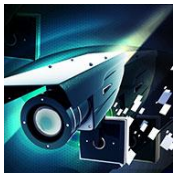
监控视频：最大的大数据

特邀编辑导言 • 黄铁军 • 2014 年 2 月



在指数增长的大数据中，监控视频是最大源头。以此为背景，本期“今日计算”围绕监控视频相关研究，从 IEEE 计算机学会数字图书馆中选择了五篇文章，并针对如何压缩和分析日益增长的巨量视频数据提供了一些参考信息。

数字宇宙中的监控视频

近年来，越来越多的摄像头出现在我们周围，例如电梯里、自动取款机角落里、办公楼墙上以及路边拍摄交通违规的摄像头，还有家里为了照看老人和孩子的摄像头，以及笔记本电脑屏幕上方和嵌在手机前后两面的摄像头。这些摄像头每天都在拍摄巨量视频并送入赛博空间。像北京或伦敦这种城市安装的摄像头数都已上百万，这些摄像头每个小时采集的视频都超过英国广播公司（BBC）或中国中央电视台（CCTV）收藏的节目总量。据国际数据集团（IDC）不久前的《The Digital Universe in 2020》（2020 年的数字宇宙）报告，全球大数据——即数字宇宙中有分析价值的部分——在 2012 年有一半是监控视频，这个比例在 2015 年会上升到 65%。

为了了解有关监控视频的研究开发活动，我用“video”和“surveillance”这两个词搜索了 IEEE Xplore（只在元数据中）和 IEEE CSDL（只限精确匹配）。搜索结果显示在 IEEE 会议、期刊和杂志论文中有 6,832 篇（Xplore）和 3,111 篇（CSDL）相关论文。图 1 是一个年度分布统计，过去十年相关论文数高速增长，表明关于监控视频的研究相当活跃。

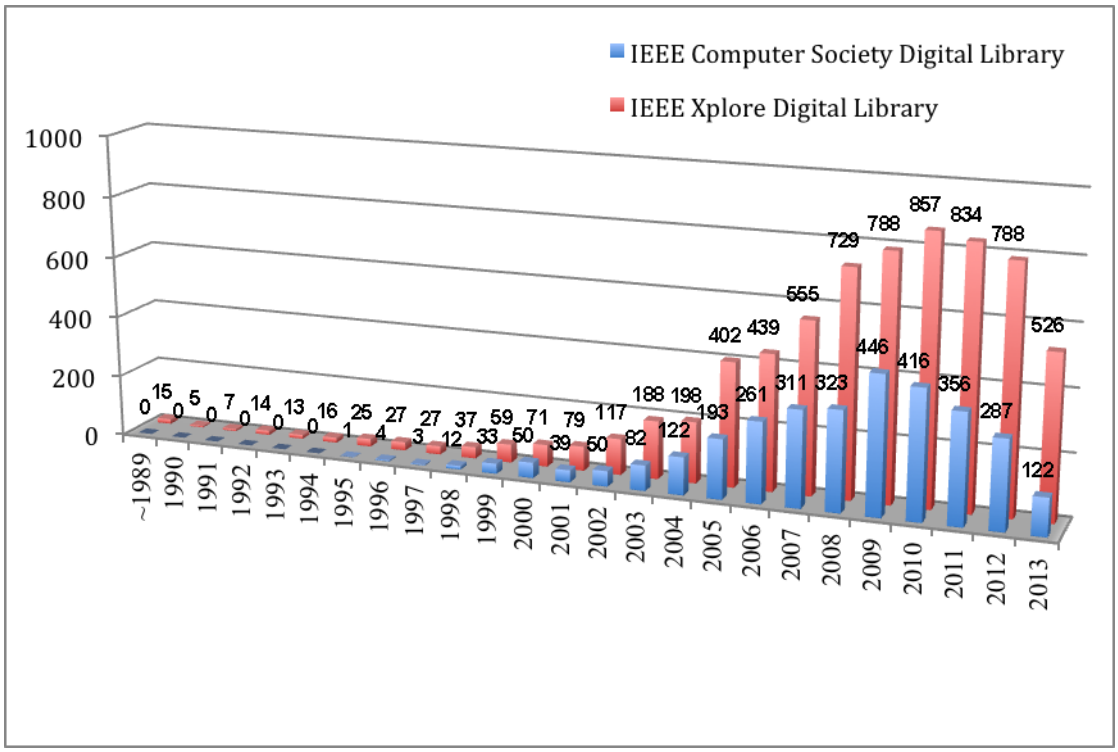


图 1. IEEE 计算机学会数字图书馆和 IEEE Xplore 论文库中元数据包含 *video* 和 *surveillance* 这两个关键词的论文数量年度统计

注："~1989"是指直到 1989 年的所有论文数。2013 年的论文数可能还会增长，因为还有部分论文可能尚未归档到数据库中。

主题论文

监控视频大数据带来了很多技术挑战，包括压缩、存储、传输、分析和识别，其中两个最急迫的问题是如何高效地传输和存储这些巨量数据以及如何智能分析和理解其中蕴藏的视觉信息。

为了降低监控视频大数据的存储和传输代价，急需更高效率的视频压缩技术。今日计算 2013 年十月号介绍的最新视频压缩标准 HEVC（High Efficiency Video Coding）可以把视频压缩到原始数据量的 3%左右，换句话说，HEVC 比 2003 年完成的 H.264/MPEG-4 AVC 标准压缩效率提高了一倍，后者的压缩效率比 1993 年完成的更早一代标准 MPEG-2/H.262 高一倍。尽管有这些进步，压缩效率十年翻一番的速度还是大大落后于物理世界中监控视频的增长速度：平均两年翻一番！

为了进一步提高压缩效率，需要在设计视频编码标准时考虑监控视频的特性。例如，不像标准视频，监控视频往往是在一个特定地点日复一日、月复一月地采集的。但是，以前的标准都没考虑监控视频中才有的这种冗余（例如不变的背景，或者出现了很多次的前景对象）。名为《Standard for Advanced Audio and Video Coding》的新 IEEE 1857 标准专门设置了一个监控档次，可以更好地消除背景冗余，效率比 AVC/H.264 高一倍，而且复杂度还更低。在《IEEE 1857 Standard Empowering Smart Video Surveillance Systems》这篇文章中，高文、我的同事和

我概要介绍了这个标准，特别是其基于背景模型的编码技术和识别友好的功能特性。这种新方法也已经用于增强 HEVC/H.265，也可以把压缩效率几乎提高一倍。（更多技术细节可参见《Background-Modeling Based Adaptive Prediction for Surveillance Video Coding》，IEEE Xplore 订户能够访问）。

就像物理宇宙一样，数字宇宙中也存在大量的暗物质：它就在那里，但是我们知之甚少。据上面提到的 IDC 报告，数字宇宙中 23%的数据如果做了标记就可以作为大数据进行分析利用，但是现有技术还远远赶不上需求：潜在有用数据中只有 3%被标记了，得到分析的甚至更少。事实上，出现在数百万摄像头中的人、车和其他运动物体是利用机器分析了自动理解复杂社会和世界的丰富资源。就像特邀编辑 Dorée Duncan Seligmann 在今日计算 2012 年四月主题中谈到的，与其他数据类型相比，对视频进行自动分析和理解的挑战性更大。本月主题增加其后发表的三篇相关文章。

人通常是监控视频分析最感兴趣的主要对象。在《Reference-Based Person Re-identification》这篇 2013 IEEE International Conference on Advanced Video and Signal-Based Surveillance （AVSS）的最佳论文中（IEEE Xplore 订户可以访问），Le An 和他的同事提出了一种基于参考的模型，通过学习一个子空间把来自不同摄像机的参考数据的相关性进行最大化，从而能够从光照差异较大的多个摄像机的视野中识别出同一个人。

再深一步就是对人的行为进行分析。Shuiwang Ji 及其同事在《3D Convolutional Neural Networks for Human Action Recognition》一文中把深度学习引进人的动作识别，提出的三维卷积神经网络通过三维卷积从空间和时间两个维度抽取特征，从而可以获得多个关联视频帧中的运动信息。对机场视频的实验表明性能比对比方法有明显提高。

在《Monocular Visual Scene Understanding: Understanding Multi-Object Traffic Scenes》中，Christian Wojek 和他的同事报告了一个创新的三维概率场景模型，它把三维几何推理和最新的多类对象检测、跟踪和场景标注集成在一起。只使用单视摄像机，该模型就可以通过推断同时完成三维场景上下文的恢复和三维多对象的跟踪。这篇文章使用车载摄像机拍摄的多个挑战性很强的视频序列进行了评估，实验表明该方法与最新三维多人跟踪和三维多类车辆跟踪方法相比，性能有明显提升。

走向场景视频时代

本月主题还包括一段来自 EMC 集团 CTO John Roese 的视频，发表了他对这一主题的见地。

[视频]

像监控视频一样，在教室、法庭以及其他特定场合采集的视频也在快速增长。

这实际上是一个“场景视频”时代的前奏：在这个时代里，绝大多数视频都是从特定场景采集的。不久的将来，无处不在的摄像机将把触角伸到人类能够触及的所有空间。

在这个新时代，“场景”将成为连接视频编码和计算机视觉的桥梁。对场景建模可以进一步提高压缩效率，就像 IEEE 1857 已经展示的那样。进而，利用视频流中已经编码的这些场景信息，前景对象的检测、跟踪和识别就会更容易。从这个意义上讲，监控视频以及其他场景视频的大规模增长对视频和视觉相关研究来说，既是巨大挑战，也是巨大机遇。

2015 年，IEEE 计算机学会的多媒体计算技术委员会（TCMC）和语义计算技术委员会（TCSEM）将共同举办首届 IEEE 多媒体大数据国际会议（the first IEEE International Conference on Multimedia Big Data），多媒体大数据研究、开发和应用这一极其活跃的领域中的前沿学者将汇集一堂。欢迎有兴趣的读者围绕快速增长的多媒体大数据，明年春天来北京，在这个新创办的盛会上开怀畅谈。

引用

T. Huang, "Surveillance Video: The Biggest Big Data," Computing Now, vol. 7, no. 2, Feb. 2014, IEEE Computer Society [online]; http://www.computer.org/portal/web/computingnow/archive/february2014.

黄铁军是北京大学信息科学技术学院教授，数字媒体技术研究所所长。他从华中科技大学图像识别与人工智能研究所获得博士学位。他的研究领域是视频编码、图像理解、数字版权管理以及相关标准制定。他入选 2010 年度教育部新世纪优秀人才计划。他是“今日计算”顾问委员和中国地区代表，负责今日计算中文翻译工作。可通过 tjhuang@pku.edu.cn 联系他。