赖于重构像素,理论并行度只能达到4.

从以上内容可以看出,现有的优化算法主要是粗粒度并行(Tiles, WPP),Jiang 等人提出的算法[95]属于细粒度并行但是并行度不高,这都严重制约了编码系统的数据吞吐能力,难以实现高通量视频编码.

2) 并行去块滤波. 除了模式选择外,视频编码中的去块滤波环节同样存在多种数据依赖,是提高 HEVC 编码数据处理能力的另一个瓶颈. 目前的并行去块滤波方法,根据是否影响编码效率,可以分为无损并行方法[96-97,18]和有损并行方法[18].

无损并行方法优先保证编码块间相关性,因此保证了编码效率,但造成并行度过低. 在去块滤波过程中,直接按左、上和右上 3 个编码块的数据依赖性进行处理[96-97]. 最大并行度如式(1)所示,W 和 H 分别表示了帧图像水平和垂直方向的编码块数目,C 是处理单元数目.

$$MP=\begin{cases}\min\left(ceil\left(\dfrac{W}{H}\right),H\right), & \min\left(ceil\left(\dfrac{W}{2}\right),H\right)<C;\\ C, & \min\left(ceil\left(\dfrac{W}{2}\right),H\right)\geqslant C.\end{cases} \tag{1}$$

Yan 等人修改了滤波边界的顺序[18],将数据依赖减少到空间临近的左和上 2 个编码块,无损并行方法虽然保证编码效率不变,但是并行度太小,不适用于 GPU 平台,达不到高通量系统的处理要求. 这个工作还提出了一种有损并行方法,修改去块滤波的滤波顺序,减少滤波块间相关性. 为了解除子任务"滤波"内部的相关性,Yan 等人[18]修改了滤波的顺序,在同一图像帧内部,所有的垂直边界先滤波,再滤波所有的水平边界. 这种方法大幅度修改了传统去块滤波的滤波顺序,大大提高了并行度,但是对编码效率的影响也比较大.

从以上分析可以看出,目前已有的模式选择和并行去块滤波不能充分利用 GPU 平台的运算单元,不适用于 GPU 平台. 主要原因是它们存在如下 2 个问题:

1) 模式选择对于帧间和帧内模式选择方法,已有的粗粒度的并行方案如 Tiles 和 WPP 未能在并行度和编码质量之间取得较好的平衡,对编码质量影响较大或者并行度不高. 解除多层次的数据依赖性,提高细粒度的并行处理能力,对在 GPU 上构建高通量的 HEVC 编码系统有重要意义.

2) 去块滤波无损并行方法,编码效率不受影响,但是并行度低,无法充分利用 GPU 的计算单元;有损并行方法,修改了滤波顺序,牺牲了编码块间的相关性,进一步提高了并行度. 设计一种并行度高、编码效率好的并行去块滤波算法对于进一步提高 HEVC 编码性能具有重要意义.

2.4 现有方法的不足

通过对国内外研究现状的分析可以看出,现有成果无法满足当前海量图像视频计算应用和服务的重大需求,体现在 3 个方面:

1) 在图像视频计算理论方面. 传统数据和任务的相关性研究多是为了提高学习任务的性能,并不适用于图像视频这种高维数据,也无法应对实际应用中数据的高并发性带来的挑战;图像视频的并行计算研究主要集中在如何利用现有并行计算硬件的体系机构特点提升图像视频计算效率,由于现有并行计算硬件的体系机构并不是针对媒体计算进行的专门设计和优化,效率提升的空间有限.

2) 在图像视频分析算法方面. 目前已有一些针对特征提取经典算法的并行化算法,但是对于目前表现很好的卷积神经网络、多核学习等模型来说,计算复杂度较高,难以满足图像视频大数据处理的实时性要求;同时针对图像视频分析与语义理解的并行化算法较少.

3) 在视频编码方面. 目前针对 HEVC 的并行方法主要是面向以前的视频编码标准,大部分方法并行度不高,不能充分利用 GPU 的计算资源. 如何在保证编码效率的情况下,提高 HEVC 并行编码的并行度,已成为亟待解决的问题.

因此,目前迫切需要开展图像视频高通量计算理论与方法的研究工作,从理论、方法和实践 3 个层次入手,在高通量图像视频计算理论、高通量图像视频分析方法、高通量视频编码方法 3 个方面展开深入研究,以应对目前海量图像视频数据高并发、高维度、大流量等特性带来的挑战,满足实际多媒体应用高精度、高效率的需求.

3 未来研究方向

高通量图像视频计算是针对实际应用的挑战,处理当前图像视频数据的大容量和高并发问题. 为了有效地开展高通量图像视频计算的研究,需要从理论分析和实际方法 2 个方面展开工作. 具体地,未来的研究方向可能从任务内在关联性与计算结构并行性的多层次匹配,图像视频分析的高通量综合优化,以及视频编解码中高通量计算、码率、失真之间的度量与转换这 3 个方面研究高通量图像视频计算理论、高通量图像视频分析和高通量视频编解码问题.

3.1 高通量图像视频计算理论

图像视频计算任务往往存在数据冗余和子任务间的冗余.为了提高图像视频计算效率,需要分析这些冗余,提出相应的高通量图像视频计算方法.首先进行相关性分析,并根据相关性分析结果提出高通量计算理论,比如并行计算理论等.

1)图像视频计算相关性分析.在海量图像视频计算中,计算的数据之间以及计算任务之间往往存在一定的相关性.利用这些相关性,可以提高计算的效率或性能.所以,为了实现高通量图像视频计算,首先需要对图像视频数据和计算任务进行相关性分析.从相关性分析的对象来说,需研究数据之间的相关性和计算任务之间的相关性.从相关性分析来说,需研究如何发现数据及任务之间的相关性,以及如何利用所发现的相关性提高图像视频计算的效率和性能.

2)图像视频计算并行理论.核心是通过软硬件的多层次高效匹配,提高图像视频计算的并行效率,具体包括:①面向图像视频计算的并行理论模型,解决图像视频并行计算中软硬件并行粒度不同而存在的层次之间的误匹配问题的理论模型.②基于图像视频计算理论模型的算法优化,结合图像视频计算中数据和任务的高相关性、多层次性等特点,深入研究图像视频计算中的层次化并行任务分解问题.③面向图像视频计算理论模型的硬件支撑,包括并行度可重构的GPU架构(即在GPU的并行度层次进行一定程度的调整,以实现在不增加硬件计算资源的前提下提升硬件的实际效率)、图像视频计算多并行度编程语言(即一种有效支持图像视频计算中软硬件多并行度的编程语言).

3.2 高通量图像视频分析

图像视频分析涉及从底层特征到高层语义的多个方面.为了提高图像视频分析的效率,需要开展图像视频特征的高通量计算、聚类和学习算法的高通量计算以及高通量图像视频语义计算3个方面的研究.

1)图像视频特征的高通量计算.需对SIFT,HOG等常用局部图像特征的算法原理及构造过程进行深入的剖析,对算法的中间过程如图像尺度空间的建立、图像特征点的提取、特征点主方向的计算和特征点描述子的计算等进行详细的梳理,并提出合理的并行化图像视频特征提取方法,提炼图像视频特征并行化构造方法的一般规律.

2)聚类和学习算法的高通量计算.为了高效地处理海量图像视频数据,需通过并行化和管道策略设计多种常见的聚类和学习算法(如基于图结构的聚类算法、支持向量机、矩阵分解、隐含狄利克雷分布LDA)的高通量计算方法,在保持聚类和学习算法精度的情况下提高算法的运算效率.

3)高通量图像视频语义计算.从高通量的角度研究面向图像视频语义计算的理论和方法,包括:

① 基于多任务多特征学习的视觉语义高通量计算,如研究低复杂度的线性多特征度量计算模型以实现多特征相似度的并行化计算方法;利用数据的类属信息对模型进行学习;研究稀疏约束L1-norm和非稀疏约束Lp-norm的多特征相似度量学习的并行化模型训练方法;建立不同的度量学习任务在不同特征表示上的信息共享机制;构建可支持高通量计算的多层次多任务学习与信息共享机制,在分布式系统上研究并行多任务学习.

② 图像视频的高通量多级计算,如对视觉数据进行分块和分布式存储,使得在不同的运算处理单元上的数据子集的相关性尽可能小;设计和构建有效的卷积特征提取运算阵列,对海量视觉数据进行并行化特征提取;实现并行化判别子模型的快速学习和动态模型更新,有效提取不同的视觉语义子集的判别信息;构建合适的多层映射机制,对分布式判别子模型决策进行选择和同步融合.

3.3 高通量视频编解码

为提高视频编解码效率,需要提出高通量视频编码的计算-码率-失真理论模型(C-R-D模型)、基于C-R-D模型的预测模式高通量优化算法以及基于C-R-D模型的高通量视频编解码并行计算方法.

1)高通量视频编码的计算-码率-失真理论模型(C-R-D模型).基于高通量的编码相对于普通编码而言,增加了一个计算量维度来进行优化.理论上通过增加计算量,可以节省码率或者减少失真,但是目前还缺少一个精准的模型来刻画计算、码率和失真之间的关系.因此要重点研究计算量与码率之间的关系以及计算量与失真之间的关系,并结合已有的率失真理论构建计算-码率-失真理论模型.

2)基于C-R-D模型的预测模式高通量优化算法.HEVC视频编码相对于以前的视频编码,将编码基本单元扩大到64×64大小的块,每个块从具体的划分到预测模式有大量参数需要决定,现有编码方案都只能采用局部串行优化算法,无法得到最优的编码性能.需研究如何设计并行算法,计算每个编码基本单元在各种划分下的运动矢量及预测误差;研究如何设计并行算法计算在各种划分下帧内预测的预测误差;并在所提出的C-R-D模型指导下,优化每个基本编码单元的块划分以及每个块的预测参数.

3) 基于 C-R-D 模型的高通量视频编解码并行计算方法. 除了基本编码单元划分和预测参数外, HEVC 编码还包括变换、熵编码、环路滤波等,这些处理在编码器和解码器中都需要,提高它们的并行计算能力对提供编解码速度有着重要的意义. 需要研究不同大小块的正变换和反变换的并行算法;研究在 C-R-D 模型指导下熵编码和解码的并行算法,尽可能在减少编码性能损失的条件下提高熵编码和解码的并行性;研究环路滤波的并行算法.

4 关键科学问题

针对第 3 节中分析出的未来研究方向,需要解决 3 个关键科学问题:

1) 如何解决图像视频高通量计算中任务内在关联性与计算结构并行性的多层次匹配问题?

在图像视频计算中存在多个层次的关联性,例如宏块间关联、条带间关联、帧间关联、任务间关联. 同时高通量计算平台本身又具备多个层次的并行性,例如 OpenCL 提供的 computer device 级并行、compute unit 级并行以及 processing element 级并行. 为了提升图像视频高通量计算的效率,必须将任务内在关联性和计算结构并行性从各个层次上进行合理的匹配.

2) 如何解决图像视频分析中多层次贯通式的高通量综合优化问题?

图像视频数据是大数据中"体量最大的大数据",如何突破图像视频高通量分析中的优化技术已经成为信息科学技术的重大挑战. 需从底层、中层、高层 3 个层次上对高通量图像视频分析进行探索,建立从特征到语义各个层次的高通量计算模型,进而实现贯通式的综合优化.

3) 如何建立视频编码中高通量计算、码率、失真之间的度量与转换模型?

高通量视频编码的核心问题是如何精确刻画计算、码率和失真的关系,需研究计算与码率之间的关系和计算与失真之间的关系,并引入传统的编码理论中码率和失真的指数关系,通过深入理论分析和大量的实验验证,建立计算、码率和失真的理论模型.

5 总 结

本文针对高通量图像视频计算问题,首先详细分析了现有的高通量图像视频计算相关方法与技术,并进一步讨论了现有高通量图像视频计算方法研究的不足,最后分析了高通量图像视频计算的未来研究方向及需解决的科学问题.

参 考 文 献

[1] Han J, Kamber M, Pei J. Data Mining, Southeast Asia Edition: Concepts and Techniques [M]. San Francisco, CA: Morgan Kaufmann, 2006

[2] Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning [M]. Berlin: Springer, 2009

[3] Hardoon D, Szedmak S, Shawe-Taylor J. Canonical correlation analysis: An overview with application to learning methods [J]. Neural Computation, 2004, 16(12): 2639-2664

[4] Zhang Hong, Wu Fei, Zhuang Yueting. Cross media correlation reasoning and retrieval [J]. Journal of Computer Research and Development, 2008, 45(5): 869-876 (in Chinese)
(张洪, 吴飞, 庄越挺. 跨媒体相关性推理与检索研究[J]. 计算机研究与发展, 2008, 45(5): 869-876)

[5] Pan Jialin, Yang Qiang. A survey on transfer learning [J]. IEEE Trans on Knowledge and Data Engineering, 2010, 22(10): 1345-1359

[6] Argyriou A, Evgeniou T, Pontil M. Multi-task feature learning [C] //Proc of Advances in Neural Information Processing Systems. Cambridge, MA: MIT Press, 2006: 41-48

[7] Argyriou A, Evgeniou T, Pontil M. Convex multi-task feature learning [J]. Machine Learning, 2008, 73(3): 243-272

[8] Evgeniou T, Pontil M. Regularized multi—task learning [C] //Proc of the 10th ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining. New York: ACM, 2004: 109-117

[9] Blitzer J, McDonald R, Pereira F. Domain adaptation with structural correspondence learning [C] //Proc of Empirical Methods in Natural Language Processing. Stroudsburg, PA: ACL, 2006: 120-128

[10] Wang Zheng, Song Yangqiu, Zhang Changshui. Transferred dimensionality reduction [C] //Proc of European Conf on Machine Learning and Knowledge Discovery in Databases. Berlin: Springer, 2008: 550-565

[11] NVIDIA. NVIDIA Launches the World's First Graphics Processing Unit: GeForce 256 [EB/OL]. (2002-01-11) [2016-12-20]. http://www.nvidia.com/object/IO_20020111_5424.html

[12] Ko S, Park S, Han H. Design analysis for real-time video transcoding on cloud systems [C] //Proc of ACM Symp on Applied Computing. New York: ACM, 2013: 1610-1615

[13] Wu Yu, Wu Chuan, Li Bo, et al. vSkyConf: Cloud-assisted multi-party mobile video conferencing [C] //Proc of the 2nd ACM SIGCOMM Workshop on Mobile Cloud Computing. New York: ACM, 2013: 33-38

[14] Zhang Weiwen, Wen Yonggang, Cai Jianfei, et al. Towards transcoding as a service in multimedia cloud: Energy-efficient job-dispatching algorithm [J]. IEEE Trans on Vehicular Technology, 2014, 63(5): 2002-2012

[15] Jokhio F, Ashraf A, Lafond S, et al. A computation and storage trade-off strategy for cost-efficient video transcoding in the cloud [C] //Proc of Euromicro Conf on Software Engineering and Advanced Applications. Piscataway, NJ: IEEE, 2013: 365-372

[16] Cho Y, Kim S, Lee J, et al. Parallelizing the H. 264 decoder on the cell BE architecture [C] //Proc of ACM Int Conf on Embedded Software. New York: ACM, 2010: 49-58

[17] Meenderinck C, Azevedo A, Juurlink B, et al. Parallel scalability of video decoders [J]. Journal of Signal Processing Systems, 2009, 57(2): 173-194

[18] Yan Chenggang, Dai Feng, Zhang Yongdong. Parallel deblocking filter for H. 264/AVC on the TILERA many-core systems [C] //Proc of Int Conf on Multimedia Modeling. Berlin: Springer, 2011: 51-61

[19] Chi C, Alvarez-Mesa M, Lucas J, et al. Parallel HEVC decoding on multi-and many-core architectures [J]. Journal of Signal Processing Systems, 2013, 71(3): 247-260

[20] Diao M, Nicopoulos C, Kim J. Large-scale semantic concept detection on manycore platforms for multimedia mining [C] // Proc of IEEE Int Parallel & Distributed Processing Symp. Piscataway, NJ: IEEE, 2011: 384-394

[21] Fang Zhenman, Yang Donglei, Zhang Weihua, et al. A comprehensive analysis and parallelization of an image retrieval algorithm [C] //Proc of IEEE Int Symp on Performance Analysis of Systems and Software. Piscataway, NJ: IEEE, 2011: 154-164

[22] Liu Keyan, Zhang Tong, Wang Lei. A new parallel video understanding and retrieval system [C] //Proc of the 2010 IEEE Int Conf on Multimedia and Expo. Piscataway, NJ: IEEE, 2010: 679-684

[23] Daugman J G. High confidence visual recognition of persons by a test of statistical independence [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 1993, 15(11): 1148-1161

[24] Zhang Baochang, Shan Shiguang, Chen Xilin, et al. Histogram of gabor phase patterns (HGPP): A novel object representation approach for face recognition [J]. IEEE Trans on Image Processing, 2007, 16(1): 57-68

[25] Meyers E, Wolf L. Using biologically inspired features for face processing [J]. International Journal on Computer Vision, 2008, 76(1): 93-104

[26] Lei Lin, Wang Zhuang, Su Yi. A new invariant feature detector based on multi-scale gabor filter bank [J]. Acta Electronic Sinica, 2009, 37(10): 2134-2139 (in Chinese)
(雷琳, 王壮, 粟毅. 基于多尺度Gabor滤波器组的不变特征点提取新方法[J]. 电子学报, 2009, 37(10): 2134-2139)

[27] Lowe D. Distinctive image features from scale-invariant keypoints [J]. International Journal of Computer Vision, 2004, 60(2): 91-110

[28] Bay H, Ess A, Tuytelaars T, et al. Speeded-up robust Features (SURF) [J]. Computer Vision and Image Understanding, 2008, 110(3): 346-359

[29] Dalal N, Triggs B. Histograms of oriented gradients for human detection [C] //Proc of IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2005: 886-893

[30] Ojala T, Pietikäinen M, Harwood D. A comparative study of texture measures with classification based on feature distributions [J]. Pattern Recognition, 1996, 29(1): 51-59

[31] Calonder M, Lepetit V, Strecha C, et al. Brief: Binary robust independent elementary features [C] //Proc of the European Conf on Computer Vision. Berlin: Springer, 2010: 778-792

[32] Alahi A, Ortiz R, Vandergheynst P. FREAK: Fast retina keypoint [C] //Proc of IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2012: 510-517

[33] Li Feifei, Perona P. A Bayesian hierarchical model for learning natural scene categories [C] //Proc of IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2005: 524-531

[34] Sivic J, Russell B C, Efros A, et al. Discovering objects and their localization in images [C] //Proc of IEEE Int Conf Computer Vision. Piscataway, NJ: IEEE, 2005: 370-377

[35] Hinton G, Salakhutdinov R. Reducing the dimensionality of data with neural networks [J]. Science, 2006, 313(5786): 504-507

[36] Krizhevsky A, Sutskever I, Hinton G. Imagenet classification with deep convolutional neural networks [C] // Proc of Neural Information Processing Systems. Lake Tahoe, Nevada: Curran Associates Inc, 2012: 1106-1114

[37] Szegedy C, Liu Wei, Jia Yangqing, et al. Going deeper with convolutions [EB/OL]. [2016-12-15]. https://arxiv.org/abs/1409.4842

[38] Tang Sheng, Gao Ke, Gu Xiaoguang, et al. High-throughput video content analysis technologies [J]. Journal of Engineering Studies, 2014, 6(3): 294-306 (in Chinese)
(唐胜, 高科, 顾晓光, 等. 高通量视频内容分析技术[J]. 工程研究——跨学科视野中的工程, 2014, 6(3): 294-306)

[39] Liu Hairong, Latecki L, Yan Shuicheng. Fast detection of dense subgraphs with iterative shrinking and expansion [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2013, 35(9): 2131-2142

[40] Rota Bulo S, Pelillo M, Bomze I. Graph-based quadratic optimization: A fast evolutionary approach [J]. Computer Vision and Image Understanding, 2011, 115(7): 984-995

[41] Bahmani B, Moseley B, Vattani A, et al. Scalable k-means++ [J]. Proceedings of the VLDB Endowment, 2012, 5(7): 22-633

[42] Lloyd S. Least squares quantization in PCM [J]. IEEE Trans on Information Theory, 1982, 28(2): 129-137

[43] Fowlkes C, Belongie S, Chung F, et al. Spectral grouping using the nystrom method [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2004, 26(2): 214-225

[44] Zhou Lin, Ping Xijian, Xu Sen, et al. Cluster ensemble based on spectral clustering [J]. Acta Automatica Sinica, 2012, 38(8): 1335-1342 (in Chinese)
(周林, 平西建, 徐森, 等. 基于谱聚类的聚类集成算法[J]. 自动化学报, 2012, 38(8): 1335-1342)

[45] Wauthier F, Jojic N, Jordan M. Active spectral clustering via iterative uncertainty reduction [C] //Proc of the 18th ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining. New York: ACM, 2012: 1339-1347

[46] Anderson P, Thor A, Benik J, et al. Pang: Finding patterns in annotation graphs [C] //Proc of the 2012 ACM SIGMOD Int Conf on Management of Data. New York: ACM, 2012: 677-680

[47] Angel A, Sarkas N, Koudas N, et al. Dense subgraph maintenance under streaming edge weight updates for real-time story identification [J]. Proceedings of the VLDB Endowment, 2012, 5(6): 574-585

[48] Wang N, Parthasarathy S, Tan K, et al. Csv: Visualizing and mining cohesive subgraphs [C] //Proc of the 2008 ACM SIGMOD Int Conf on Management of Data. New York: ACM, 2008: 445-458

[49] Motzkin T S, Straus E G. Maxima for graphs and a new proof of a theorem of turan [J]. Canadian Journal of Mathematics, 1965, 17(4): 533-540

[50] Weibull J W. Evolutionary Game theory [M]. Cambridge, MA: MIT Press, 1997

[51] Liu Hairong, Yan Shuicheng. Robust graph mode seeking by graph shift [C] //Proc of the 27th Int Conf on Machine Learning. Madison, WI: Omnipress, 2010: 671-678

[52] Frey B J, Dueck D. Clustering by passing messages between data points [J]. Science, 2007, 315(5814): 972-976

[53] Comaniciu D, Meer P. Mean shift: A robust approach toward feature space analysis [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2002, 24(5): 603-619

[54] Snoek C, Worring M. Early versus late fusion in semantic video analysis [C] //Prof of ACM Int Conf on Multimedia. New York: ACM, 2005: 399-402

[55] Blum A, Mitchell T. Combining labeled and unlabeled data with co-training [C] //Proc of Int Conf on Computational Learning Theory. New York: ACM, 1998: 92-100

[56] Sindhwani V, Niyogi P, Belkin M. Beyond the point cloud: From transductive to semi-supervised learning [C] //Prof of the 22nd Int Conf on Machine Learning. Madison, WI: Omnipress, 2005: 824-831

[57] Rosenberg D, Sindhwani V, Bartlett P, et al. Multiview point cloud kernels for semisupervised learning [J]. IEEE Signal Processing Magazine, 2009, 26(5): 145-150

[58] Lanckriet G, Cristianini N, Bartlett P, et al. Learning the kernel matrix with semi-definite programming [J]. Journal of Machine Learning Research, 2004, 5: 27-72

[59] Bach F, Lanckriet, Jordan M. Multiple kernel learning, conic duality, and the SMO algorithm [C] //Proc of the 21st Int Conf on Machine Learning. Madison, WI: Omnipress, 2004: 1-8

[60] Sonnenburg S, Ratsch G, Schafer C, et al. Large scale multiple kernel learning [J]. Journal of Machine Learning Research, 2006, 7: 1531-1565

[61] Rakotomamonjy A, Bach F, Canu S, et al. SimpleMKL [J]. Journal of Machine Learning Research, 2008, 9: 2491-2521

[62] Bach F. Consistency of the group Lasso and multiple kernel learning [J]. Journal of Machine Learning Research, 2008, 9: 1179-1225

[63] Cortes C, Mohri M, Rostamizadeh A. L2 regularization for learning kernels [C] //Proc of the 25th Conf on Uncertainty in Artificial Intelligence. Montreal, Quebec, Canada: AUAI, 2009: 109-116

[64] Varma M, Ray D. Learning the discriminative power-invariance trade-off [C] //Proc of the 11th Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2007: 1-8

[65] Kloft M, Brefeld U, Sonnenburg S, et al. Efficient and accurate Lp-norm multiple kernel learning [C] //Proc of Neural Information Processing System. Lake Tahoe, Nevada: Curran Associates Inc, 2009: 997-1005

[66] Vishwanathan S, Sun Zhaonan, Theera-Ampornpunt N, et al. Multiple kernel learning and the SMO algorithm [C] //Proc of the 24th Neural Information Processing System. Lake Tahoe, Nevada: Curran Associates Inc, 2010: 2361-2369

[67] Gonen M, Alpaydin E. Localized multiple kernel learning [C] //Proc of the 25th Int Conf on Machine Learning. Madison, WI: Omnipres, 2008: 352-359

[68] Yang Jingjing, Li Yuanning, Tian Yunhong, et al. Group sensitive multiple kernel learning for object categorization [C] //Proc of the 12th Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2009: 436-443

[69] Suzuki T, Tomioka R. SpicyMKL: A fast algorithm for multiple kernel learning with thousands of kernels [J]. Machine Learning, 2011, 85: 77-108

[70] Cortes C, Mohri M, Rostamizadeh A. Generalization bounds for learning kernels [C] //Proc of the 27th Int Conf on Machine Learning. Madison, WI: Omnipress, 2010: 247-254

[71] Miller G. WordNet: A lexical database for English [J]. Communications of ACM, 1995, 38(11): 39-41

[72] Torralba A, Murphy K, Freeman W. Sharing visual features for multi-class and multi-view object detection [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2007, 29(5): 854-869

[73] Hwang S, Grauman K, Sha F. Learning a tree of metrics with disjoint visual feature [C] //Proc of Neural Information Processing Systems. Lake Tahoe, Nevada: Curran Associates Inc, 2011: 621-629

[74] Parameswaran S, Weinberger K. Large margin multi-task metric learning [C] //Proc of Neural Information Processing Systems. Lake Tahoe, Nevada: Curran Associates Inc, 2010: 1867-1875

[75] Weinberger K, Saul L. Distance metric learning for large margin nearest neighbor classification [J]. Journal of Machine Learning Research, 2009, 10: 207-244

[76] Hwang S, Sha F, Grauman K. Sharing features between objects and their attributes [C] //Proc of IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2011: 1761-1768

[77] Hinton G, Salakhutdinov R. Reducing the dimensionality of data with neural networks [J]. Science, 2006, 313(5786): 504-507

[78] Hinton G. Learning multiple layers of representation [J]. Trends in Cognitive Sciences, 2007, 11(10): 428-434

[79] Bengio Y, Courville, Vincent P. Representation learning: A review and new perspectives [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2013, 35(8): 1798-1828

[80] Le Cun, Bengio Y. Convolutional networks for images, speech, and time-series, in Arbib, M. A. (Eds)[G] //The Handbook of Brain Theory and Neural Networks. Cambridge, MA: MIT Press, 1995

[81] Krizhevsky A, Sutskever I, Hinton G. Imagenet classification with deep convolutional neural networks [C] //Proc of Neural Information Processing Systems. Lake Tahoe, Nevada: Curran Associates Inc, 2012: 1106-1114

[82] Yu Dong, Deng Li, Seide F. The deep tensor neural network with applications to large vocabulary speech recognition [J]. IEEE Trans on Audio, Speech, and Language Processing, 2013, 21(2): 388-396

[83] Bengio Y, Senécal J. Adaptive importance sampling to accelerate training of a neural probabilistic language model [J]. IEEE Trans on Neural Networks, 2008, 19(4): 713-722

[84] Sullivan G, Ohm J, Han W, et al. Overview of the high efficiency video coding (HEVC) standard [J]. IEEE Trans on Circuits and Systems for Video Technology, 2012, 22(12): 1649-1668

[85] Clare G, Henry F, Pateux S. Wavefront parallel processing for HEVC encoding and decoding, JCTVC-F274 [R]. San Jose, CA: Joint Collaborative Team on Video Coding (JCT-VC), 2011

[86] Fuldseth A, Horowitz M, Xu S, et al. Tiles for managing computational complexity of video encoding and decoding [C] //Proc of Picture Coding Symp. Piscataway, NJ: IEEE, 2012: 389-392

[87] Zhou M. AHG10: Configurable and CU-group level parallel merge/skip, JCTVC-H0082 [R]. San Jose, CA: Joint Collaborative Team on Video Coding (JCT-VC), 2012

[88] Chi C, Alvarez M, Juurlink B, et al. Parallel scalability and efficiency of HEVC parallelization approaches [J]. IEEE Trans on Circuits and Systems for Video Technology, 2012, 22(12): 1827-1838

[89] Leupers R, Eeckhout L, Martin G, et al. Virtual manycore platforms: Moving towards 100+ processor cores [C] //Proc of Design, Automation & Test in Europe Conf & Exhibition (DATE). Piscataway, NJ: IEEE, 2011: 715-720

[90] Bini E, Buttazzo G, Eker J, et al. Resource management on multicore systems: The ACTORS approach [J]. IEEE Micro, 2011, 31(3): 72-81

[91] Annavaram M. A case for guarded power gating for multi-core processors [C] //Proc of the 17th IEEE Int Symp on High Performance Computer Architecture (HPCA). Piscataway, NJ: IEEE, 2011: 291-300

[92] Yu Qin, Zhao Liang, Ma Siwei. Parallel AMVP candidate list construction for HEVC [C] //Proc of Visual Communications and Image Processing (VCIP). Piscataway, NJ: IEEE, 2012: 1-6

[93] Zhao Yanan, Song Li, Wang Xiangwen, et al. Efficient realization of parallel HEVC intra encoding [C] //Proc of the 2013 IEEE Int Conf on Multimedia and Expo Workshops. Piscataway, NJ: IEEE, 2013: 1-6

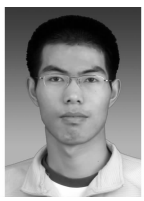
[94] Yan Chenggang, Zhang Yongdong, Dai Feng, et al. Efficient parallel HEVC intra-prediction on many-core processor [J]. Electronics Letters, 2014, 50(11): 805-806

[95] Jiang Jie, Guo Longbao, Mo Wei, et al. Block-based parallel intra prediction scheme for HEVC [J]. Journal of Multimedia, 2012, 7(4): 289-294

[96] Chi C, Juurlink B, Meenderinck C. Evaluation of parallel H. 264 decoding strategies for the cell broadband engine [C] //Proc of the 24th ACM Int Conf on Supercomputing. New York: ACM, 2010: 105-114

[97] Lee J Y, Lee J J, Park S. Multi-core platform for an efficient H. 264 and VC-1 video decoding based on macroblock row-level parallelism [J]. IET Circuits, Devices & Systems, 2010, 4(2): 147-158

Tang Jinhui, born in 1981. PhD, professor, PhD supervisor. Senior member of IEEE. His main research interests include large-scale multimedia search, social media mining, and computer vision.



Li Zechao, born in 1985. PhD, associate professor. His main research interests include large-scale multimedia understanding, social media mining, etc.

Liu Shaoli, born in 1987. PhD, associate professor. His main research interests include computer architecture, machine learning, parallel computing and video processing.

Qin Lei, born in 1977. PhD, associate professor. His main research interests include image/video processing, computer vision, and pattern recognition.