

智慧城市视频大数据的 趋势和挑战

田永鸿：北京大学信息技术学院，教授，北京，100871

黄铁军：北京大学信息技术学院，教授，北京，100871

高文：北京大学信息技术学院，院士，教授，北京，100871

摘要：智慧城市和大数据已经上升为我国的国家战略。为进一步促进我国城市化发展，迫切需要推进城市大数据特别是视频大数据的研究，充分挖掘我国城市数据资源的巨大价值。分析当前城市视频大数据技术发展总体趋势，并重点阐述城市视频大数据的关键技术挑战。

关键词：智慧城市；视频大数据；编码；识别；计算平台

中图分类号：TP391

文献标识码：A

文章编号：1001-683X(2016)01-0082-05

0 引言

自2008年IBM提出“Smart City”的概念以来，“智慧城市”这一新型城市发展理念在全球快速兴起，并已经成为全球城市发展的热点。目前，我国的城市化过程正处于历史关口，推动智慧城市建设将成为我国经济社会发展的重要抓手。截至2012年底，我国已有超过230个城市在“十二五”规划或政府工作报告中提出建设智慧城市，计划投资总额将超过2万亿元。与欧美的智慧城市建设主要关注“管理与服务的智能化”问题不同，我国当下的时代特征是“四化”（即工业化、信息化、城镇化、农业现代化）同步发展，城市化发展在“物理世界”、“人类社会”和“信息世界”所组成的三元空间中都面临着巨大变化，因此我国智慧城市发展的内涵与实践要丰富得多，面临着巨大的发展机遇和技术挑战。

一般认为，智慧城市（或称为智能城市，即 Intelligent City），是把新一代信息技术充分运用在城市的各行各业之中，形成基于知识创新的城市信息化高级形态，实现信息化、工业化与城镇化深度融合，从而有助于缓解“大城市病”，提高城镇化质量，实现精细化和动态管理^[1]。这里的新一代信息技术是指包括新一代互联网、云计算、智能传感、通信、遥感、卫星定位、地理信息系统等技术及其融合，依托它们可实现城市中一切实体和行为的智能化感知、识别、定位、跟踪和监管，并在此基础上通过机器学习和大数据分析处理等技术来实现海量信息的处理和决策支持，从而使城市达到“智慧”的状态。

在面向智慧城市的新一代信息技术中，大数据具有非常重要的意义和作用。它能在城市的各个方面得到广泛应用，是连接城市信息基础设施与城市智能化应用系统之间的桥梁。同时，智慧城市面向国家发展的重大战略需求，是对大数据的整合性、有效性和价值性的重要检验。在此主要从面向智慧城市的大数据（特别是视频大数据）研究发展趋势与技术挑战等方面进行探讨，以期达到抛砖引玉的目的。

1 大数据——“智慧城市”发展的助推剂

近年来,伴随着城市化、信息化进程不断发展,城市信息基础设施得到了飞速发展。然而,要实现智慧城市,关键还需要在城市信息基础设施的基础上实现“智慧政府”、“智慧安保”、“智慧交通”、“智慧医疗”等各种智能化应用。在信息基础设施获取的数据和各种智能应用之间,大数据扮演了连接两者桥梁的重要角色。目前,大数据已经渗透到了医疗卫生、科研、教育、公共安全、交通、物流、环境、气候、商务、金融、能源、制造等城市发展的各方面,城市大数据中蕴含着巨大的价值,能促进产业创新,催生新型的网络经济和数据经济。美国信息技术与创新基金会在2013年11月的报告中甚至宣称,如果说石油是20世纪经济发展的助推剂,那么数据将成为21世纪经济发展的助推剂。

智慧城市面向国家发展的重大战略需求,是对大数据的整合性、有效性和价值性的重要检验,也是检验大数据方法与技术的重要试金石。目前,我国城市的大数据远远没有得到充分地聚集和利用。如何聚集、统合各类城市大数据,并对数据进行高效地分析和挖掘,充分提取城市大数据的价值并服务于我国智慧城市建设,是我国智慧城市发展亟待解决的一个关键的核心问题。

在各类城市大数据中,图像视频是“体量最大的大数据”。据思科统计,视频内容约占互联网总流量的90%,而在迅速发展的移动网络中,视频流量的比例也高达64%,并以超过130%的年复合增长率增长。可见,图像视频数据在大数据中占据着主导地位,因此图像视频的处理是大数据应用的关键所在。而且,相对于文本、语音等数据,图像视频的数据量更大、维度更高,其表达、处理、传输和利用的技术挑战性更大。因此,以图像视频为对象的大数据研究将成为新的发展方向和研究热点。

2 智慧城市视频大数据技术发展总体趋势

近年来,我国通过平安城市建设实现了城域范围的视频监控系统。据统计,2013年全国监控摄像头的规模已经超过3 000万台。以监控视频为例,从应用、数据、技术3个维度分析当

前城市视频大数据的总体发展趋势(见图1)。

(1)从应用维度来看,城市视频感知技术与系统在安防、交管上已有初步成功应用,但在城市运行(如城市范围内的运行态势分析)和社会服务(如重点景区的实时直播、家庭看护)方面才刚刚起步。因此,如何从“管、控”到“服务”是城市视频大数据应用发展的重点。

(2)从数据维度来看,基于单摄像头的城市视频感知分析技术与系统已初步可应用。一个典型的例子就是交通电子警察的监控与违章发现。然而,随着城市视频监控规模不断扩大和应用需求的爆炸式增长,处理跨摄像头视频数据、大规模摄像头网络数据,甚至融合各类视频图像及关联数据的城市视频大数据就成为当务之急。

(3)从技术维度来看,过去20年已基本解决监控视频摄像系统的数字化问题,近5年开始解决大规模监控摄像头的高清化问题。然而在现阶段,大量布设的监控摄像节点的智能程度低,信息不能实时处理,从而造成以人工监测为主的应用模式效率低下,对突发事件的反应缓慢。因此,视频监控的智能化是今后相当长一段时间内的研究与应用重点。更进一步,由于目前仍缺少广域范围内多摄像机视频数据的协同处理与计算技术,因而难以深度利用和挖掘这些广域视频中丰富的人、物、行为乃至事件信息。因此大数据智能化(简称大数据化)是未来若干年的必然研究与发展趋势。

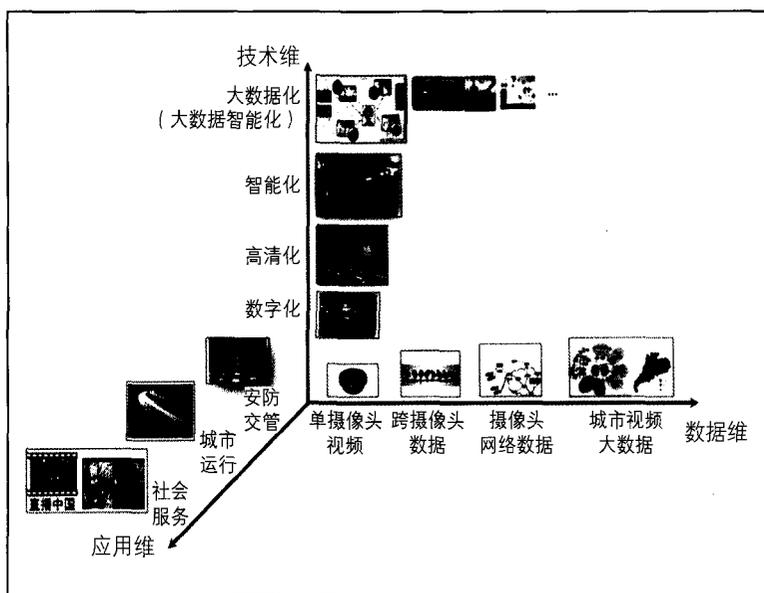


图1 城市视频大数据技术发展总体趋势

城市视频大数据的主体是监控视频，但广义的城市视频大数据还包括会议视频、家庭视频、教学视频、法庭视频等，它们往往采用固定摄像头对某个特定的场景（如家庭、教室、会议室、法庭、交通路口等）在一段时间内进行拍摄。由于这类视频都有较为固定的场景特性，场景的拍摄并没有预定义的剧本或意图，称之为“场景视频”（Scenic Video）。与其他类型视频（如新闻视频、影视视频、体育视频）相比，场景视频是机器视觉研究的最可能突破口，也是解决视频大数据问题的最佳实验数据。首先，场景视频是通过摄像头长期注视单一场景而采集的视频，场景中的“变”与“不变”要素能有效地进行建模、分析与推理，从而有可能达到高效压缩、场景理解甚至精确识别；其次，场景视频的体量巨大，是城市视频大数据中的主要研究对象，因此场景视频处理与分析技术的发展能直接推动城市视频大数据的技术水平。

从更大的视角来看，信息技术领域正在孕育重大技术突破与变革，视频大数据是主要的推动力之一及最佳的应用问题。一方面，认知与脑科学、机器视觉等相关领域正在发生从“量变”到“质变”的过程，预期能在未来若干年内获得理论上的突破。另一方面，近年来以深度学习和电子大脑为代表的人工智能技术获得了迅猛的发展，产生了包括Google Brain、百度大脑、IBM的类人脑芯片TrueNorth等一批标志性的阶段成果。这些成果将为场景视频的处理与分析提供强有力的计算与处理能力。

3 智慧城市视频大数据的关键技术挑战

要充分利用城市视频大数据来助力智慧城市发展，还需要依靠高效的视频大数据分析处理方法作为支撑，更有效地挖掘出城市视频大数据的巨大价值。城市视频大数据是内容最丰富、最活跃的一类大数据，集中体现了大数据体量大、变化快、多样性和真实性的特点，并具有非结构化、实时性、精细化、交叉性强等

新特点。对城市视频大数据的实时处理和创新利用，特别是全市意义上的大数据实时综合利用，需要重点解决“存不下”、“找不到”和“看不清”3个关键技术挑战，从而突破大数据高效处理和深度分析的架构与技术。

3.1 存不下 (compression issue)

所谓存不下，是指城市视频大数据的数据量增长速度远远超过视频压缩编码技术的发展速度，从而造成所需的存储开销呈指数级增长。根据美国国际数据公司（IDC）的估计，数据每2年翻一番，10年将增长50倍^[2]。而按照视频编码领域的“摩尔定律”，每10年左右视频压缩效率大概能提升1倍（从1995年的MPEG-2到2005年的H.264/AVC，再到2013/2014年的H.265/AVS2，概莫如此）^[3]。如图2所示，2010年之前视频压缩率基本能满足视频数据量的增长速度；但之后，视频数据量的增长速度远远超过了视频压缩率的提升速度，并且这一鸿沟随着时间推移变得越来越大。

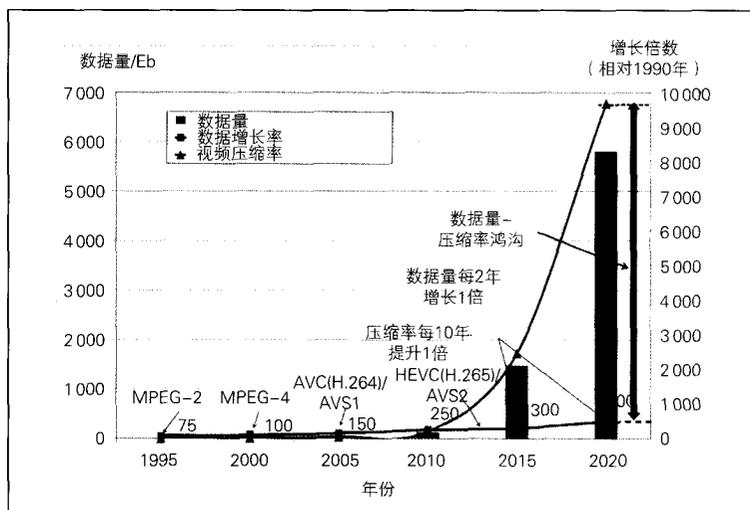


图2 视频大数据的“数据量-压缩率鸿沟”^[3]

存不下问题带来最直接的后果是视频存储成本的急剧增加。以北京为例，目前有50万监控摄像头正在运行，而主流的监控视频编码压缩仍然采用通用的编码标准（如H.264/AVC HP），因此每个高清摄像头每小时将产生3.6 Gb数据。如果每个摄像头至少存储一个月数据，则全北京的监控视频总存储量约1 620 PB。按平均2美元/GB的存储价格计算，北京仅为监控视频所需的存储成本约200亿元。除存储资源外，日益增长的城市视频大数据同时还会带来极大的网络资源和能源消耗。如此巨大的资源消耗，其根源在于现有存储、传输、压缩和编码方法没有充分利用视频大数据中数据间的相关性以及信息分散存储和大规模冗余。因此，研究开发图像视频的超高效表达与压缩编码技术已经迫在眉睫。

国家音视频编解码标准AVS在全球范围内率先启动了针对监控视频的编码技术标准化。这项标准化工作的缘起在于监控视频拥有其自身的特性，其内部的数据相关性可以被进一步挖掘来提升压缩效率。例如大多数固定摄像头拍摄的监控视频中，场景的背景信息相对稳定，这种场景冗余可以被编码算法利用起来有效提升压缩效率。基于上述思想，AVS的监控视频编解码标准AVS-S2在同等编码质量的情况下，比H.264 HP平均节省码率50%~54%。2013年6月，包括AVS-S2在内的AVS视频编解码标准由IEEE颁布为编号1857的国际标准^[3]。在上述基础上，新一代国家音视频编解码标准AVS2进一步扩展到对场景视频的编码，并将其作为基本档次的一部分。目前，AVS2在场景视频编码方面的性能已经大幅超过2013年底颁布的国际标准HEVC/H.265（编码效率提升约40%）。

3.2 找不到 (multi-camera search issue)

从多种多样的大数据中快速发掘出有价值的信息和知识是一个科学挑战问题，也是大数据应用的根本问题。由于图像视频的非结构化特性，如何从视频大数据中“快速找到想要的目标或对象”就成为视频大数据面临的重大难题。现有技术大多采用局部的、单个原始信号层面的表达和分析技术手段，未能利用视频大数据带来的全局信息，因而不能提取有效的信息和知识，所以往往不得不依赖于人工辅助，导致智能自动化程度低，数据处理效率低下。例如目前上海市的监控系统，实行24 h视频监控巡逻模式，需要多达4 000余名专职人员参与到视频监控中，以期在发现可疑情况时及时进行应对和处置。显然，“找不到”问题已经极大地制约了城市视频大数据的有效利用与大规模应用。

多摄像头网络中监控视频的对象标识与检索是有效解决城市视频大数据“找不到”问题的关键，其中特别是目标物体（车辆、行人等）的检测、跟踪与标识。网络监控视频中目标检测与再标识技术（Object Re-identification）的典型应用实例有：2012年初“苏湘渝系列枪击案”，从全南京一个多月、多达几百TB的监控视频中查找案犯周克华的片段；2013年长春“盗车杀婴案”，对于车辆的快速检索与跟踪。从技术角度讲，目标再标识，特别是行人再标识，旨在将某特定区域中多个监控摄像头所监测到的、位于不同位置的同一个目标

（如行人）关联起来，从而形成一个准确和稳定的行人标识（而不需要明确地知道他/她是谁）。针对行人再标识问题，近年来学术界开展了大量的研究，主要的研究思路可分成3种^[4]：基于表观（appearance-based）的方法、基于行为（behavior-based）的方法和其他方法。然而，由于对象间的遮挡、在不同地理位置摄像机的视角位置和光照条件变化等因素，导致了在不同监控摄像头间的目标物体外观外貌特征会产生剧烈的变化，因而目标再标识问题具有极大的技术挑战性，当前的性能水平与实际应用仍有相当大的距离。

在教育部学位与研究生教育发展中心和中国科协青少年科技中心的支持下，数字音视频编解码产业技术创新战略联盟在“中国研究生创新实践大赛”之“研究生智慧城市技术与创意设计大赛”中举办了视频技术挑战赛。比赛的主题即是校园场景下行人和车辆的跨摄像头检测、跟踪和检索。首届比赛从2014年5月1日开始，分技术擂台赛和现场决赛两个阶段。来自全国37所高校和科研院所的132支队伍注册参赛并获取数据，其中56支队伍在线提交擂台赛结果，27支队伍进入了8月22—23日在北京航空航天大学举行的现场决赛。这一系列比赛是目前国内外视频领域比赛中参加队伍数目最多的算法比赛，对于促进大数据时代计算机科学人才培养、推动视频大数据分析技术的研究发展具有重要的意义。

3.3 看不清 (pattern recognition issue)

除搜索应用外，特定目标与行为的识别是城市视频大数据的另一重要应用。车牌识别和人脸识别在过去20多年来得到了长足发展，其中车牌识别在交通电子警察系统中已经大规模使用，而人脸识别则在某些特定应用场景下（例如机场安检）也有很好地应用。然而，全天候真实场景下的人脸与行为识别仍然极具挑战性。其根本原因在于基于模式识别的监控视频分析技术通常基于特定场景摄像头采集的数据，很难推广到不同光照/视角/遮挡程度/室内外等全天候监控视频中。

除更高性能的监控视频解析模型与算法外，城市视频大数据需要突破传统的视频内容分析与处理架构，实现视频编码与分析这两个领域的深度融合。传统的监控视频内容分析大多采用压缩然后分析（Compress-Then-Analyze, CTA）架构，即摄像头端仅进行视频压缩，压缩后的码流通过网络传输到后端服务器，然后在解码后

重构的图像上进行分析 and 识别。

CTA架构具有如下问题：（1）视频压缩的目的是为了节省传输带宽和存储，因此不可避免地对编码后的图像质量造成影响，而在重构图像上进行内容分析与识别也必然受其影响。这一结论已被文献[3]中的实验所证实：在不同的编码参数QP（QP值越大，视频编码码率越小，但重构图像的质量越低）情况下，特征提取和人脸识别任务的性能均随QP增大而迅速降低。（2）视频内容分析与识别任务集中在后端服务器，不利于关注目标、异常行为与事件的实时处理与报警；而与此同时，集中式计算模型耗资源、耗能，大量摄像头端的计算能力并没有得到充分利用。

与之相对照，一种更适合城市视频大数据的新型计算架构则是分析然后压缩（Analyze-Then-Compress, ATC），即利用摄像头进行浅层内容分析与识别，包括背景建模、对象检测与描述、特征提取等，将其分析结果压缩传输到后台进行深度分析，并按需传送视频内容。面向城市视频大数据应用的ATC框架见图3。显然，面向城市视频大数据应用的ATC框架需要新型的视频技术标准支持，包括分析友好的视频编码标准、内容描述标准和特征压缩标准。AVS-S2及AVS2是目前国内外首批对监控视频内容分析友好的视频编码标准，其中新增的背景建模技术和感兴趣区域提取等两项技术，是国外任何其他编码标准所没有使用的^[5]。目前，AVS工作组正在进行监控视频的内容描述标准制定工作，并准备启动视频特征压缩标准的制定，从而将对城市视频大数据看不清的问题提供更好的解决方案。

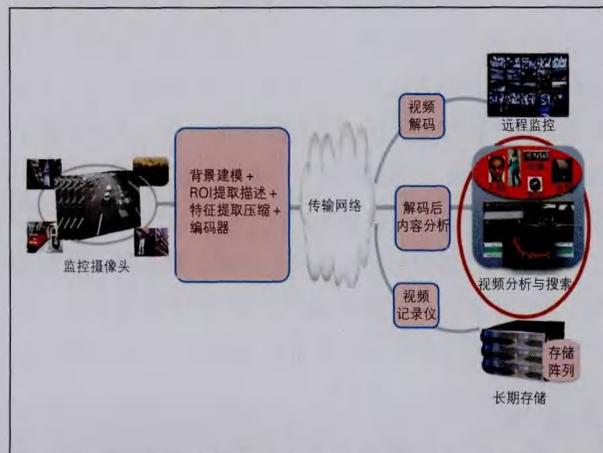


图3 面向城市视频大数据应用的ATC框架

4 结束语

综上所述，智慧城市和大数据已经上升为我国的国家战略。我国是全球人口最多的国家，拥有全球最丰富的数据资源。要推进我国城市化的发展，就迫切需要推进城市大数据特别是视频大数据的研究，充分挖掘我国城市数据资源的巨大价值。城市视频大数据的新型计算模式、新应用服务等对视频感知编码、分析挖掘、计算平台等领域都提出了新的挑战，需要建立高效、可靠的视频大数据技术体系，应对新需求来布局城市视频大数据技术的研发，从而提升花巨资建立的城市视频感知网的应用效能，真正实现“智慧城市”这一战略目标。

参考文献

- [1] 宋刚, 邹伦. 创新2.0视野下的智慧城市[J]. 城市发展研究, 2012(9): 53-61.
- [2] J Gantz, D Reinsel. The Digital Universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East[R], 2012.
- [3] Wen Gao, Yonghong Tian, Tiejun Huang, et al. IEEE 1857 Standard Empowering Smart Video Surveillance Systems[J]. IEEE Intelligent Systems, 2014, 29(5): 30-39.
- [4] S G Gong, M Cristani, S C Yan, et al. Person Re-Identification[M]. Springer, 2014.
- [5] 黄铁军, 张贤国, 田永鸿, 等. 支持监控视频高效压缩与识别的IEEE 1857标准[J]. 电子产品世界, 2013(7): 22-29.

责任编辑 卢敏

收稿日期 2015-12-11