

# 视频大数据研究综述

刘祥凯 张 云 张 欢 李 娜 樊春玲 谢祖庆 朱林卫

(中国科学院深圳先进技术研究院高性能计算技术研究中心 深圳 518055)

**摘 要** 科学技术与生产力的发展带来了数据量的高速增长, 其中视频图像等多媒体数据占了很大的比重。如何高效处理这些海量数据并从中快速挖掘有价值的信息是当前的研究热点。通常大数据具有四个特点, 即数据量大、需要快速响应、数据类型多样和价值密度低。视频大数据同样具有以上特点, 但其特殊性在于数据冗余更大, 需要进行高效的压缩编码与分析处理。总的来说, 视频大数据的研究内容包括了视频数据表示、智能视频分析、视频压缩与传输、视频显示与评价等方面。在发展趋势上, 视频数据的表示将向真实感与智能化两个方向发展; 智能视频分析技术将会借助深度神经网络获得更准确的识别分类结果; 视频压缩技术在提升压缩效率的同时也会探索降低编码复杂度的方法, 并通过结合人眼视觉感知特性的编码算法来减少视频大数据的视觉冗余; 视频显示设备将伴随着视频数据表示形式的改变而进行相应的升级换代; 视频质量的评价准则将由单一的图像质量评价向更加综合全面的用户体验质量评价发展。

**关键词** 大数据; 视频分析; 视频编码; 视频质量评价; 用户体验质量

**中图分类号** TP 37 **文献标志码** A

## A Review on Video Big Data

LIU Xiangkai ZHANG Yun ZHANG Huan LI Na FAN Chunling XIE Zuqing ZHU Linwei

(*Research Center for High Performance Computing, Shenzhen Institutes of Advanced Technology,  
Chinese Academy of Sciences, Shenzhen 518055, China*)

**Abstract** The developments of science and technology have brought rapid growth of data, of which video and image data account for a high percentage. How to efficiently handle these data and find valuable information from them is a hot topic. Big data are characterized by four Vs: volume, velocity, variety, and value, representing large amounts of data, quick data processing, various data types, and low value density, respectively. Video big data share all these characteristics, and often come with much greater data redundancy than other types of data. As a result, they call for more efficient techniques for compression and processing. The

收稿日期: 2015-11-20 修回日期: 2016-01-04

基金项目: 国家自然科学基金项目(61471348); 广东省特支计划-科技创新青年拔尖人才(2014TQ01X345); 深圳市孔雀人才创业创新项目(KQCX20140520154115027); 广东省自然科学基金-博士启动(2015A030310262)

作者简介: 刘祥凯, 博士研究生, 研究方向为视频信号处理、质量评价; 张云(通讯作者), 博士, 副研究员, 硕士研究生导师, 研究方向为多媒体信号处理与通信, E-mail: yun.zhang@siat.ac.cn; 张欢, 博士研究生, 研究方向为字典学习与稀疏编码; 李娜, 博士后, 研究方向为图像与视频分析; 樊春玲, 博士研究生, 研究方向为图像与视频分析; 谢祖庆, 硕士研究生, 研究方向为多媒体系统设计; 朱林卫, 研究助理, 研究方向为视频编码。

research of video big data is primarily carried out along four dimensions: video data representation, intelligent video analysis, video compression and transmission, and video display and quality evaluation. Recent trends show that video representation is becoming more realistic and intelligent, and video analysis more accurate in identification and classification thanks to the deep neural networks. At the same time, video compression promises to be more efficient with new methods to reduce coding complexity, and less redundant with the help of visual perception aware coding algorithms. In accordance with more advanced video representation, video display devices are undergoing hardware upgrades, guided by a comprehensive methodology of video quality evaluation that is centered around quality of experience, instead of the traditional criteria developed for image quality assessment.

**Keywords** big data; video analysis; video coding; video quality assessment; quality of experience

## 1 引言

随着互联网和移动端智能设备的快速发展,多媒体视频与图像应用越来越广泛,已经影响人们生活的方方面面,包括国家/城市安全、医疗、教育、通讯、工业生产以及文化影视等,可谓“无处不在”且起着“至关重要”的作用。这些视频应用每时每刻产生海量的视频数据,目前已约占人们通讯数据量的80%。在国外,仅YouTube一个视频网站就拥有10亿用户,每分钟有300小时视频被上传,5000万小时视频被观看。在中国,网络视频用户超过4.33亿,其中手机视频用户有3.13亿<sup>[1]</sup>。至2014年,移动视频数据占移动数据通信总量的55%,每个月数据1.38 EB,即 $1.38 \times 10^6$  TB。如果以25 Mbps的1080i高清HD视频观看,5分钟1 GB,每个月数据1.38 EB可以观看14000年。思科Cisco预期至2019年,移动视频数据将占设备数据通信总量的72%以上<sup>[2]</sup>。除了互联网视频之外,监控视频是视频数据的主要来源。中国大约有3000万台监控摄像机,每月将生成60 EB的视频数据,根据互联网数据中心的统计报告,至2015年监控视频数据已占全球视频数据的一半以上。全球视频数据呈现爆炸式增长,容量每两年翻一

番,急需有效的视频处理、编码压缩、分析与管理等技术。本文将从四个方面来对视频大数据的相关研究进展进行介绍,包括视频数据的表示、智能视频分析、视频压缩与存储、视频显示与质量评价,各模块的关系链如图1所示。

## 2 视频与图像数据表示形式

视频与图像数据表示形式的发展趋势可大致分为两个方向:更真实和更高效,如图2所示。一方面,超高清分辨率、高动态(High Dynamic Range, HDR)、广色域(Wide Color Gamut)和三维(Three Dimensional, 3D)视频等技术使得视频表示效果越来越接近人眼对世界的真实感受,更为逼真;另一方面,针对监控视频、屏幕内容视频等具有独特内容特性的视频数据以及针对智能视频分析等特殊应用需求,需要设计专用的更高效视频数据表示方式。

### 2.1 更真实的数据表示

为获得更真实的观看体验,目前视频数据表示形式正在向更高分辨率、更高帧率、高动态、广色域和3D视频方向发展。国际电信联盟(International Telecommunications Union, ITU)在2012年推出的ITU-R BT.2020建议书<sup>[3]</sup>中定

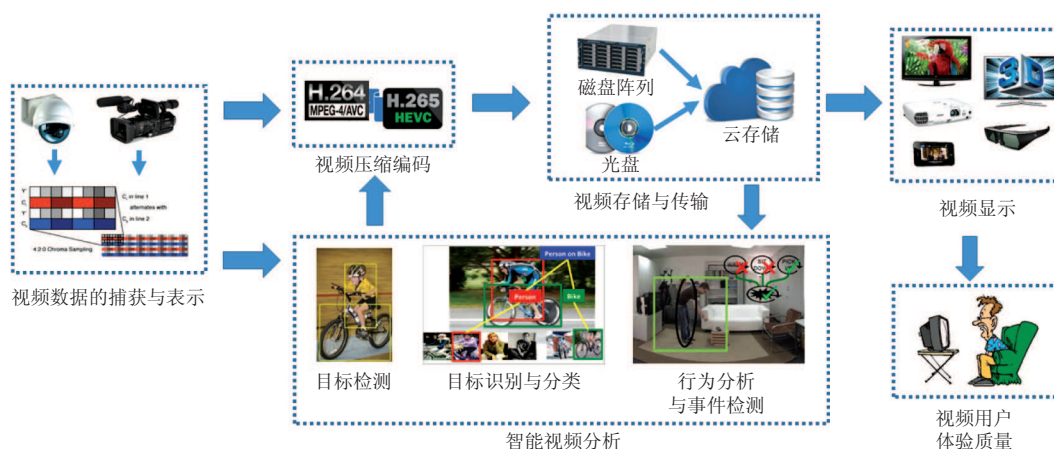


图 1 视频大数据研究框图

Fig. 1 Framework of video big data research

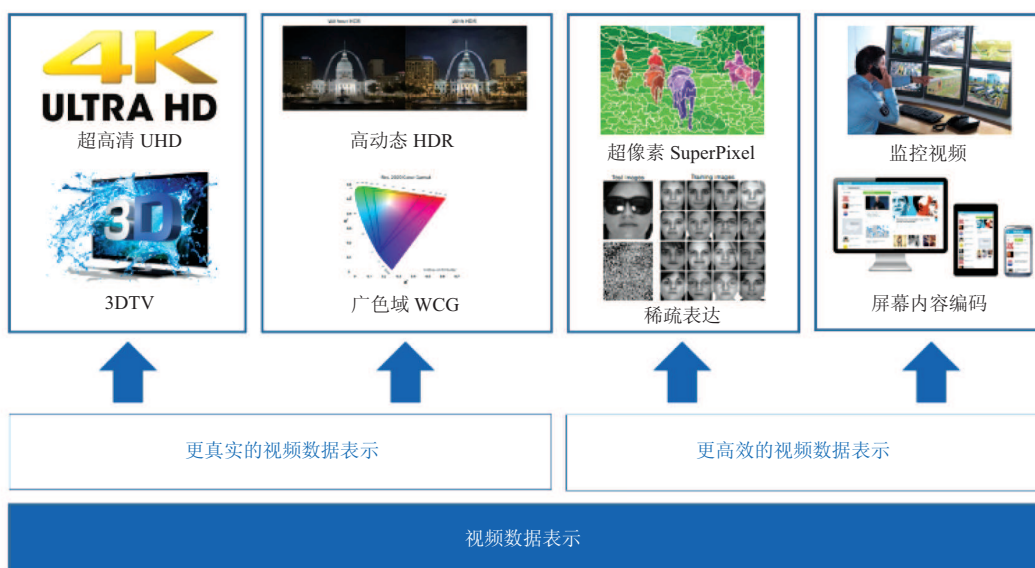


图 2 视频数据表示

Fig. 2 Video data representation

义了超高清电视(Ultra High Definition, UHD)的 4K(3 840×2 160)和 8K(7 680×4 320)分辨率标准,并将视频的最高帧率范围由原来的 25 Hz 提升到了 120 Hz。画面细腻程度和运动图像的连贯程度将达到人眼可分辨极限。另外,BT.2020 标准将电视的色域范围扩展到了之前标准的两倍,已经可以覆盖到 CIE1931 视觉标准色域的 63.3%。为提供更高的亮度和色彩动态范围,更

真实的图像对比度,该标准将目前的 8 比特每通道(256 级)增加到 10 比特和 12 比特。然而,伴随着更加真实震撼的视觉体验,随之而来的是视频数据量的大幅增长。4K 分辨率 60 Hz 帧率的 HDR 视频数据量的大小是目前全高清视频(1 080 p)的 12 倍。

人们双眼所观察到的真实世界是三维立体的,但传统视频只能提供平面 2D 画面,缺乏

深度感和立体感。双目立体以及多视点 3D 视频在原有的 RGB 颜色数据之外增加了场景的深度(Depth)维度,对视频数据表示方式带来的变革不亚于从无声电影到有声电影的跨越。三维场景表示主要可划分为基于图像的表达方式(Image Based Representation)和基于几何的表达方式(Geometry Based Representation),如图 3 所示。基于图像的表达方式主要通过密集相机拍摄 3D 场景,并通过插值等图像处理方法生成虚拟视点图像,其优点是无需三维场景建模,缺点是需比较密集的相机阵列进行图像采集,数据量巨大。典型的基于图像的 3D 场景表示方式有光线空间(Ray Space)、光场(Light Field)和光亮度(Lumigraph)等。基于几何的三维表示方式目前主要应用于游戏和动画等,大多数情况下是通过三维网格(3D Mesh)、点云等进行描述:先对真实场景中的物体建立模型,然后通过渲染技术将图像纹理渲染于三维模型表面。其优点是可以实现任意视点角度图像的绘制,数据量较小;缺点是自然场景建模困难,真实性较差。鉴于以上分析,结合基于图像和基于几何的表达方式提出多

路视频加深度信息(Multiview Video plus Depth)格式<sup>[4]</sup>:通过多视点图像提供高逼真纹理内容,同时通过深度图表示三维几何信息。目前该格式已被最新 3D 视频系统标准采纳为其主流表示形式。

## 2.2 更高效的数据表示

目前,在视频数据中,有的来自于摄像机拍摄(如自然场景),有的来自于计算机生成(如 Flash 动画、游戏视频、屏幕内容视频),但即使同样都是由摄像机拍摄的视频,其内容特性也会随视频场景的不同而差异显著。目前视频主要基于块和像素的表示方式,缺乏对特定场景视频内容特性的考虑,如监控、桌面视频以及医疗等特定场景,难以实现压缩效率最大化。监控视频,由于背景信息在较长时间内保持不变,可采用基于背景建模和运动对象编码的表示方式<sup>[5,6]</sup>,便于视频后续高效压缩、分析以及检索。屏幕内容视频,主要由计算机产生,一般具有明显像素边缘和较多高频分量,且内容相关性较弱。屏幕内容视频数据理想表示及编码方法需可针对非连续色调区域提高压缩效率<sup>[7,8]</sup>。

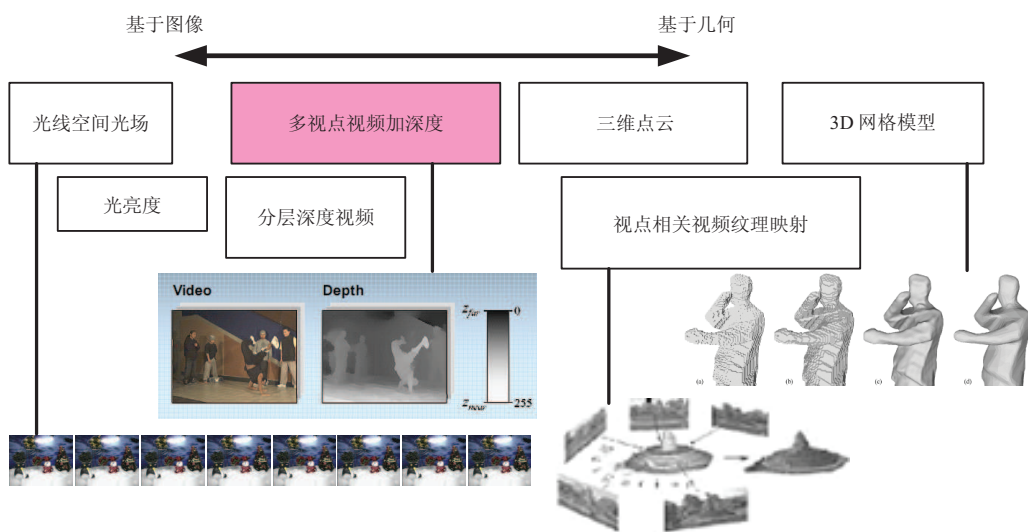


图 3 三维场景的数据表示方式

Fig. 3 Data representation of 3D scene



随着计算机视觉与模式识别等研究领域的快速发展, 视频的对象检测、识别与跟踪以及分类、检索等智能分析技术具备了广阔的应用前景。进行视频对象识别与分类时, 传统基于块和像素的表示方式缺乏视频对象物理意义, 不利于理解, 因此, 需要更能够体现视频内容特征的数据表示方式。例如超像素(Superpixel)分割表示方式, 它有利于视频对象分割与提取, 而图像的稀疏表示不仅可以用于底层图像去噪和高分辨率重建, 还可用于高层语义的对象识别与分类。

超像素指由具有相似性质、空间上相邻的像素所组成的块<sup>[9]</sup>。超像素因其局部一致性, 在人眼视觉感兴趣尺寸上保存了分割所需大部分结构信息。图像采用超像素代替像素表示后, 信息单元数目大量减少, 能提取出更稳定的特征, 方便应用更加复杂的图像处理算法。超像素块的获取方法可以分为两类<sup>[10]</sup>: 基于图论的方法和基于梯度下降的方法。其中, 基于图论的方法有归一化分割方法<sup>[11]</sup>、最小生成树分割方法<sup>[12]</sup>、超像素晶阵方法<sup>[13]</sup>和基于熵率的方法<sup>[14]</sup>等; 基于梯度下降的方法有分水岭方法<sup>[15]</sup>、均值漂移方法<sup>[16]</sup>和基于几何流的算法<sup>[17]</sup>等。对应于图像中的超像素, 视频应用中引入了超体素(Super-voxel)的概念, 主要考虑时间维度上视频内容的一致性<sup>[18]</sup>并应用在后续目标分割与跟踪等环节<sup>[19]</sup>。

自然图像不同于随机的二维信号, 一般符合高斯模型等特有的统计特征。由于图像具有能量局部集中性<sup>[20]</sup>和稀疏表达形式, 一幅图像可通过限制其表达系数的稀疏性由一组基函数的线性叠加组合表示, 这组基函数可以称为字典, 该字典是超完备的且其字典基是线性相关的<sup>[21]</sup>。图像的稀疏表达是根据图像信号本身具有的稀疏性特点, 将图像数据表示成一组字典基的稀疏线性表示方法。基于图像可以在完备字典的基础上进行稀疏分解的假设, 图像去噪、超分辨率重建可以看成是信号恢复问题<sup>[22,23]</sup>。从降维的角度来看,

稀疏表达可用于降维后空间的信号分类<sup>[24]</sup>。通过在构造字典的优化目标函数里同时权衡重构表达能力、分类性能和表达系数稀疏性三个指标, 图像稀疏表达在目标识别、分类、跟踪等问题上有许多成功应用<sup>[25,26]</sup>。

### 3 智能视频分析

#### 3.1 视频分析的层次和系统结构

全球视频数据正在呈现爆炸式增长, 如何从海量视频数据中快速找到感兴趣目标成为一项很有挑战性的工作。智能视频分析技术力图解决“什么对象”、“在什么地方”、“发生什么行为”等问题。视频分析的层次和系统结构如图 4 所示, 大致可分为感知层、认知层和高层语义层。其中, 感知层提供的信息包括色彩、形状、纹理、声音信号等, 实现低层次语义信息提取; 认知层的内容主要是识别视频中所显示的一些对象(如人物、动物、植物)以及这些对象的运动及相互关系等; 高层语义的推理主要根据已有专家知识库, 辅以模式规则发现等, 实现高层语义与知识提取。

#### 3.2 智能视频分析技术

在视频分析系统中, 感知层和认知层是目前实施智能视频分析技术的关键和研究重点, 主要包括目标检测、目标跟踪、目标识别与分类等研究内容。

##### 3.2.1 目标检测

目标检测一直是计算机视觉中的关键问题, 过去十几年研究工作主要关注特定目标检测和显著目标检测, 尤其 ImageNet 2012-2015、PSACAL VOC 等数据集的出现较好地推动了图像和视频中一般目标检测的研究。面向图像的目标检测研究, 主要围绕目标不同程度的形变特征表示, 代表性的有基于可形变部分模型和空间金字塔匹配模型的目标检测器。对于面向视频的

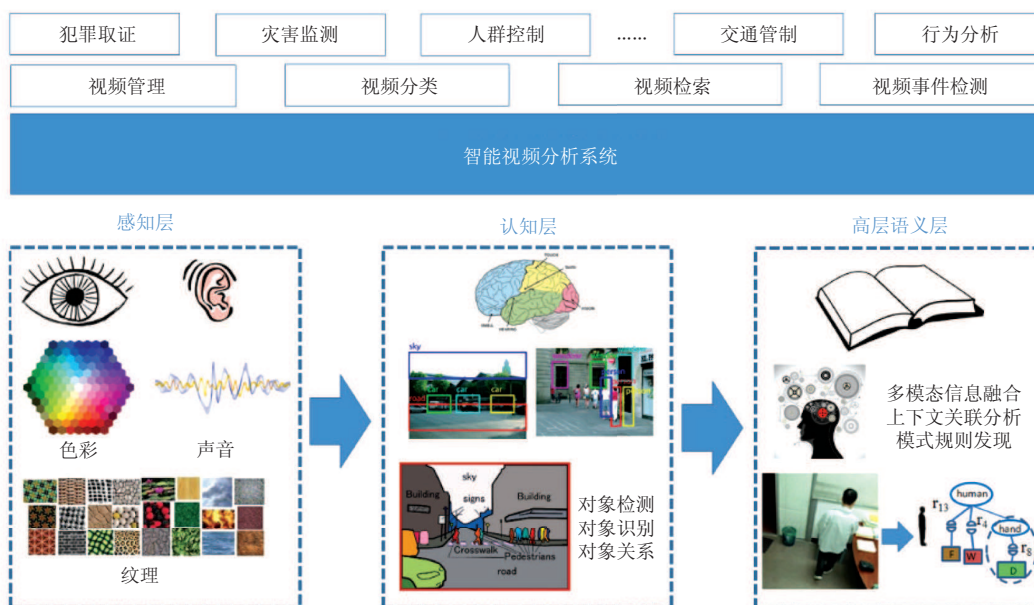


图4 视频分析的层次和系统结构

Fig. 4 Structure of video analysis system

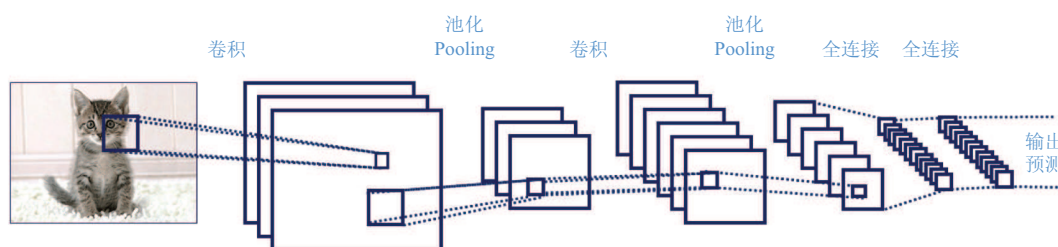


图5 卷积神经网络示意图

Fig. 5 Convolutional neural network

目标检测，目标检测器一般首先基于图像集合训练，然后针对视频帧序列的时序约束和目标动态变化实现目标检测器在线优化<sup>[27]</sup>。

近年来，随着深度学习模型如卷积神经网络(Convolutional Neural Network, CNN)<sup>[28]</sup>的应用，图像中一般性目标检测的性能得到了很大提升，CNN网络样例如图5所示。具有代表性的目标检测是2014年Girshick等<sup>[29]</sup>提出的区域卷积神经网络(Region-based Convolutional Neural Network, R-CNN)框架：首先对每幅输入图像抽取若干个候选窗口，然后利用CNN网络学习候选窗口的特征，最后采用支持向量机(Support

Vector Machine, SVM)分类器进行分类。R-CNN方法相比其他方法精度较高，但训练和测试时间开销比较大。

### 3.2.2 目标跟踪

在目标跟踪方面，卡尔曼滤波、隐马尔科夫、粒子滤波器等数学模型常被用来进行运动状态估计。代表性的跟踪方法有基于光流的Kanade-Lucas-Tomasi(KLT)算法<sup>[30]</sup>，但前提假设包括：相邻帧之间的亮度恒定、时间连续或者运动比较“微小”以及保持空间一致性。Zdenek等<sup>[31]</sup>提出一种单目标长时间跟踪算法，主要包括追踪器、检测器和机器学习三个模块。该算法

将传统的跟踪算法和检测算法相结合用于解决目标形变、部分遮挡等问题。同时,通过一种改进的在线学习机制不断更新跟踪模块的“显著特征点”和目标检测模块的目标模型及相关参数,从而使得跟踪效果更加稳定。然而,视频目标跟踪方面目前还没有统一体系,能够适用于复杂场景的跟踪算法还有待进一步的研究和探索。

### 3.2.3 目标识别与分类

传统的以贝叶斯决策理论、线性分类器、非线性分类器为代表的判别决策理论与机器学习是应用广泛的模式识别方法,例如,支持向量机就是一种二类线性分类器,也可扩展到非线性应用。除传统的机器学习算法以外,深度学习模型是近年目标识别与分类的研究热点。

深度学习网络首先利用卷积神经网络对图像块提取局部特征,然后将提取的特征输入到线性 SVM 分类器进行分类。Simonyan 等<sup>[32]</sup>提出一种基于两个 CNN 模型的双流卷积神经网络对视频进行分类,其中一个 CNN 用于提取原始图像空域特征,另一个 CNN 将多幅图像计算得到的光流图像作为输入,附加作为视频时域信息。该算法加入运动信息提高视频分类精度。Joe 等<sup>[33]</sup>在双流卷积神经网络之后加入基于长短时记忆单元(Long-Short Term Memory, LSTM)<sup>[34]</sup>的递归神经网络(Recurrent Neural Network, RNN)<sup>[35]</sup>。RNN 可更好地对时序信息进行建模,同时 LSTM 解决梯度爆发和消失问题。Wu 等<sup>[36]</sup>提出一种混合深度学习框架将时空信息用正则化约束融合起来。该框架包括三部分:提取空间图像特征的 CNN 与 LSTM 网络,提取堆叠运动光流信息的 CNN 与 LSTM 网络和基于正则化约束的两种特征融合网络。整个混合框架在数据集 UCF-101 上分类准确度达到 90.1%。

目前,深度学习网络在计算机视觉领域的目标识别与分类表现已经超过了传统模式识别方法,并且具有较好的迁移学习特性,可通过对已

训练深度学习模型进行微调后用于解决其他分类问题。但是,深度学习网络的训练非常耗时,参数多且调整最优参数的过程复杂。针对视频的运动等特征,目前还没有理想可用的初始化网络,有待于进一步研究以促进视频分类、识别的发展和

## 4 视频的压缩与存储技术

随着视频图像的广泛应用,视频数据容量急剧增加,高效的视频编码压缩技术变得至关重要。视频压缩的目的是为了保证在维持一定质量的前提下最大化视频压缩比。目前,主要通过预测编码技术消除视频空间和时间上的相关性,即时空冗余;通过熵编码等消除信号冗余;通过变换量化的有损压缩以及色彩空间变换等消除部分感知冗余。由此,通过减少数据冗余的方式,最大化压缩比。

### 4.1 视频压缩标准的发展

目前国际上主要有两大组织负责视频编码标准的制定,一个是国际标准化组织的运动图像专家组(Motion Picture Expert Group, MPEG),另一个是国际电信联盟的视频编码专家组(Video Coding Expert Group, VCEG)。图 6 为国际视频压缩标准的发展历程与时间节点。MPEG 制定的标准为 MPEG-X 系列,主要应用于存储、广播电视、流媒体等多媒体领域内的相关标准,包括 MPEG-1、MPEG-2、MPEG-4 等。其中, MPEG-1 为 VCD 采用的压缩标准, MPEG-2 为 DVD 采用的压缩标准。VCEG 主要面向视频电话、视频会议等实时视频通信领域,制定的标准为 H.26x 系列,如 H.261、H.262、H.263 等。MPEG 和 VCEG 的专家组成员于 2003 年组成联合视频专家组(Joint Video Team, JVT),联合制定了视频编码标准 H.264/AVC(Advanced Video Coding)<sup>[37]</sup>,亦即 MPEG-4/AVC 的第 10 部分,在保持与



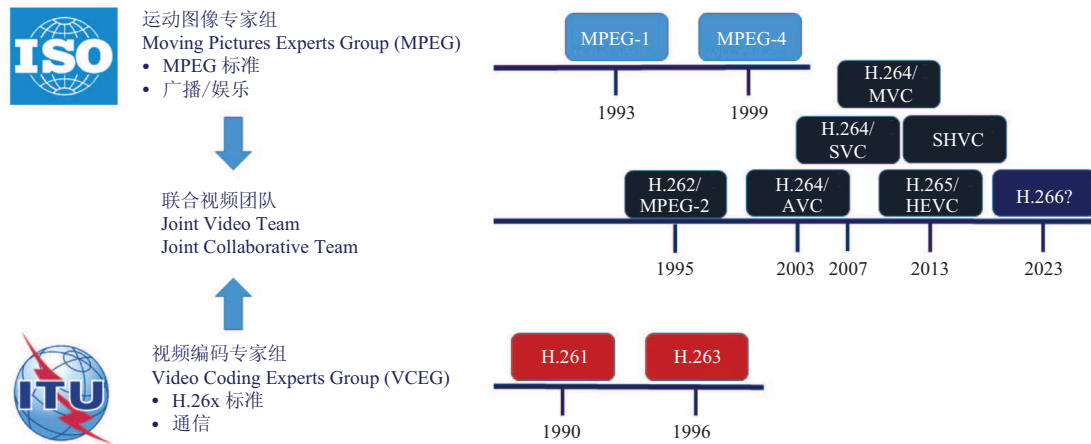


图6 视频压缩标准的发展历程

Fig. 6 Development of video coding standards

MPEG-2 相同视频质量的前提下将视频的压缩比提高一倍。为了进一步扩展 H.264/AVC 标准的功能和应用, JVT 相继公布了 3 个扩展标准, 它们分别是高保真范围扩展 Frext、4:4:4 格式扩展和可分级视频编码<sup>[38]</sup>。基于 H.264/AVC 的可分级视频编码系统主要通过一次编码形成一个基本层码流和多个增强层码流, 并根据网络条件选择码流传输与解码以提高网络适应性。考虑到多视点与 3D 视频的应用前景, JVT 将多视点视频编码作为 H.264/AVC 的第四个扩展标准<sup>[39]</sup>以支持 3D 视频压缩与视点选择等扩展功能。

2010 年, MPEG 和 VCEG 再度联手成立了视频编码联合专家组 (Joint Collaborative Team, JCT), 并于 2013 年 1 月正式发布面向高清和超高清等视频应用的新一代视频编码标准<sup>[40]</sup>——高效视频编码标准 (High Efficiency Video Coding, HEVC), 即 H.265。该标准相比于 H.264/AVC 压缩效率又提高一倍, 即相同视频质量前提下码率降低 50%。2014 年 10 月, JCT 颁布了 HEVC 第二版, 增加了 3 个扩展附录。其中, 附录 F 为多层扩展 “Common Specification for Multi-Layer Extension”, 简称 MLxT; 附录 G 为多视点扩展 “Multiview HEVC”, 简称 MV-HEVC; 附

录 H 为可分级扩展 “Scalable HEVC”, 简称 SHVC<sup>[41]</sup>。3D 扩展 “3D-HEVC” 拟定为 HEVC 的附录 I。从概念上说, 附录 G、H、I 这 3 项 HEVC 扩展都隶属于多层视频编码技术<sup>[42]</sup>, 其共同的编码结构和高层语法规则由附录 F 定义。

在国内, 2002 年 12 月开始, 国家信息产业部科学技术司成立 “数字音视频编码技术标准工作组” (简称 AVS 工作组) 致力于制定具有自主知识产权的视频编码标准, 并于 2003 年 12 月成功完成基准档次视频编码标准的制定<sup>[43]</sup>, 在 2006 年公布为国家标准。此后, AVS 工作组又陆续制定了面向移动、监控等应用的编码标准, 并于 2012 年通过成立 AVS 技术应用联合推进工作组制定了面向高清晰数字电视广播的 “AVS+” 标准, 如今 “AVS+” 已在数字电视广播等应用中取得重要进展。此外, 2011 年底 AVS 还启动了面向超高清应用的新一代视频编码标准 AVS2.0<sup>[44]</sup>的制定工作, 目前已接近完成, 正在制定的 AVS2.0 标准已经取得相比于 AVS1.0 标准 40% 左右的码率节省。

依据三代视频压缩标准的发展历程, 大约每隔十年视频的压缩效率提高一倍。但是, 视频数据量的增加速度远比视频压缩效率的提高快得



多, 全球视频数据总量每两年翻一番, 所以迫切需要更有效的视频压缩方法以解决视频传输与存储等问题。

## 4.2 国内外视频编码算法的研究

由于视频标准定义了码流的语法与语义约束解码器, 编码器则是开放式的, 在标准制定后可在符合码流语法和语义的前提下, 仍不断进步和提升, 以进一步提高压缩效率。现有高效编码主要研究可以归结为两类优化: 一是如何进一步提高编码效率; 二是如何有效降低编码复杂度。

(1) 高效视频编码方法: 为了提高编码性能, 最新高效视频编码标准采纳了多项先进编码技术<sup>[40]</sup>, 主要包括四叉树单元结构、残差四叉树变换结构、像素自适应差值以及自适应熵编码技术等。为了充分利用图像的空间相关性, HEVC 中共定义了 35 种帧内预测方向和更先进的帧间插值与预测技术<sup>[45]</sup>。为了有效分配资源和改善编码性能, 研究人员提出背景参考帧图像重组和预测技术<sup>[6]</sup>、自适应滤波技术<sup>[46]</sup>以及码率控制技术<sup>[47]</sup>等, 实现编码效率的最大化。以上方法主要通过大量增加编码模式、优化参数和全遍历搜索技术等消除冗余, 以计算量换取增益。鉴于挖掘视频中存在的视觉冗余是压缩效率的另一增长点, 研究人员提出了感兴趣区域编码<sup>[48]</sup>、恰可察觉失真编码<sup>[49]</sup>、非对称立体视频编码<sup>[50]</sup>等算法, 笔者所在课题组将视频编码与视觉感知相融合, 提出高效编码和码率分配方法<sup>[51-62]</sup>, 有效提升网络适应性和编码效率。

(2) 低复杂度高效视频编码方法: 鉴于高效视频编码 HEVC 过程引入分层可变尺寸编码单元、多尺度预测单元、变换单元、多参考帧运动估计等技术, 引入越来越多的模式与自适应参数, 以增加计算量的方式提高预测精度并换取编码增益, 却极大地增加了计算复杂度, 不利于规模化处理和实现。为了降低编码复杂度, 研究人员针对高效编码器提出帧内编码预测方向的快速

确定、帧内模式的快速自适应选择、运动估计最佳匹配块的快速搜索<sup>[63]</sup>、多参考帧的选择<sup>[64]</sup>等。主要通过利用概率统计和时空相关性等信息, 并结合优化理论实现编码过程早期终止和决策。笔者所在课题组也在低复杂度编码优化中从事大量相关研究工作<sup>[63-67]</sup>。随着编码模式、参数的不断增加, 传统基于少量特征的统计分析方法在区分性和普适性上渐渐落后, 基于数据驱动和学习的高效编码框架是进一步的研究方向。

## 4.3 视频存储技术

视频存储技术按存储媒介大致可分为以光盘为载体和以硬盘为载体两类。光盘视频存储技术主要面向消费领域, 一张 VCD 光盘容量大约为 650 MB, DVD 容量大约 4.7 GB, 蓝光光盘的存储容量则可以达到 50 GB。在视频监控领域, 常见的视频存储设备有硬盘录像机 (Digital Video Recorder, DVR)、网络视频录像机 (Network Video Recorder, NVR) 以及存储区域网络 (Storage Area Network, SAN)。其中, DVR 集编码、存储和摄像头云台控制等功能于一身, 适用于小型监控系统; NVR 主要用于控制网络摄像机 (IP Camera, IPC) 并负责存储 IPC 输出的数字视频码流, 适用于远距离监控; 存储区域网络 SAN 是通过专用高速网络将多个存储设备和服务器连接起来的专用存储系统, 可分为 FC-SAN 和 IP-SAN, 前者主要采用光纤互联, 后者采用 IP 协议互联。SAN 具有很强的存储扩展能力和故障隔离能力, 如图 7(a) 所示, 适用于大规模监控网络的视频数据集中存储。

目前, 视频点播网站普遍采用了云存储系统来保存海量的视频数据。常用的云存储系统有开源的 Hadoop 分布式系统内置的 HDFS 文件系统, 如图 7(b) 所示, 该存储系统会把一个完整的视频文件按固定大小分割为若干块, 然后将各个文件块保存在不同主机的硬盘上, 且提供数据块多机备份机制, 提高了数据安全性。云存储系

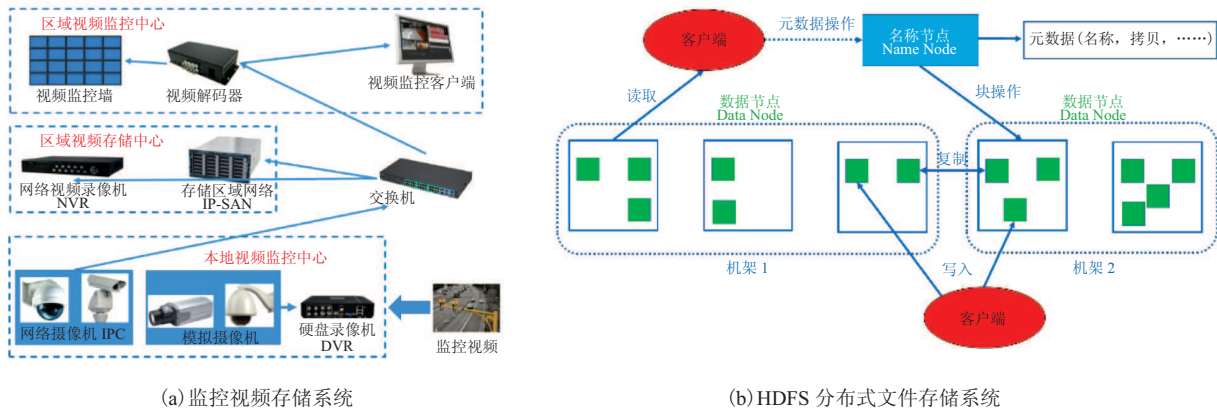


图7 视频存储系统

Fig. 7 Video storage system

系统能将异构存储设备组建成一个巨大的虚拟存储池，按需动态扩张存储容量，满足海量视频存储需求，提供 TB 级的视频输出带宽，而且由于云存储和云计算的紧密结合，可在视频大数据之上构建各种服务与应用。

## 5 视频显示与用户体验质量

### 5.1 视频显示技术的发展

视频的显示技术伴随着视频数据的表示、压缩和传输技术的进步而不断发展。4K 分辨率的显示器逐渐得到普及，而投影显示则会将画面的尺寸放大到数百寸。具有更高亮度和对比度的显示技术可以让视频图像更丰富和逼真的视觉感受。3D 显示器和全息投影技术会把观众从 2D 平面的束缚中解放出来，从而获得更多场景和物体的三维信息。目前视频显示技术的发展趋势包括：

#### (1) 超高清显示

根据 ITU 建议书，1 920×1 080 为全高清 Full HD，3 840×2 160 为超高清 UHD。UHD 又被称为 4K 分辨率，此外 ITU-R BT.2020 还定义了 8K (7 680×4 320) 分辨率标准。目前支持 4K 分辨率 60 Hz 帧率的超高清蓝光标准 (UHD Blu-ray) 和 HDMI 2.0a 标准将逐渐得到推广。

#### (2) 高动态与广色域显示

人眼能感知的光线亮度范围通常在 0~10 000 nit，但目前的视频显示设备通常只具备 0~500 nit 的亮度显示范围。同时，旧电视的标准色域范围只能覆盖标准视觉色域的 33.5%。这些都使观众无法观看到接近人眼真实感受的视频。目前针对高动态范围 HDR 与广色域 WCG，国际标准组织与各大公司都在制定相关技术规范。要实现高动态 HDR 显示，需要显示器具备 1 000 nit 以上的显示亮度。激光光源投影利用激光波长可选择性和高光谱亮度的特点，可以使显示色域覆盖标准视觉色域的 90%，在广色域显示方面具有优势。

#### (3) 3D 显示

3D 显示相比于 2D 平面显示增加了深度感，可以给观看者带来更加真实的临场感。目前 3D 显示技术大致分为眼镜式 3D、裸眼 3D 和全息 3D 等<sup>[68]</sup>。眼镜式 3D 分为色差式、偏光式以及快门式。其中，色差式眼镜通过不同颜色的滤光片使双眼分别看到不同颜色的左右图像，该方式会存在图像颜色失真；偏光式 3D 采用不同偏振方向的偏光镜片来分离左右偏振光图像，偏光式 3D 显示器的缺点是图像亮度会减半；快门式 3D 则通过 120 Hz 以上刷新率的显示器配合液晶镜

片的闭合与打开来快速切换左右视点图像, 该方式的好处是可以看到完整分辨率的图像, 但缺点是亮度较暗且存在轻微闪烁。裸眼 3D 显示技术目前根据显示原理主要以光栅式和透镜式为主。其中, 光栅式 3D 通过控制屏幕表层的液晶层形成垂直排列的栅状条纹来将左右眼看到的画面分开; 透镜式 3D 通过屏幕上的一层柱状透镜使每个柱状透镜下面的左右图像的像素分别投向左右眼。多视点裸眼 3D 显示器可以将多个视点的图像的像素分别投射到多个方向, 从而拓展了裸眼 3D 显示的观看区域和角度。裸眼 3D 显示器的缺点是损失了图像分辨率, 并且视点越多则分辨率下降得越严重; 此外, 各视区之间存在视点适配导致的跳变, 影响视点切换过程中的立体感知。全息 3D 投影技术利用光的干涉和衍射原理可以还原出全息摄影记录下的三维物体信息, 该方式可以得到更加真实的 3D 感。但由于全息投影需要在拍摄时记录下拍摄对象完整的三维信息, 数

据量巨大, 所以目前只能应用于一些特定场合, 如 3D 文物展示。

## 5.2 视频用户体验质量评价

各种视频压缩、通讯与处理技术必不可少, 而这些处理势必造成视频图像质量的失真, 所以需要视频图像的质量进行准确的评价, 由此, 能够符合人眼视觉感知特性的视频质量评价准则对指导视频的压缩与处理算法的优化具有重要意义。

视频质量评价分为主观质量评价和客观质量评价(如图 8)。视频的最终接收端是人眼, 所以由人来进行主观质量评价是最准确的。但是主观实验费时费力, 无法应用于实时的视频处理系统, 所以通过算法进行自动的客观视频质量评价很有必要。主观质量评价实验的作用则变成为客观评价算法提供训练数据与测试数据。

### (1) 主观视频质量评价

国际电信联盟的无线电通信部门 ITU-R 制订了电视图像质量的主观评价方法建议书 ITU-R

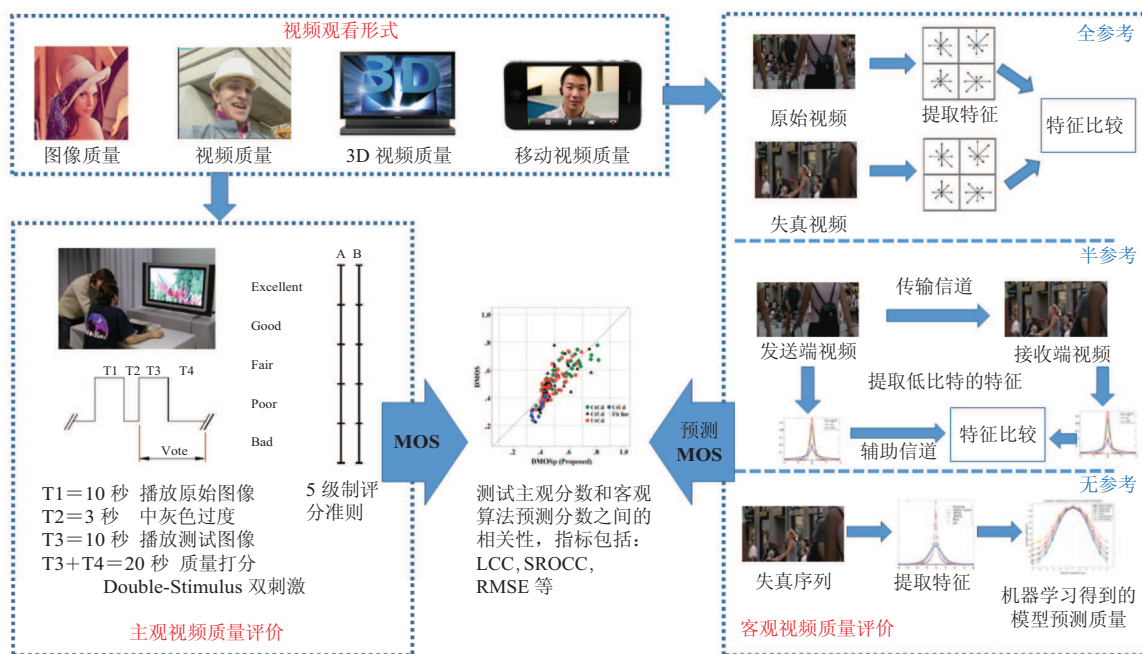


图 8 视频图像主客观质量评价

Fig. 8 Subjective and objective image/video quality assessment



BT.500<sup>[69]</sup>，而同属于 ITU 的电信标准化部门 ITU-T 也制订了面向多媒体应用的主观视频质量评价方法建议书 ITU-T P.910<sup>[70]</sup>。各种主观视频质量评价方法的核心区别在于使用单刺激还是双刺激。双刺激是指观看者在给每一个视频打分之前都需要观看一遍原始参考视频，而单刺激是指每次观看一个单独的视频并对其质量进行打分。

ITU-R BT.500 和 ITU-T P.910 对主观视频质量评价的实验环境、视频源、测试者以及整个质量评价流程都进行了详细的规定。

### (2) 客观视频质量评价

客观视频质量评价方法可以按照是否需要参考原始视频分为全参考、半参考和无参考三种<sup>[71]</sup>。全参考的视频质量评价算法需要输入原始视频和处理过后的失真视频<sup>[72-75]</sup>。由于很多情况下在对视频质量进行评价时其原始图像是不可获得的，此时则需要使用无参考视频质量评价方法。在一些视频传输系统中，为了在接收端对视频进行质量评价，需要在发送端将原始视频的一些图像统计特征一起发送到接收端用于评价所接收视频的失真，该类方法属于半参考视频质量。无论使用哪种方法，最终目的都是使评价结果尽可能接近人眼对视频质量的真实感受。

### (3) 从图像质量到用户体验质量

随着视频数据、显示设备以及观看方式的越来越复杂多样，视频质量评价的趋势从单纯的图像质量评价发展为更加综合的用户体验质量 (Quality of Experience, QoE) 评价。早期人们观看视频主要通过电视，随着技术的发展，视频类型以及观看视频的方式越来越多样化。网络点播视频、移动端视频播放、3D 立体视频等新的视频观看形式使得影响用户体验质量 QoE 的因素不再是单一的图像质量<sup>[76,77]</sup>。ITU-R BT.2021 针对 3D 立体视频的 QoE 评价，建议在原有的图像质量评价基础之上增加深度感质量与视觉舒适度的评价<sup>[78]</sup>。移动端播放视频时的网络延迟与

丢包、设备因素、环境因素等也会影响用户的 QoE<sup>[79]</sup>。所以，在对用户体验质量 QoE 进行评价时，有必要从多个维度并结合大数据相关技术综合考虑。

## 6 总结与展望

本文从视频数据表示、智能视频分析、视频压缩与存储、视频显示与质量评价等方面对多媒体视频大数据研究进行了综合介绍。视频大数据不仅拓宽了人类的视野，使我们能够更好地感受和认知身边的世界，还埋藏着大量有价值的信息等待人们去挖掘利用。可以预见，未来视频大数据的研究将会有以下趋势：

(1) 增强现实与虚拟现实的视频显示技术。视频的表现形式不仅是向着越来越高的分辨率和帧率发展，也不仅是提高图像的亮度和色域范围。虚拟现实与增强现实的显示技术会带来更加身临其境的显示效果，如 Oculus 的虚拟现实眼镜和 Microsoft 的增强现实全息 3D 眼镜 HoloLens。这类显示技术在带给人们更加震撼的视觉效果的同时，也会改变人与智能设备之间的交互方式。虚拟的 3D 影像将与真实环境叠加在一起，并且人们可以通过眼神、声音、手势与虚拟影像进行交互。目前，该类设备距离一般消费者还有一段距离，并且还需要挖掘开发更多的应用场景。

(2) 基于感知的高效编码技术。经历三代的视频压缩技术的发展，时空冗余、信息冗余的挖掘已接近极限，通过增加计算量的方式换取编码增益代价大，且提升空间小。由于人们观看视频过程中，人眼存在着较大的感知冗余，如何有效地将视频编码与视觉感知融合，进一步挖掘人眼视觉冗余，提高压缩效率并降低计算复杂度是未来视频压缩领域重要研究方向。

(3) 视频编码与视频分析技术的结合。目前



视频压缩编码与智能视频分析还是两个相对独立的过程,但实际上这两个过程都需要对视频信号的数据特性进行分析统计。视频分析所使用的数据源都是压缩编码之后的视频,如果能够在视频的压缩编码阶段就对视频的内容进行一些描述编码,无疑将大大提高视频分析与检索的效率。具体做法可以是在视频数据的预处理阶段增加图像分割和目标检测等分析过程,在视频压缩编码阶段增加对背景模型、目标对象的描述编码。

(4) 基于深度神经网络的智能视频分析。虽然目前针对智能视频分析中的目标检测、跟踪、识别和分类等问题,研究人员们已经提出了各种各样的方法,但是还没有统一的理论框架或体系,实际应用中还存在很多难点亟待突破,如光照变化、阴影、复杂背景、遮挡、相机抖动以及物体的非刚体形变等都会影响目标检测和跟踪,以及后续的识别准确度。非重叠场景下的目标跟踪问题也是一个新的挑战。值得注意的是,基于深度神经网络的机器学习技术已经在图像视频分析方面取得了较好的效果。目前的深度学习主要还是依赖有标记的数据,但在未来,深度学习的趋势将会是同时依赖未标记数据和标记数据,直到最终实现基于较少标记数据,真正实现模拟人和动物的认知世界的过程。

(5) 更全面的用户体验质量评价。目前视频主观质量评价主要依据的 ITU-R BT.500 标准还是针对标清(704×480)和高清(1 920×1 080) CRT 显示设备所制定的。随着技术的进步,视频显示设备已经从传统的高清标清电视机发展到 4K 超高清电视、智能手机、投影仪、3D 显示器甚至虚拟现实眼镜等设备,信号的传输也从以往的模拟信号转变为全数字网络传输。这就要求视频质量评价的标准也需要与时俱进,不仅要考虑新的视频表现形式,还需要从单一的图像质量评价向综合全面的用户体验质量评价发展。具体包括从网络因素、设备因素、环境因素、图像质量

因素和个人因素等多个方面全面综合地考察视频服务质量。

总的来说,目前全球的大数据研究尚处于发展初期,而视频大数据研究更是刚刚起步。可以预见,视频大数据研究会在人们的生活、生产等诸多领域都会产生深远影响并发挥重要作用。

## 参 考 文 献

- [1] 中国互联网络信息中心. 第 35 次中国互联网络发展状况统计报告 [R/OL]. 2015-02-03 [2015-12-20]. [http://www.cnnic.net.cn/hlwfzyj/hlwxzbg/hlwtjbg/201502/t20150203\\_51634.htm](http://www.cnnic.net.cn/hlwfzyj/hlwxzbg/hlwtjbg/201502/t20150203_51634.htm).
- [2] Cisco visual networking index: global mobile data traffic forecast update 2014-2019 White Paper [R/OL]. 2015-02-03 [2015-12-20]. [http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white\\_paper\\_c11-520862.html](http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white_paper_c11-520862.html).
- [3] ITU-R Recommendation BT.2020. Parameter values for ultra-high definition television systems for production and international programme exchange [S]. <https://www.itu.int/rec/R-REC-BT.2020/en>.
- [4] Muller K, Merkle P, Wiegand T. 3-D video representation using depth maps [J]. *Proceedings of the IEEE*, 2011, 99(4): 643-656.
- [5] Zhang XG, Huang TJ, Tian YH, et al. Background-modeling-based adaptive prediction for surveillance video coding [J]. *IEEE Transactions on Image Processing*, 2014, 23(2): 769-784.
- [6] Zhang XG, Tian YH, Huang TJ, et al. Optimization the hierarchical prediction and coding in HEVC for surveillance and conference videos with background modeling [J]. *IEEE Transactions on Image Processing*, 2014, 23(10): 4511-4526.
- [7] Zhu WJ, Ding WP, Xu JZ, et al. Screen content coding based on HEVC framework [J]. *IEEE Transactions on Multimedia*, 2014, 16(5): 1316-1326.
- [8] Ma Z, Wang W, Xu M, et al. Advanced screen content coding using color table and index map [J]. *IEEE Transactions on Image Processing*, 2014,

- 23(10): 4399-4412.
- [9] Ren XF, Malik J. Learning a classification model for segmentation [C] // Proceedings of the Ninth IEEE International Conference on Computer Vision, 2003: 10-17.
- [10] 王春瑶, 陈俊周, 李炜. 超像素分割算法研究综述 [J]. 计算机应用研究, 2014, 31(1): 6-12.
- [11] Shi J, Malik J. Normalized cuts and image segmentation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(8): 888-905.
- [12] Felzenszwalb PF, Huttenlocher DP. Efficient graph-based image segmentation [J]. International Journal of Computer Vision, 2004, 59(2): 167-181.
- [13] Moore AP, Prince S, Warrell J, et al. Superpixel lattices [C] // IEEE Conference on Computer Vision and Pattern Recognition, 2008: 23-28.
- [14] Liu MY, Tuzel O, Ramalingam S, et al. Entropy rate superpixel segmentation [C] // IEEE Conference on Computer Vision and Pattern Recognition, 2011: 2097-2104.
- [15] Vincent L, Soille P. Watersheds in digital spaces: an efficient algorithm based on immersion simulations [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1991, 13(6): 583-598.
- [16] Comaniciu D, Meer P. Mean shift: a robust approach toward feature space analysis [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24(5): 603-619.
- [17] Levinshstein A, Stere A, Kutulakos KN, et al. Turbopixels: fast superpixels using geometric flows [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 31(12): 2290-2297.
- [18] Xu CL, Corso JJ. Evaluation of super-voxel methods for early video processing [C] // IEEE Conference on Computer Vision and Pattern Recognition, 2012: 1202-1209.
- [19] Yang F, Lu H, Yang MH. Robust superpixel tracking [J]. IEEE Transactions on Image Processing, 2014, 23(4): 1639-1651.
- [20] Field DJ. What is the goal of sensory coding? [J]. Neural Computation, 1994, 6(4): 559-601.
- [21] Olshausen BA, Field DJ. Emergence of simple-cell receptive field properties by learning a sparse code for natural images [J]. Nature, 1996, 381(6583): 607-609.
- [22] Elad M, Aharon M. Image denoising via sparse and redundant representations over learned dictionaries [J]. IEEE Transactions on Image Processing, 2006, 15(12): 3736-3745.
- [23] Yang JC, Wright J, Huang TS, et al. Image super-resolution via sparse representation [J]. IEEE Transactions on Image Processing, 2010, 19(11): 2861-2873.
- [24] Wright J, Ma Y, Mairal J, et al. Sparse representation for computer vision and pattern recognition [J]. Proceedings of the IEEE, 2010, 98(6): 1031-1044.
- [25] Liu LQ, Zhang L, Liu HR, et al. Toward large-population face identification in unconstrained videos [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2014, 24(11): 1874-1884.
- [26] Hu WM, Li W, Zhang XQ, et al. Single and multiple object tracking using a multi-feature joint sparse representation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(4): 816-833.
- [27] Andriluka M, Roth S, Schiele B. People-tracking-by-detection and people-detection-by-tracking [C] // IEEE Conference on Computer Vision and Pattern Recognition, 2008: 1-8.
- [28] Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks [C] // Advances in Neural Information Processing Systems, 2012: 1097-1105.
- [29] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C] // IEEE Conference on Computer Vision and Pattern Recognition, 2014: 580-587.
- [30] Lucas BD, Kanade T. An iterative image registration technique with an application to stereo vision [C] // Proceedings of the 7th International Joint Conference on Artificial Intelligence, 1981: 674-679.
- [31] Kalal Z, Mikolajczyk K, Matas J. Tracking-learning-detection [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 34(7): 1409-1422.
- [32] Simonyan K, Zisserman A. Two-stream convolutional networks for action recognition in videos [C] // Advances in Neural Information Processing Systems, 2014: 568-576.
- [33] Ng JYH, Hausknecht M, Vijayanarasimhan S, et al. Beyond short snippets: deep networks for video

- classification [J]. arXiv preprint arXiv:1503.08909, 2015.
- [34] Hochreiter S, Schmidhuber J. Long short-term memory [J]. *Neural Computation*, 1997, 9(8): 1735-1780.
- [35] Donahue J, Hendricks LA, Guadarrama S, et al. Long-term recurrent convolutional networks for visual recognition and description [J]. arXiv preprint arXiv:1411.4389, 2014.
- [36] Wu ZX, Wang X, Jiang YG, et al. Modeling spatial-temporal clues in a hybrid deep learning framework for video classification [J]. arXiv preprint arXiv:1504.01561, 2015.
- [37] Kalva H. The H.264 video coding standard [J]. *IEEE Multimedia*, 2006, 13(4): 86-90.
- [38] Schwarz H, Marpe D, Wiegand T. Overview of the scalable video coding extension of the H.264/AVC standard [J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2007, 17(9): 1103-1120.
- [39] Vetro A, Wiegand T, Sullivan GJ. Overview of the stereo and multiview video coding extensions of the H.264/MPEG-4 AVC standard [J]. *Proceedings of the IEEE*, 2011, 99(4): 626-642.
- [40] Sullivan GJ, Ohm J, Han WJ, et al. Overview of the high efficiency video coding (HEVC) standard [J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2012, 22(12): 1649-1668.
- [41] Boyce JM, Ye Y, Chen J, et al. Overview of SHVC: scalable extensions of the high efficiency video coding (HEVC) standard [J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2015, 26(1): 20-34.
- [42] 朱秀昌, 王国刚, 陈杰, 等. HEVC 标准的多层视频编码扩展 [J]. *南京邮电大学学报(自然科学版)*, 2015, 35(3): 1-10.
- [43] 马思伟. AVS 视频编码标准技术回顾及最新进展 [J]. *计算机研究与发展*, 2015, 52(1): 27-37.
- [44] Ma SW, Huang TJ, Reader C, et al. AVS2-Making video coding smarter [Standards in a nutshell] [J]. *IEEE Signal Processing Magazine*, 2015, 32(2): 172-183.
- [45] Ugur K, Alshin A, Alshina E, et al. Motion compensated prediction and interpolation filter design in H.265/HEVC [J]. *IEEE Journal of Selected Topics in Signal Processing*, 2013, 7(6): 946-956.
- [46] Ugur K, Andersson K, Fuldseth A, et al. High performance, low complexity video coding and the emerging HEVC standard [J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2010, 20(12): 1688-1697.
- [47] Seo CW, Moon JH, Han JK. Rate control for consistent objective quality in high efficiency video coding [J]. *IEEE Transactions on Image Processing*, 2013, 22(6): 2442-2454.
- [48] 蒋刚毅, 朱亚培, 郁梅, 等. 基于感知的视频编码方法综述 [J]. *电子与信息学报*, 2013, 35(2): 474-483.
- [49] Luo Z, Song L, Zheng S, et al. H.264/Advanced video control perceptual optimization coding based on JND-directed coefficient suppression [J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2013, 23(6): 935-948.
- [50] Ekmekcioglu E, De Silva V, Pesch PT, et al. Visual attention model aided non-uniform asymmetric coding of stereoscopic video [J]. *IEEE Journal of Selected Topics in Signal Processing*, 2014, 8(3): 402-414.
- [51] Zhang Y, Kwong S, Hu S, et al. Efficient multiview depth coding optimization based on allowable depth distortion in view synthesis [J]. *IEEE Transactions on Image Processing*, 2014, 23(11): 4879-4892.
- [52] Zhang Y, Kwong S, Xu L, et al. Regional bit allocation and rate distortion optimization for multiview depth video coding with view synthesis distortion model [J]. *IEEE Transactions on Image Processing*, 2013, 22(9): 3497-3512.
- [53] Zhang Y, Jiang GY, Yu M, et al. Depth perceptual region-of-interest based multiview video coding [J]. *Journal of Visual Communication and Image Representation*, 2010, 21(5-6): 498-512.
- [54] Zhang Y, Jiang GY, Yu M, et al. Stereoscopic visual attention based bit allocation optimization for multiview video coding [J]. *EURASIP Journal on Advances in Signal Processing*, 2010, 2010: 60.
- [55] Xu L, Kwong S, Zhang Y, et al. Low-complexity encoder framework for window-level rate control optimization [J]. *IEEE Transactions on Industrial Electronics*, 2012, 60(5): 1850-1858.
- [56] Hu SD, Kwong S, Zhang Y, et al. Rate-distortion optimized rate control for depth map based 3-D video coding [J]. *IEEE Transactions on Image Processing*, 2013, 22(2): 585-594.
- [57] Xu L, Kwong S, Wang HL, et al. A universal rate control scheme for video transcoding [J]. *IEEE*

- Transactions on Circuits and System for Video Technology, 2011, 22(4): 489-501.
- [58] Wang X, Kwong S, Zhang Y. Applying game theory to rate control optimization for hierarchical B-pictures [J]. IEEE Transactions on Broadcasting, 2013, 59(4): 591-601.
- [59] Wang X, Kwong S, Xu L, et al. Generalized Nash bargaining solution to rate control optimization for spatial scalable video coding [J]. IEEE Transactions on Image Processing, 2014, 23(9): 4010-4021.
- [60] Yuan H, Kwong S, Ge C, et al. Interview rate distortion analysis based coarse to fine bit allocation algorithm for 3D video coding [J]. IEEE Transactions on Broadcasting, 2014, 60(4): 614-625.
- [61] Zhu LW, Zhang Y, Yu M, et al. View-spatial-temporal post-refinement for view synthesis in 3D video systems [J]. Signal Processing: Image Communication, 2013, 28(10): 1342-1357.
- [62] Zhu L, Zhang Y, Wang X, et al. View synthesis distortion elimination filter for depth video coding in 3D video broadcasting [J]. Multimedia Tools and Applications, 2015, 74(15): 5935-5954.
- [63] Pan ZQ, Zhang Y, Kwong S. Efficient motion and disparity estimation optimization for low complexity multiview video coding [J]. IEEE Transactions on Broadcasting, 2015, 61(2): 166-176.
- [64] Zhang Y, Kwong S, Jiang GY, et al. Efficient multi-reference frame selection algorithm for hierarchical B pictures in multiview video coding [J]. IEEE Transactions on Broadcasting, 2011, 57(1): 15-23.
- [65] 张冠军, 朱林卫, 张云. 基于模式预测的低复杂度高清视频帧内编码方法 [J]. 集成技术, 2014, 3(2): 42-52.
- [66] Zhang Y, Kwong S, Wang X, et al. Machine learning-based coding unit depth decisions for flexible complexity allocation in high efficiency video coding [J]. IEEE Transactions on Image Processing, 2015, 24(7): 2225-2238.
- [67] Zhang Y, Kwong S, Jiang GY, et al. Statistical early termination model for fast mode decision and reference frame selection in multiview video coding [J]. IEEE Transactions on Broadcasting, 2012, 58(1): 10-23.
- [68] Smolic A, Kauff P, Knorr S, et al. Three-dimensional video postproduction and processing [J]. Proceedings of the IEEE, 2011, 99(4): 607-625.
- [69] ITU-R Recommendation. BT.500-13. Methodology for the subjective assessment of the quality of television pictures [S]. <http://www.itu.int/rec/R-REC-BT.500-13-201201-I>.
- [70] ITU-T Recommendation P.910. Subjective video quality assessment methods for multimedia applications [S]. <https://www.itu.int/rec/T-REC-P.910/en>.
- [71] Chikkerur S, Sundaram V, Reisslein M, et al. Objective video quality assessment methods: a classification, review, and performance comparison [J]. IEEE Transactions on Broadcasting, 2011, 57(2): 165-182.
- [72] Hu SD, Jin L, Wang HL, et al. Compressed image quality metric based on perceptually weighted distortion [J]. IEEE Transactions on Image Processing, 2015, 24(12): 5594-5608.
- [73] Xu HY, Jiang GY, Yu M, et al. Binocular combination and fractional differential based 3D image quality assessment [C] // 2015 7th International Workshop on Quality of Multimedia Experience, 2015: 1-5.
- [74] Wang X, Jiang GY, Zhou JM, et al. Visibility threshold of compressed stereoscopic image: effects of asymmetrical coding [J]. The Imaging Science Journal, 2013, 61(2): 172-182.
- [75] Jiang G, Zhou J, Yu M, et al. Binocular fusion and suppression based objective quality assessment method for stereoscopic images [J]. Multimedia Tools and Applications, 2015, 74(18): 8197-8218.
- [76] Chen YJ, Wu KS, Zhang Q. From QoS to QoE: a tutorial on video quality assessment [J]. IEEE Communications Surveys and Tutorials, 2015, 17(2): 1126-1165.
- [77] Liu XK, Zhang Y, Hu SD, et al. Subjective and objective video quality assessment of 3D synthesized view with texture/depth compression distortion [J]. IEEE Transactions on Image Processing, 2015, 24(12): 4847-4861.
- [78] ITU-R Recommendation BT.2021-1. BT.2021: Subjective methods assessment stereoscopic 3DTV systems [S]. <http://www.itu.int/rec/R-REC-BT.2021-1-201502-I>.
- [79] Xue JT, Chen CW. Mobile video perception: new insights and adaptation strategies [J]. IEEE Journal of Selected Topics in Signal Processing, 2014, 8(3): 390-401.