





多媒体通信

Multimedia Communications

第2章 多媒体数据压缩国际标准

 2019年9月15日 



MIT 《技术评论》 每年初评选当年度十大突破性技术

◆ MIT Technology Review

- 美国专业的科技商业杂志，创刊于1899年，由麻省理工学院（MIT）主办，以推介、评析科技创新为宗旨，展示最新科技成果及其商业潜能，内容涉及互联网、通讯、计算机技术、能源、新材料、生物医学和商务科技等领域。



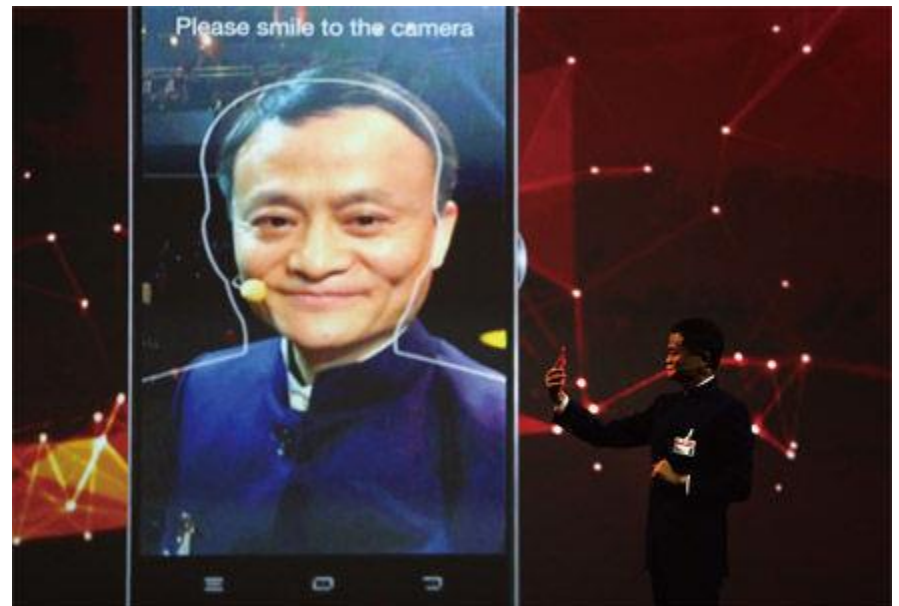
10 BREAKTHROUGH TECHNOLOGIES



《麻省理工科技评论》2017年十大突破性技术 刷脸支付（Paying with Your Face）

指纹支付基本目前已是各大智能手机配备的技术。但指纹支付在某些场景下，比如说手上沾水等情况就没法好好地玩耍了。在人脸识别技术日益精确地情况下，人脸成为网络交易等相关领域盯上的下一个“指纹”。支付宝此前就推出了刷脸支付的功能。

主要公司：阿里巴巴、
百度、科大讯飞、旷视
Face++
成熟期：现在





《麻省理工科技评论》2017年十大突破性技术 自动驾驶货车（Self-Driving Trucks）

自动驾驶货车的优势很明显，除了能够出色高效完成运输任务外，有研究表明自动驾驶货车可降低 15 % 排放量，还能减少交通事故，避免交通拥堵。另外，自动驾驶货车行驶 10 万英里（16.1 万公里）可以节约 5000 英镑（约合人民币 43000 元）。

主要公司：

Otto、沃尔沃（Volvo）、
戴姆勒（Daimler AG）、
皮特比尔特（Peterbilt）
、百度（Baidu）

成熟期：5—10年





《麻省理工科技评论》2017年十大突破性技术 360 度自拍（ The 360-Degree Selfie ）

不久的将来，人们除了以现有的图片、文字和小视频进行交互外，更增添了新的沟通表达方式，比如 360 度全景图片、360 度全景视频。

主要研究者：日本理光（Ricoh）、360fly、三星、JK Imaging（柯达 Pixpro 相机的制造厂商）、IC Real Tech（ALLie 相机的制造厂商）、Humaneyes Technologies（全景相机 Vuze 的制造厂商）
成熟期：现在





《麻省理工科技评论》2017年十大突破性技术 僵尸物联网 (Botnets of Things)

黑客们的攻击范围正在变得愈来愈广。随着物联网时代的到来，其攻击的对象不再局限于过去的电脑、手机等设备，凡是能联网的设备，都可成为其潜在攻击源。

去年十月，Twitter、Paypal、Spotify 等多个美国网站被迫中断服务，其背后的罪魁祸首是一种被称作“物联网破坏者”的Mirai病毒。它就是通过互联网搜索物联网设备来传播病毒，当它扫描到一个物联网设备，比如网络摄像头、智能开关等后，就会尝试使用默认密码进行登陆，该病毒自带60个通用密码。一旦登录成功，这台物联网设备就进入“肉鸡”名单，开始被黑客操控攻击其他网络设备，这次一共有超过百万台参与了DDoS攻击。

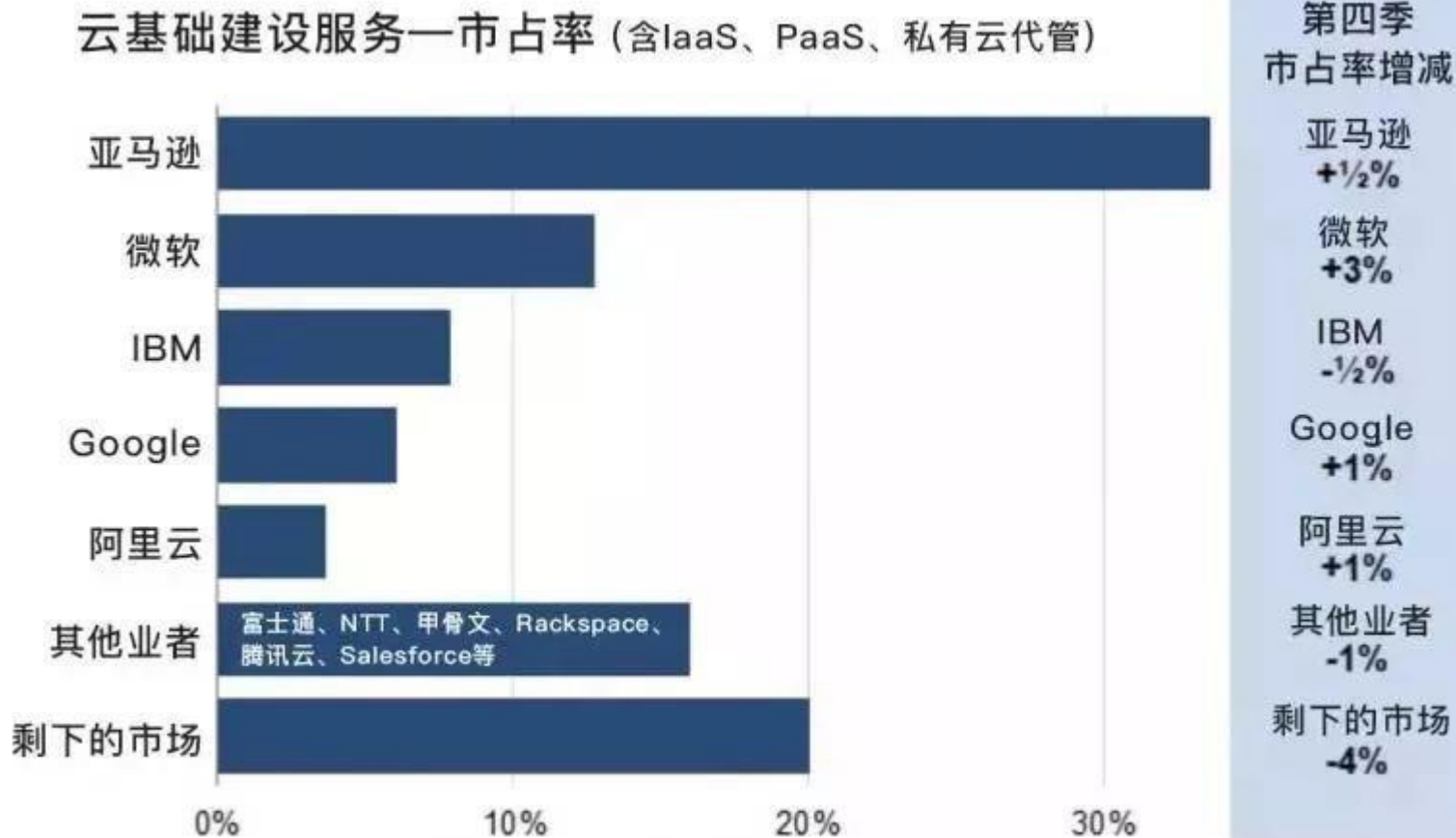
而这种病毒能轻易获得、感染设备后会进行大量自动化设置，危害极大。另外，伴着万物互联的设备逐渐增多，病毒的扩散阻止起来会愈发困难。





《麻省理工科技评论》2018年十大突破性技术 给所有人的人工智能（AI for Everybody）

突破性：基于云的人工智能技术（Cloud-based AI）使得这项技术的使用更便宜、更容易。重要性：目前，AI的使用由少数几家公司主导，但作为一种基于云的服务，它能够被更多的人使用，从而推动经济增长。关键研究者：亚马逊、谷歌、微软。





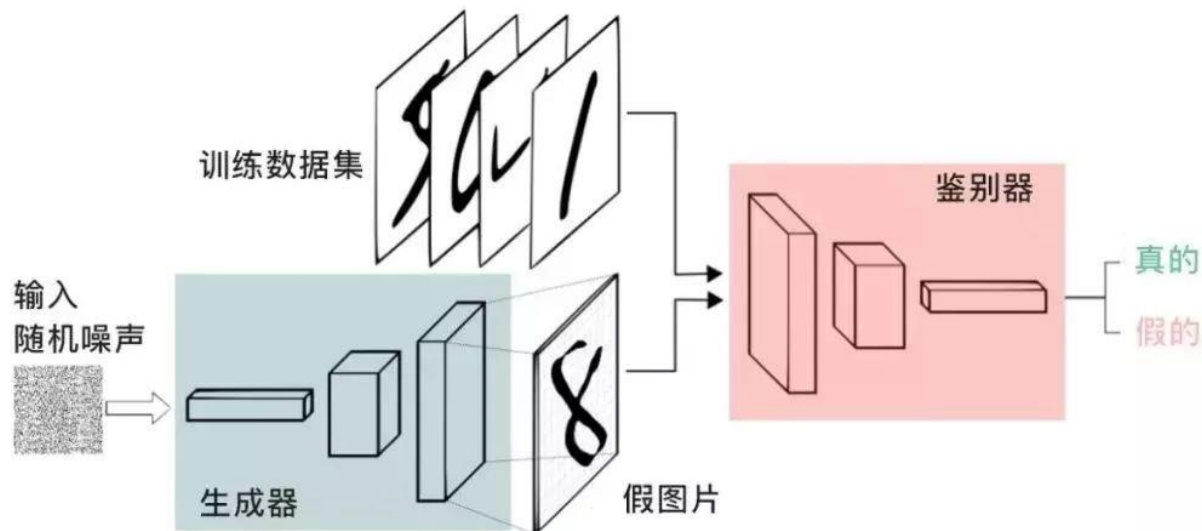
《麻省理工科技评论》2018年十大突破性技术 对抗性神经网络（Dueling Neural Networks）

技术突破：两个 AI 系统可以通过相互对抗来创造超级真实的原创图像或声音，而在此之前，机器从未有这种能力

重大意义：这给机器带来一种类似想象力的能力，因此可能让它们变得不再那么依赖人类，但也把它们变成了一种能力惊人的数字造假工具

忧虑：用猫的图片来训练，无法生成狗的图片。对参数非常敏感。

主要研究者：Google Brain、DeepMind、英伟达、中科院自动化所、百度、阿里巴巴、腾讯、商汤科技、依图科技、云从科技、旷世科技等。





《麻省理工科技评论》2018年十大突破性技术 传感城市（Sensing City）

入选理由：Alphabet 旗下的 Sidewalk Labs 计划创建一个高科技社区来重新思考到底应该如何建设和运营一座城市。传感城市更宜居、环保。收集各种信息，软件和系统开源。给城市装上全新的数字肢体和感官。

技术突破：多伦多的一个街区有望成为全球首个成功将尖端城市设计与前沿科技融合在一起的地方。

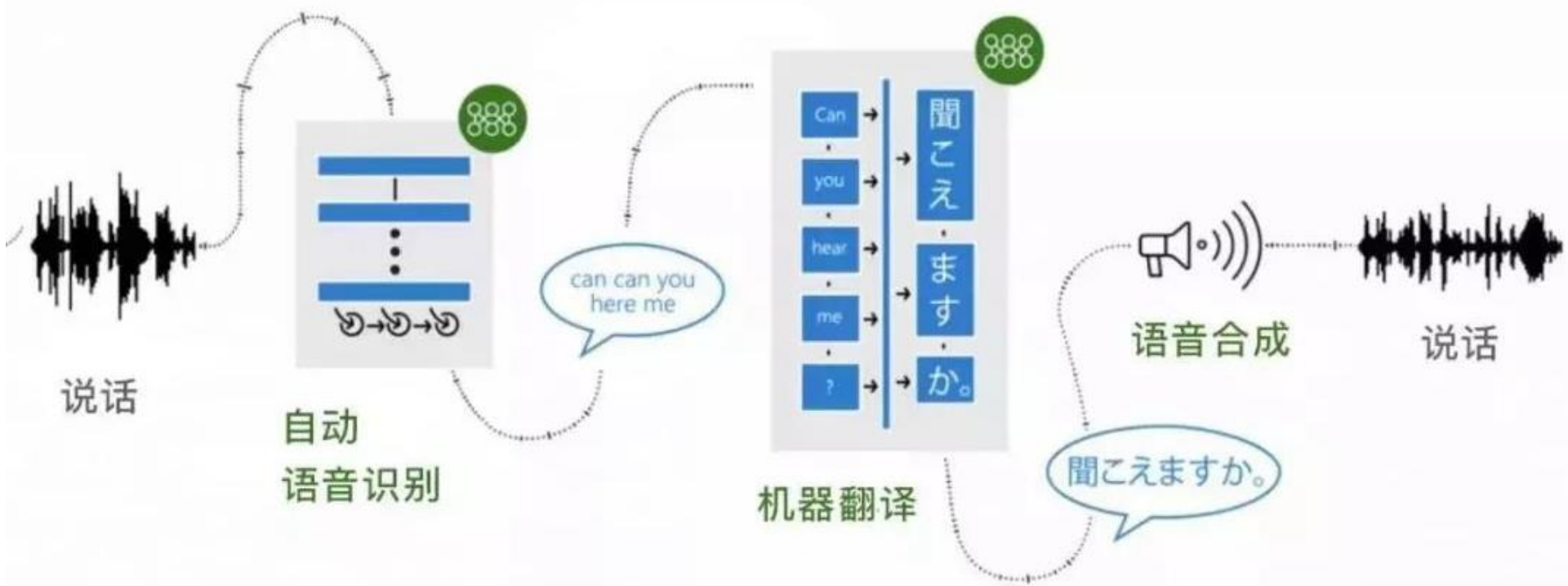


主要研究者：
Alphabet 旗下的
Sidewalk Labs、多
伦多 Waterfront、
阿里巴巴等。



《麻省理工科技评论》2018年十大突破性技术 即时翻译耳塞（Babel-Fish Earbuds）

在风靡一时的科幻经典《银河系漫游指南》中，你把一条黄色的巴别鱼塞到耳朵里，就可以听到实时翻译。在现实世界中，谷歌已经研究出了一个过渡性的解决方案：一副叫做 Pixel Buds 价值 159 美元的耳塞。这副耳塞可以在 Pixel 智能手机上通过谷歌翻译应用进行实时翻译。需要一个人佩戴耳塞，另一个人手持手机。佩戴耳塞的人用自己的语言讲话——默认是英语——然后谷歌翻译应用就会对所讲的话进行翻译，并在智能手机上大声播放。手持手机的人回应后，回答被翻译，然后在耳塞中播放。



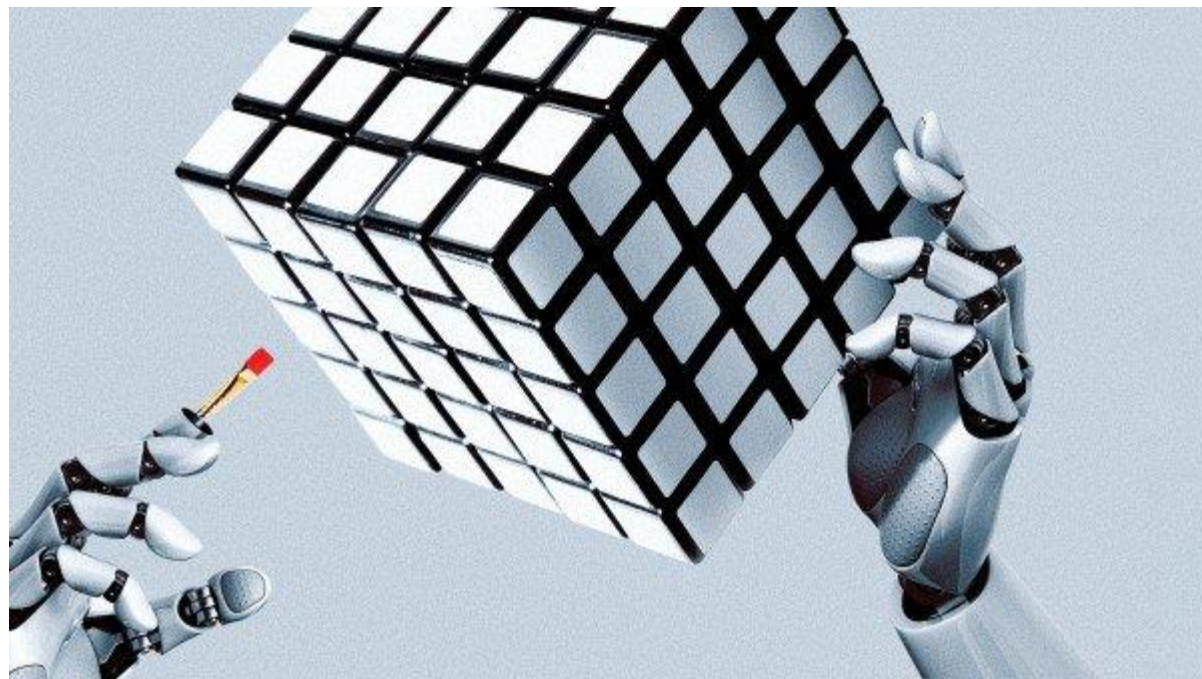


《麻省理工科技评论》2019年十大突破性技术 流利对话的 AI 助手

重大意义： 机器正在通过自我学习学会应对这个现实世界。如果机器人能学会应对混乱的现实世界，那么它们就可以胜任更多的任务。

主要研究者： OpenAI（人工智能非营利组织）、卡内基梅隆大学、密歇根大学、加州大学伯克利分校

成熟期： 3-5年





《麻省理工科技评论》2019年十大突破性技术 流利对话的 AI 助手

重大意义：随着监管机构的批准和相关技术的进步，人们可以轻松通过可穿戴设备持续监测自己的心脏健康。可检测心电图的智能手表可以预警如心房颤动等潜在的危及生命的心脏疾病。

主要研究者：苹果、AliveCor、Withings

成熟期：现在



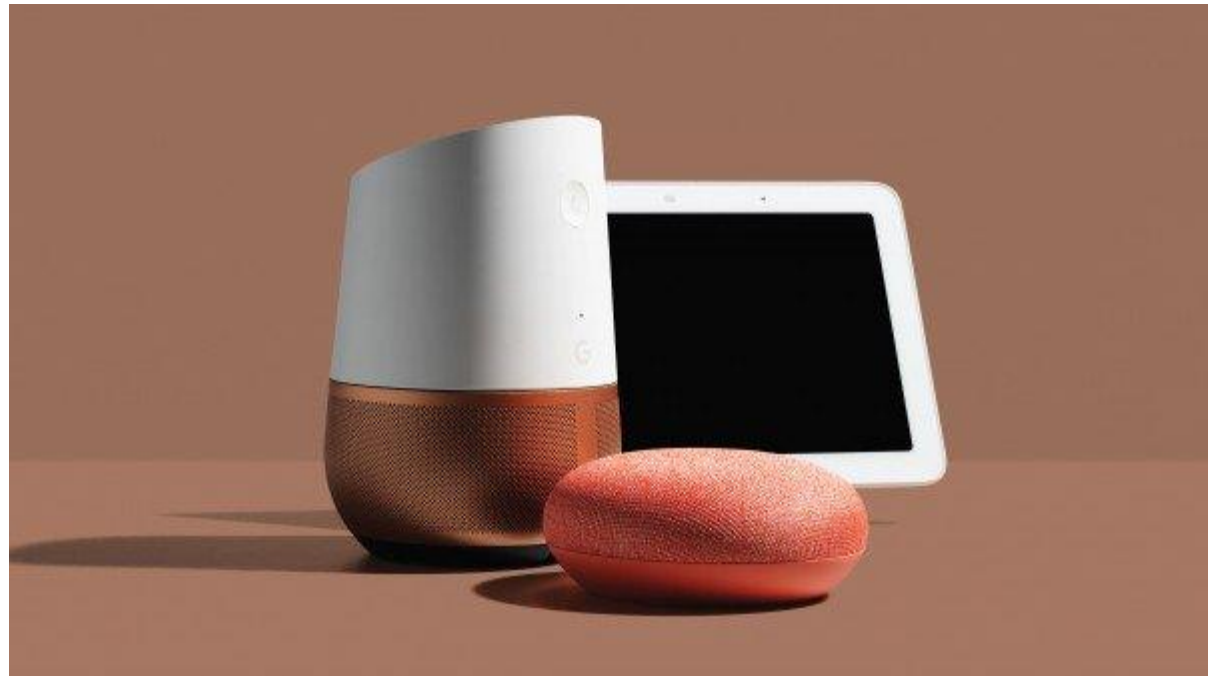


《麻省理工科技评论》2019年十大突破性技术 流利对话的 AI 助手

重大意义：捕捉单词之间语义关系的新技术正在使机器更好地理解自然语言。人工智能助手现在可以执行基于对话的任务，如预订餐厅或协调行李托运，而不仅仅是服从简单命令。

主要研究者：谷歌、阿里巴巴、亚马逊

成熟期：1-2 年后





2016年《麻省理工科技评论》 全球十大突破性技术

多样性/集成性有关
交互性有关
网络有关
媒体处理有关

- ◆1、免疫工程。杀伤性T细胞可被用来消灭癌症。
- ◆2、精确编辑植物基因。精确地编辑植物基因组，不留下外源DNA。
- ◆3、语音接口。将语音识别和自然语言理解相结合，为世界上最大的互联网市场创造切实可行的语音接口。
- ◆4、可回收火箭。可以发射有效载荷至轨道并安全着陆的火箭。
- ◆5、知识分享型机器人。有一种机器人，可以学习任务，同时将知识传送到云端，以供其他机器人学习。
- ◆6、DNA应用商店。DNA测序商业模式让在线获取基因信息成为可能。
- ◆7、SolarCity的超级工厂。通过一种简化的、低成本的制造工艺生产出高效的太阳能电池板。
- ◆8、Slack通信软件。便捷易用通信软件正取代电子邮件成为新的工作协同工具。
- ◆9、特斯拉自动驾驶仪。汽车可以在各种环境下安全自驾。
- ◆10、空中取电。突破技术：新型无线装置，能够利用周边的无线电信号（如Wi-Fi）为自身供电并进行通信。



2017年《麻省理工科技评论》 全球十大突破性技术

多样性/集成性有关
交互性有关
网络有关
媒体处理有关

- ◆ 强化学习（Reinforcement Learning）
- ◆ 360度自拍（The 360-Degree Selfie）
- ◆ 基因疗法2.0（Gene Therapy 2.0）
- ◆ 细胞图谱（The Cell Atlas）
- ◆ 自动驾驶货车（Self-Driving Trucks）
- ◆ 刷脸支付（Paying with Your Face）
- ◆ 太阳能光伏电池（Hot Solar Cells）
- ◆ 实用型量子计算机（Practical Quantum Computers）
- ◆ 治愈瘫痪（Reversing Paralysis）
- ◆ 僵尸物联网（Botnets of Things）



2018年《麻省理工科技评论》 全球十大突破性技术

多样性/集成性有关
交互性有关
网络有关
媒体处理有关

- ◆ 给所有人的人工智能（AI for Everybody）
- ◆ 对抗性神经网络（Dueling Neural Networks）
- ◆ 人造胚胎（Artificial Embryos）
- ◆ 基因占卜（Genetic Fortune-Telling）
- ◆ 传感城市（Sensing City）
- ◆ 即时翻译耳塞（Babel-Fish Earbuds）
- ◆ 完美网络隐私（Perfect Online Privacy）
- ◆ 零碳排放天然气（Zero-Carbon Natural Gas）
- ◆ 3D金属打印机（3-D Metal Printing）
- ◆ 材料的量子飞跃（Materials" Quantum Leap）



2019年《麻省理工科技评论》 全球十大突破性技术

多样性/集成性有关
交互性有关
网络有关
媒体处理有关

- ◆ 灵巧机器人 (Robot dexterity)
- ◆ 核能新浪潮 (New-wave nuclear power)
- ◆ 早产预测 (Predicting preemies)
- ◆ 肠道显微胶囊 (Gut probe in a pill)
- ◆ 定制癌症疫苗 (Custom cancer vaccines)
- ◆ 人造肉汉堡 (The cow-free burger)
- ◆ 捕获二氧化碳 (Carbon dioxide catcher)
- ◆ 可穿戴心电图 (An ECG on your wrist)
- ◆ 无下水道卫生间 (Sanitation without sewers)
- ◆ 流利对话的 AI 助手 (Smooth-talking AI assistants)



Gartner Top 10 Strategic Technology Trends

Gartner每年10月评选10个战略性技术

Gartner，研究与顾问咨询公司，成立于1979年，总部设在美国康涅狄克州斯坦福。其研究范围覆盖全部IT产业，就IT的研究、发展、评估、应用、市场等领域，为客户提供客观、公正的论证报告及市场调研报告，协助客户进行市场分析、技术选择、项目论证、投资决策。





Gartner's top 10 technology trends for 2016

Trend No. 1: The Device Mesh

Trend No. 2: Ambient User Experience

Trend No. 3: 3D-Printing Materials

Trend No. 4: Information of Everything

Trend No. 5: Advanced Machine Learning

Trend No. 6: Autonomous Agents and Things

Trend No. 7: Adaptive Security Architecture

Trend No. 8: Advanced System Architecture

Trend No. 9: Mesh App and Service Architecture

Trend No. 10: Internet of Things Architecture and Platforms

多样性/集成性有关

交互性有关

网络有关

媒体处理有关



Gartner's top 10 technology trends for 2017

◆ Intelligent

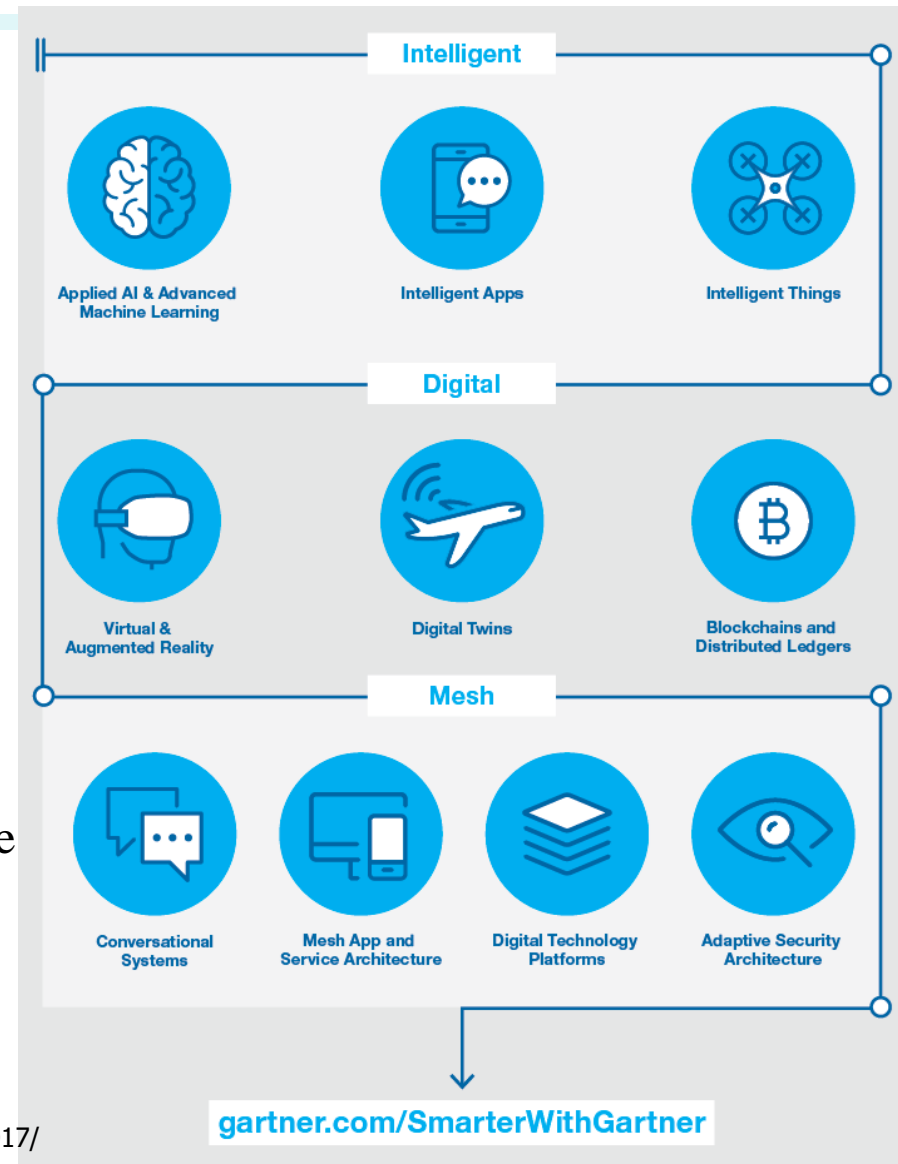
- 1: AI & Advanced Machine Learning
- 2: Intelligent Apps
- 3: Intelligent Things

◆ Digital

- 4: Virtual & Augmented Reality
- 5: Digital Twin
- 6: Blockchain

◆ Mesh

- 7: Conversational Systems
- 8: Mesh App and Service Architecture
- 9: Digital Technology Platforms
- 10: Adaptive Security Architecture





Gartner's top 10 technology trends for 2018

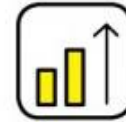
- AI foundation
- Intelligent apps and analytics
- Digital twin
- Cloud to the edge
- Conversational platforms
- **Immersive experience**
- Blockchain
- Event driven
- Continuous adaptive risk and trust



Intelligent



AI Foundations



Intelligent Apps and Analytics



Intelligent Things



Digital



Digital Twins



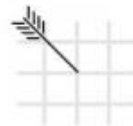
Cloud to the Edge



Conversational Platform



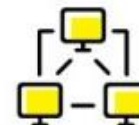
Immersive Experience



Mesh



Blockchain



Event-Driven



Continuous Adaptive Risk and Trust



Top 10 Strategic Technology Trends For 2019

Blockchain, quantum computing, augmented analytics and artificial intelligence will drive disruption and new business models.

- 1: Autonomous things**
- 2: Augmented analytics**
- 3: AI-driven development**
- 4: Digital twins**
- 5: Empowered edge**
- 6: Immersive technologies**
- 7: Blockchain**
- 8: Smart spaces**
- 9: Digital ethics and privacy**
- 10: Quantum computing**





课程内容

- ◆ 第1章 多媒体通信概述
- ◆ 第2章 多媒体数据压缩国际标准
- ◆ 第3章 多媒体同步机制
- ◆ 第4章 多媒体网络QoS
- ◆ 第5章 多媒体传输网络
- ◆ 第6章 多媒体通信终端、系统与国际标准
- ◆ 第7章 多媒体通信技术最新进展



第2章 多媒体数据压缩国际标准

- ◆ 2.1 多媒体数据压缩编码的重要性和分类
 - 2.1.1 为什么要压缩?
 - 2.1.2 为什么能压缩?
 - 2.1.3 数据压缩编码的两大类
- ◆ 2.2 常见数据压缩方法分类与基本原理
- ◆ 2.3 音频压缩标准
- ◆ 2.4 静态图像压缩编码的国际标准
- ◆ 2.5 视频压缩的国际标准
- ◆ 2.6 可伸缩性编码和分布式编码



为什么要压缩?

未压缩基本音源的数据率示例

◆多声道系统：5.1、7.1

质量	采样频率(kHz)	样本精度(bit/s)	单声道/ 立体声	数据率 (kB/s)	声音频率范围
电话	8	8	单声道	8	200~3400 Hz
AM	11.025	8	单声道	11.0	50~7000Hz
FM	22.050	16	立体声	88.2	20~15000Hz
CD	44.1	16	立体声	176.4	20~20000 Hz
DAT	48	16	立体声	192.0	20~20000 Hz
HDCD	96	24	立体声	384	20~20000 Hz

注：* 电话使用 μ 律编码，动态范围为 13^{26} 比特，而不是8比特



为什么要压缩？

未压缩图像的数据率示例

- ◆ 例如，扫描仪产生的图像
 - 分辨率：300×600dpi、600×1200dpi、1200×2400dpi甚至更高
 - 色彩深度：24位、36位甚至更高
 - A4纸张：8.27英寸×11.69英寸
- ◆ $(8.27*1200) * (11.69*2400) *(24/8) \approx 280\text{M Bytes}$
- ◆ 如，2000万像素的数码相机
 - $20000000 * 3 \approx 180\text{M Bytes}$



为什么要压缩？

无压缩视频源的数据率需求

◆ 例如：一个 512×512 分辨率的一分钟视频需要如下的存储容量：

$$512 \times 512 \times 8 \times 3 = 6291456 = 6.3 \text{M bit/s}$$

$$6.3 \times 30 \text{帧/s} = 188 \text{M bit/s}$$

$$188 / 8 = 23.5 \text{MByte/s}$$

那么一张650MByte的光盘就只能存储

$$650 \text{M byte 光盘} / 23.5 \text{M bytes/s} = 27.5 \text{秒}$$

◆ 虽然4.7G的DVD已经非常普遍，但是视频→HDTV



为什么要压缩？ URL的编码方式

【百度面试题】一个大的含有50M个URL的记录，一个小的含有500个URL的记录，找出两个记录里相同的URL。

- 先用包含500个URL的文件创建一个hash_set。
- 然后遍历50M的URL记录，如果URL在hash_set中，则输出此URL并从hash_set中删除这个URL。
- 所有输出的URL就是两个记录里相同的url。

URL示例

http://baike.baidu.com/link?url=NGckZcowcm9U0k3ObaLLoUjVRWI-xZ1IXzIyB9_4lgHHuReW3X9bvmlsSXDgHVIClywp4sKwemSAHMqzmncMSa

URL的信息指纹：（网络爬虫需要存储URL）假定网址的平均长度为一百个字符，那么存贮 200 亿个网址本身至少需要 2 TB，即两千 GB 的容量。如果能够找到一个函数，将这 200 亿个网址随机地映射到128 二进位即 16 个字节的整数空间，这样每个网址只需要占用 16 个字节。



小结：为什么要压缩 海量数据与有限处理能力的矛盾

◆多媒体信息包括文本、数据、声音、动画、图形以及视频等多种媒体信息，经过数字化处理后其数据量非常大，如果不进行数据压缩处理，计算机系统就无法对它进行存储和交换；

◆多媒体信息中的图像、音频和视频这些媒体具有很大的压缩潜力。因为这些数据中存在着空间冗余、时间冗余、结构冗余、视觉冗余、图像区域的相同性冗余、纹理的统计冗余等，它们为数据压缩技术的应用提供了可能的条件。

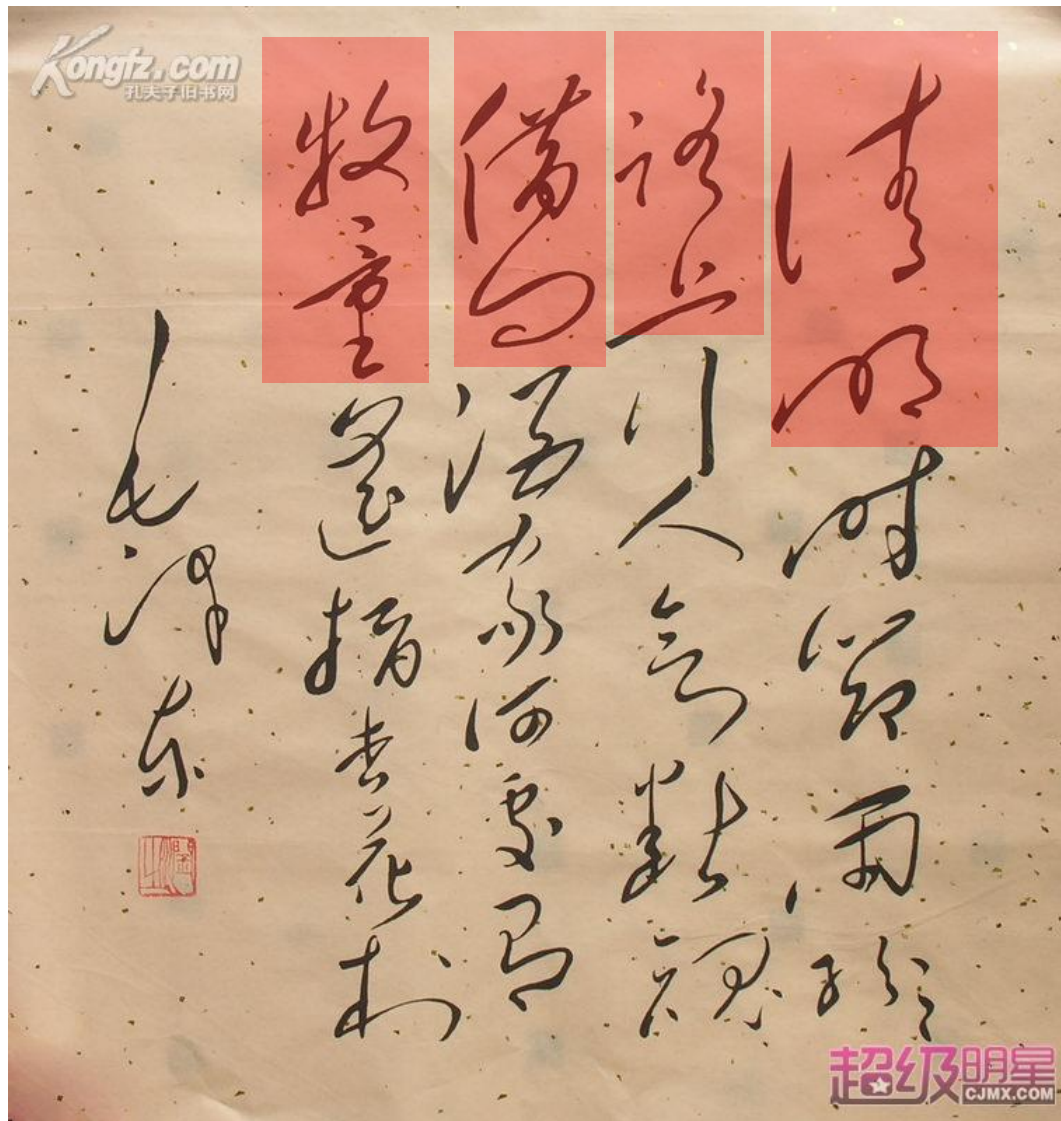


第2章 多媒体数据压缩国际标准

- ◆ 2.1 多媒体数据压缩编码的重要性和分类
 - 2.1.1 为什么要压缩?
 - 2.1.2 为什么能压缩?
 - 2.1.3 数据压缩编码的两大类
- ◆ 2.2 常见数据压缩方法分类与基本原理
- ◆ 2.3 音频压缩标准
- ◆ 2.4 静态图像压缩编码的国际标准
- ◆ 2.5 视频压缩的国际标准
- ◆ 2.6 可伸缩性编码和分布式编码



为什么能压缩 冗余





为什么能压缩 信息量与数据量的关系

◆ $I = D - du$

◆ I—信息量； D—数据量； du—冗余量

◆ du在平时说话时是大量存在的。例如，中文广播员一分钟读180个汉字。（1）采用文字存储：一个汉字两个字节，360个Byte。（2）采用音频存储：采样1分钟， $8k \times 60 = 480 \text{ kByte}$

◆ **$480 \text{ kByte} / 360 \text{ Byte} = 1000$ 倍的冗余**



为什么能压缩 信息量与信息熵

- ◆ **信息量**：指从N个相等的可能事件中选出一个事件所需要的信息度和含量。
- ◆ **信息熵**：指一团数据所带的信息量，平均信息量就是信息熵（entropy）。
- ◆ 例如：从64个数中选出某一个数，可先问“是否大于32？”消除半数的可能，这样只要6次就可选出某数。这是因为每提问一次都会得到1比特的信息量。因此，在64个数中选定某一数所需的信息量是 $\log_2 64=6(\text{bit})$ 。
- ◆ 设从N个数中选任意一个数X的概率为 $P(x)$ ，假定选定任意一个数的概率都相等， $P(x)=1/N$ ，因此定义信息量： $I(x)=\log_2 N = -\log_2(1/N) = -\log_2 P(x) = I[P(x)]$



为什么能压缩 图像中的冗余信息

◆ 空间冗余

- 例：一幅书法作品的图像。图像中背景全部是白色。

◆ 结构冗余

- 图像有非常强的纹理结构。如草席图结构上存在冗余。

◆ 知识冗余

- 图像的理解与某些知识有关。例：人脸图像有同样的结构：嘴的上方有鼻子，鼻子上方有眼睛，鼻子在中线上……

◆ 视觉冗余

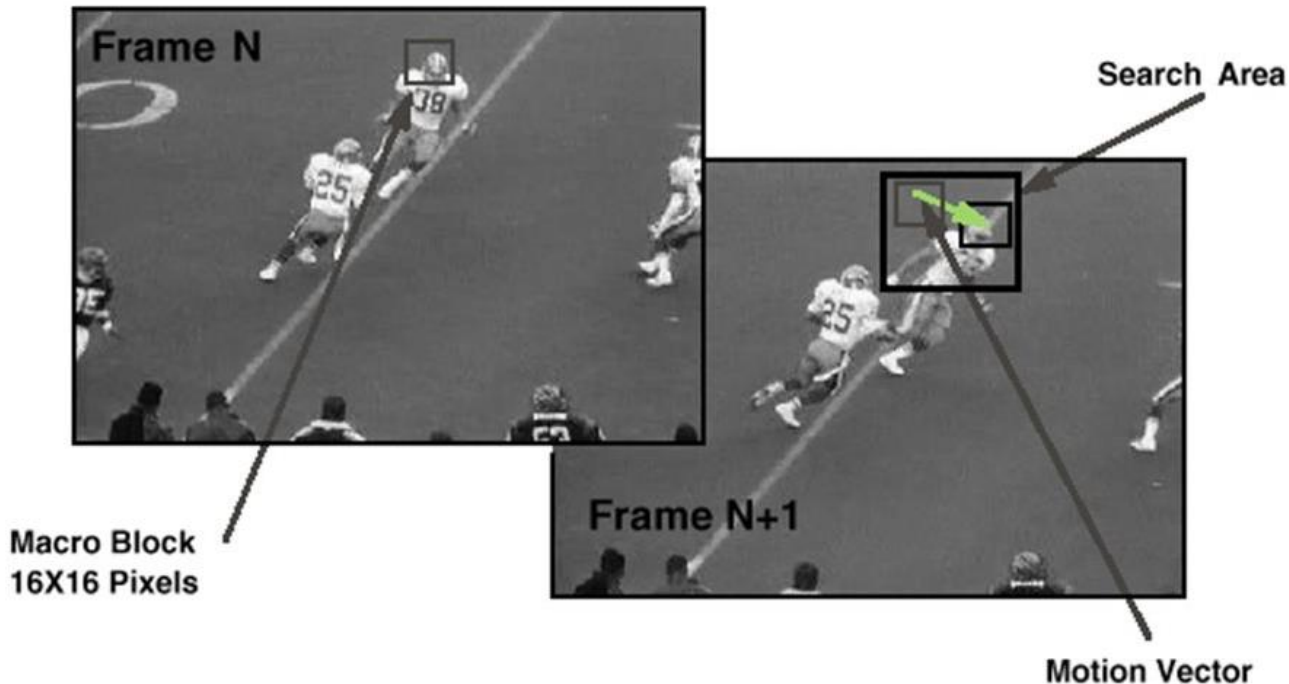
- 视觉冗余是非均匀、非线性的。例，对亮度变化敏感，而对色度的变化相对不敏感；在高亮度区，人眼对亮度变化敏感度下降对物体边缘敏感，内部区域相对不敏感；对整体结构敏感，而对内部细节相对不敏感。



为什么能压缩 视频中的冗余信息（时间冗余）

◆以连续视频为例，它的每一帧画面是由若干个像素组成的，因为动态图像通常反映的是一个连续的过程，它的相邻的帧之间存在着很大的相关性，从一幅画面到下一幅画面，背景与前景就可以没有太多的变化。

连续多帧画面在很大程度上是相似的，而这些相似的信息（或称作冗余信息）为数据的压缩提供了基础。





小结：为什么能压缩 冗余无处不在

- ◆ 面向模拟信号的处理：如模拟功放
 - 小信号放大器、模拟滤波器、功率放大器、
- ◆ 面向数字信号的处理：如数字功放
 - 数字滤波器、D/A、功率放大器
- ◆ 面向信息的处理？
 - 信息的表示
 - 人的感知特性
 - 网络中信息的提取
 - 源自网络信息的冗余去除



第2章 多媒体数据压缩国际标准

- ◆ 2.1 多媒体数据压缩编码的重要性和分类
 - 2.1.1 为什么要压缩?
 - 2.1.2 为什么能压缩?
 - 2.1.3 数据压缩编码的两大类
- ◆ 2.2 常见数据压缩方法分类与基本原理
- ◆ 2.3 音频压缩标准
- ◆ 2.4 静态图像压缩编码的国际标准
- ◆ 2.5 视频压缩的国际标准
- ◆ 2.6 可伸缩性编码和分布式编码

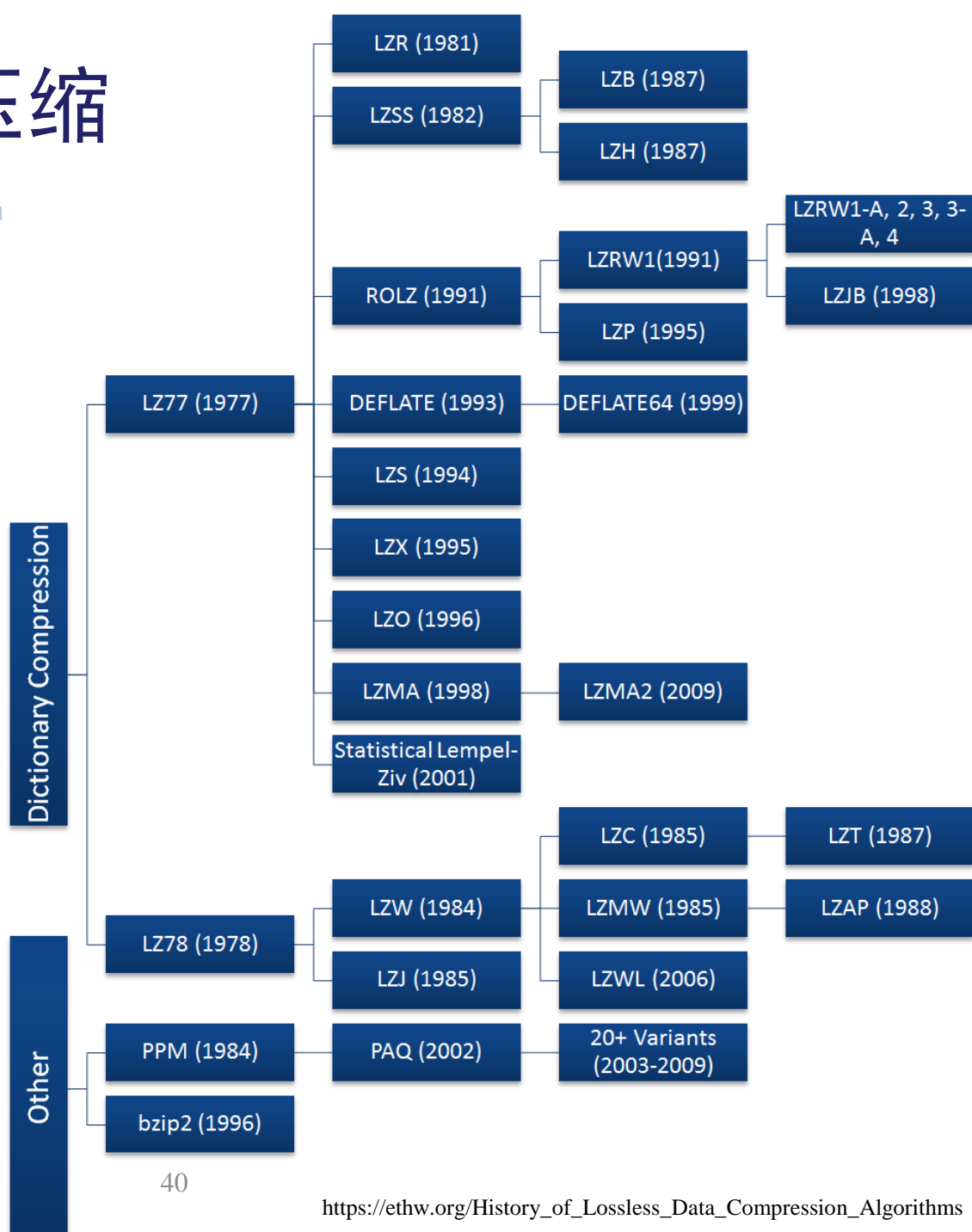


2.1.3 数据压缩编码的两大类

- ◆一类是冗余压缩法；也称**无损压缩**或无失真压缩。它是为保留原始多媒体对象（包括图像、语音和视频）而设计的。在无损压缩中，数据在压缩或解压缩过程中不会改变或损失，**解压缩产生的数据是对原始对象的完整复制**。
- ◆二类是熵压缩法；也称**有损压缩**或有失真压缩。当音频/图像/视频的冗余度很少时，用无损压缩技术不能得到可接受的结果，这时就要采用有损压缩。有损压缩会造成一些信息的损失，**关键问题是看这种损失对图像质量带来的影响**。只要这种损失被限制在允许的范围内，有损压缩就是可接受的。



无损数据压缩





无损数据压缩的朴素思想 基于字典

"But I don't want to go among mad people," Alice remarked.
 "Oh, you can't help that," said the Cat: "we're all mad here. I'm mad. You're mad."
 "How do you know I'm mad?" said Alice.
 "You must be," said the Cat, or you wouldn't have come here."
 - Lewis Carroll, Alice in Wonderland

Input:		a	b	ba	d	ab	baa	baad
Dictionary Index	0	1	2	3	4	5	6	7
Output	NULL	(0,a)	(0,b)	(2,a)	(0,d)	(1,b)	(3,a)	(6,d)



无损数据压缩的朴素思想

Run-Length Encoding

Input: AAABBBCCCDEEEEEEA
AAAAAAAAAAAAAAAAAAAA
AAAAAA

Output: 3A2B4C1D6E38A



无损数据压缩的朴素思想

Burrows-Wheeler Transform, 1994

Input	Rotations	Alpha-Sorted Rotations	Output
HAHAHA&	HAHAHA&	AHAHA& <u>H</u>	<u>HHH&AAA</u>
	&HAHAHA	AHA&HA <u>H</u>	
	A&HAHAH	A&HAHA <u>H</u>	
	HA&HAHA	HAHAHA <u>&</u>	
	AHA&HAH	HAHA&H <u>A</u>	
	HAHA&HA	HA&HAH <u>A</u>	
	AHAHA&H	&HAHAH <u>A</u>	



影响数据压缩的几个要素：压缩比

◆对压缩前后的文件大小和数据量进行比较，作为压缩率的衡量指标。人们普遍希望压缩的倍数越高越好，压缩的速度越快越好，并且希望压缩所耗费的资源、时间对用户来讲是透明的。但同时人们又希望确保数据压缩的精度，即压缩完了以后，解压缩的数据和原来的数据最好没有什么差别，没有什么数据损失。然而追求压缩比率和追求精度往往是矛盾的，因此就需要在这两者之间权衡取舍。



经典无损数据压缩算法的压缩比

Date	bpc	scheme	authors
May 1977	3.94	LZ77	Ziv, Lempel
1984	3.32	LZMW	Miller and Wegman
1987	3.30	LZH	Brent
1987	3.24	MTF	Moffat
1987	3.18	LZB	Bell
.	2.71	GZIP	.
1988	2.48	PPMC	Moffat
.	2.47	SAKDC	Williams
Oct 1994	2.34	PPM*	Cleary, Teahan, Witten
1995	2.29	BW	Burrows, Wheeler
1997	1.99	BOA	Sutton
1999	1.89	RK	Taylor



影响数据压缩的几个要素：**速度**

◆ **压缩**和**解压缩**的速度是压缩系统的两项单独的性能度量。在有些应用中，压缩和解压缩都需要实时进行，如电视会议的图像传输。在有些应用中，压缩可以用非实时压缩，而只要解压缩是实时的，这种压缩称为非对称压缩，如多媒体CD-ROM的节目制作。从目前开发的压缩技术来看，一般压缩的计算量比解压缩要大。

◆ 压缩的速度不仅与采用的压缩方法有关，而且与快速算法的**计算量**有关，如果在算法上有较大的突破，无疑将对多媒体的开发与应用产生很大的影响。



影响数据压缩的几个要素：**质量**

- ◆有损压缩可获得较大的压缩比，但压缩比过高，还原后的音频/图像/视频质量就可能降低。音频/图像/视频质量的评估法常采用客观评估和主观评估两种方法。
- ◆**客观评估**则是通过一种具体的算法来统计多媒体数据压缩结果的损失。例如采用纯误差测度，即原始图像与恢复图像之间的简单数学统计差别。。
- ◆**主观评估**基于人的视觉感知，因为观察者作为最终视觉信宿，他们能对恢复音频/图像/视频的质量作出直观的判断。方法是让观察者通过观测一系列恢复音频/图像/视频，并与原始信息进行比较，再根据损伤的可见程度进行评级，以判断哪种压缩方法的失真少。

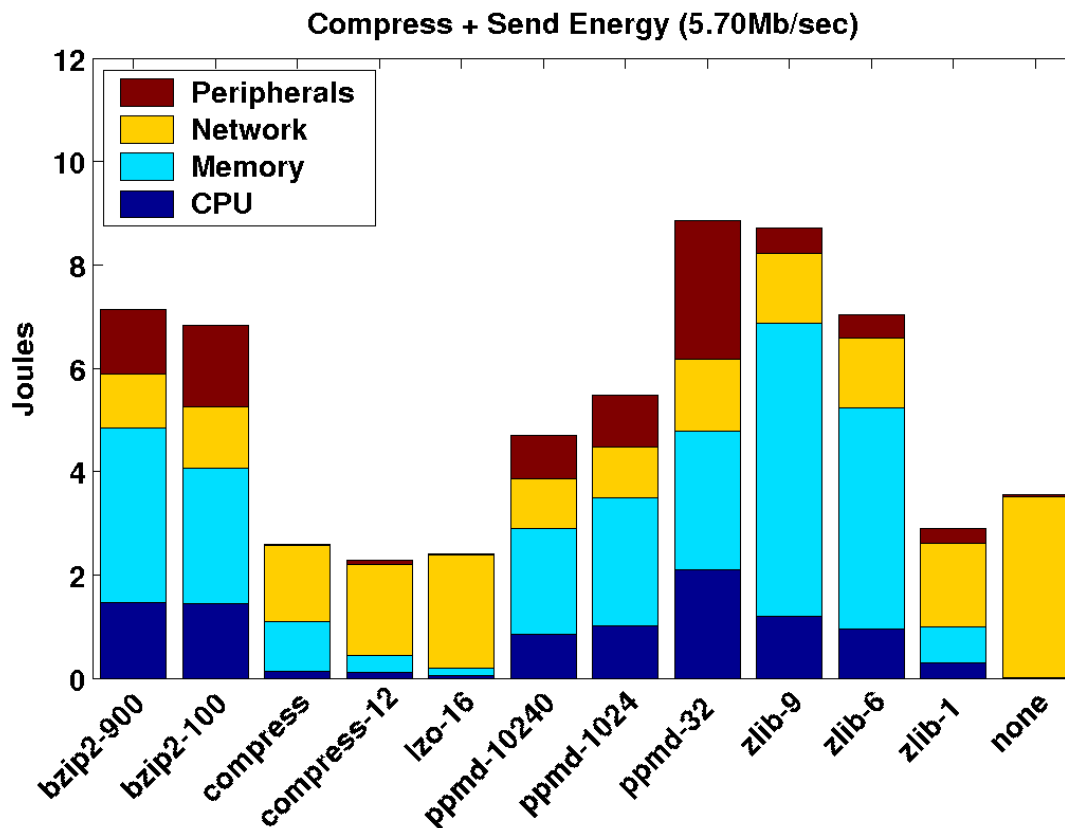
MC影响数据压缩的几个要素：硬件/软件

- ◆采用什么样的硬件与软件去执行压缩/解压缩，与采用压缩方案和算法的复杂程度有着密切的关系。
- ◆设计精巧的简单算法可以在简单的硬件上执行，且执行速度很快。而设计复杂的算法需要在功能强大的硬件和软件的支持下才能运行。
- ◆但仅靠算法来提高压缩/解压缩的速度还是有限的。在大多数情况下，不得不依靠硬件本身提供的功能去完成，例如采用专用多媒体处理芯片。因此在压缩/解压缩系统中，速度和硬件之间的选择就显得十分重要。



低碳经济下的新要素：功耗

- Parameters: size of input blocks, size of data structures, amount of effort
- Use such a chart to choose best compressor for platform+data combo





小结：压缩的需求、可能性

◆ 为什么要压缩？

- 音频/图像/视频的海量数据

◆ 为什么能压缩？

- 冗余

◆ 数据压缩编码的两大类

□ 无损压缩

- 压缩的思路较为朴素

□ 有损压缩

- 与人感知能力有关系

□ 衡量准则的Tradeoff

- 压缩比与质量
- 速度与成本



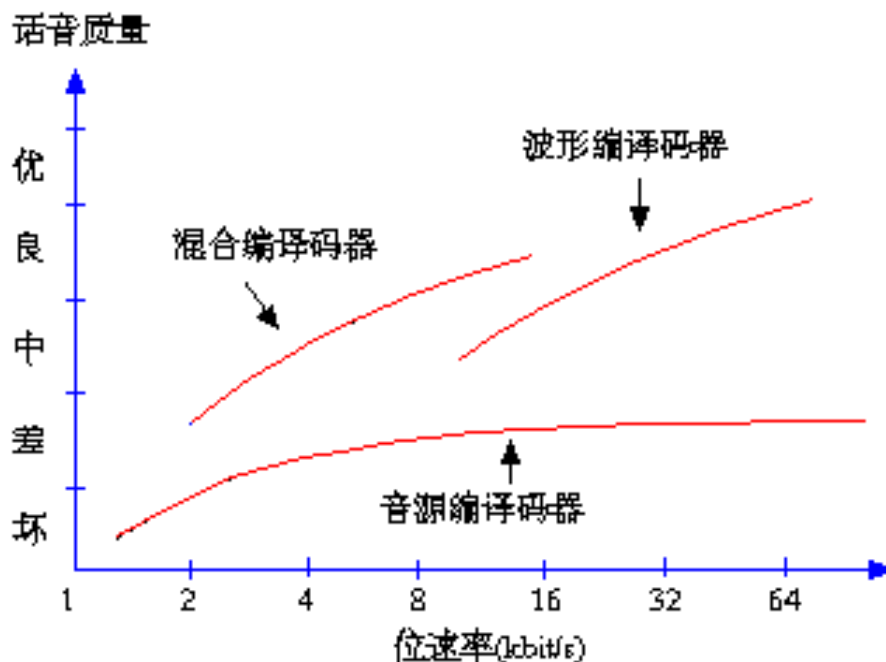
第2章 多媒体数据压缩国际标准

- ◆ 2.1 多媒体数据压缩编码的重要性和分类
- ◆ 2.2 常见数据压缩方法分类与基本原理
 - 2.2.1 量化与向量量化原理
 - 2.2.2 统计编码原理
 - 2.2.3 Huffman编码
 - 2.2.4 算术编码
 - 2.2.5 预测编码
 - 2.2.6 变换编码
- ◆ 2.3 音频压缩标准
- ◆ 2.4 静态图像压缩编码的国际标准
- ◆ 2.5 视频压缩的国际标准
- ◆ 2.6 可伸缩性编码和分布式编码



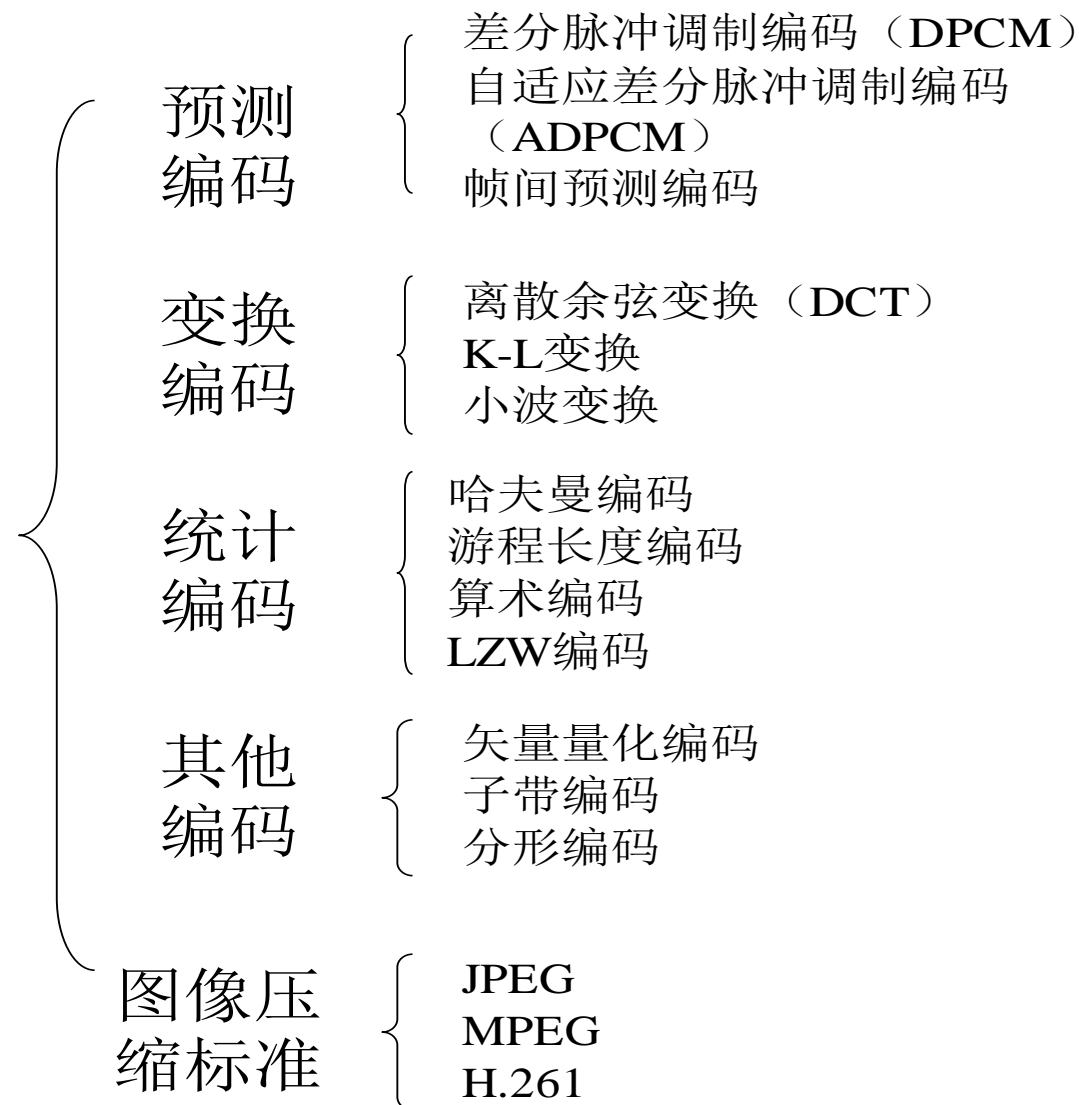
话音编译码器分类的一般方法

- ◆ **波形编码**：不利用声音的任何知识，数据率较高，实现简单
- ◆ **音源编码**：从声音的波形中提取生成话音的参数，数据率可以很低，实现复杂
- ◆ **混合编码**：以上两种思想的结合





图像压缩编码分类的一般方法





数据压缩方法按编码算法原理分类

- ◆ 量化与向量量化编码
- ◆ 统计编码
- ◆ 预测编码
- ◆ 变换编码
- ◆ 信息熵编码
- ◆ 子带编码
- ◆ 结构编码
- ◆ 基于知识的编码



第2章 多媒体数据压缩国际标准

- ◆ 2.1 多媒体数据压缩编码的重要性和分类
- ◆ 2.2 常见数据压缩方法分类与基本原理
 - 2.2.1 量化与向量量化原理
 - 2.2.2 统计编码原理
 - 2.2.3 Huffman编码
 - 2.2.4 算术编码
 - 2.2.5 预测编码
 - 2.2.6 变换编码
- ◆ 2.3 音频压缩标准
- ◆ 2.4 静态图像压缩编码的国际标准
- ◆ 2.5 视频压缩的国际标准
- ◆ 2.6 可伸缩性编码和分布式编码



量化原理：量化的目的

◆ 量化处理是使数据比特率下降的一个强有力的措施。脉冲编码调制（PCM）的量化处理是采样之后进行，从理论分析的角度，图像灰度值是连续的数值，而我们通常看到的是以（0~255）的整数表示图像灰度，这是经A/D变换后的以256级灰度分层量化处理了的离散数值，这样可以用 $\log_2 256 = 8$ 比特表示一个图像像素的灰度值，或色差信号值。

MC 量化的成本？



Part#	Resolution (Bits)	Throughput Rate	US-Price-1000-to-4999
Results: 564			
<u>AD9253-105</u>	14	105MSPS	\$99.45
<u>AD9633-105</u>	12	105MSPS	\$57.8
<u>AD9628-105</u>	12	105MSPS	\$27.57
<u>AD9284</u>	8	250MSPS	\$25
<u>AD9286</u>	8	500MSPS	\$36



量化原理：量化的本质

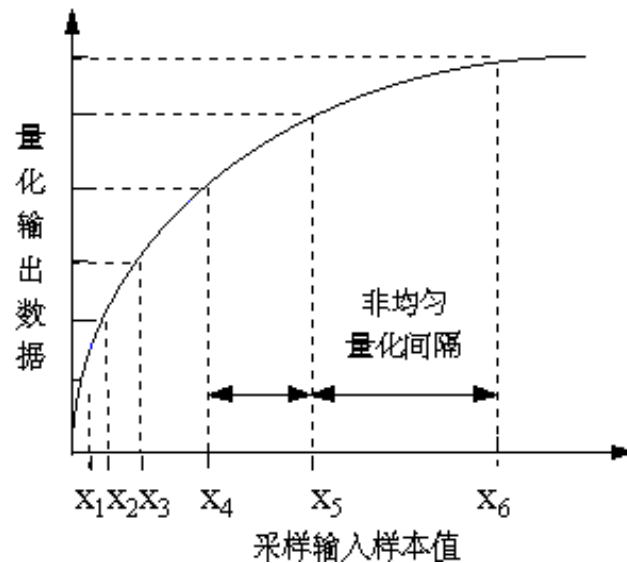
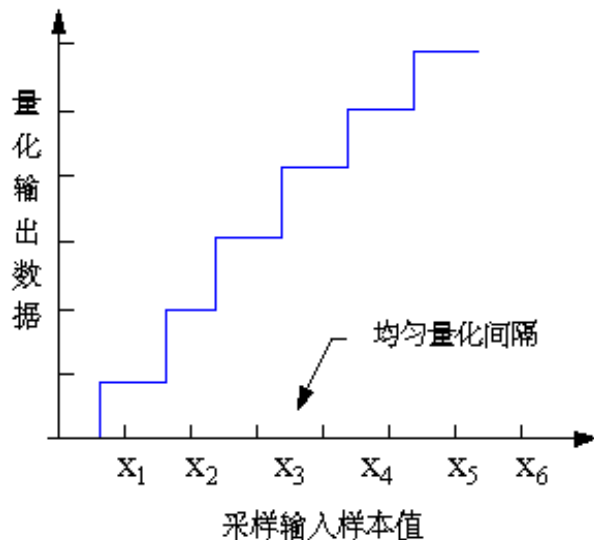
◆ 量化输入值的动态范围很大，需要很多的比特数表示一个数值，量化输出只能取有限个整数，称作量化级，希望量化后的数值用较少的比特数便可表示。每个量化输入被强行归到与其接近的某个输出，即量化到某个级。量化处理总是把一批输入，量化到一个输出级上，所以**量化处理是个多对一的处理过程**，是个**不可逆过程**，量化处理中有信息丢失，或者说，会引起量化误差（量化噪声）。

◆ **变换域的量化**：数据压缩编码中的量化处理，不是指A/D变换后的量化，而是指以PCM码作为输入，经正交变换、差分、或预测处理后，熵编码之前，对正交变换系数、差值或预测误差的量化处理。

◆ 通常设计量化器有下述两种情况：

- 给定量化分层级数，满足量化误差最小。
- 限定量化误差，确定分层级数，满足以尽量小的平均比特数，表示量化输出。

◆ 根据量化方式可分为：均匀量化、非均匀量化。

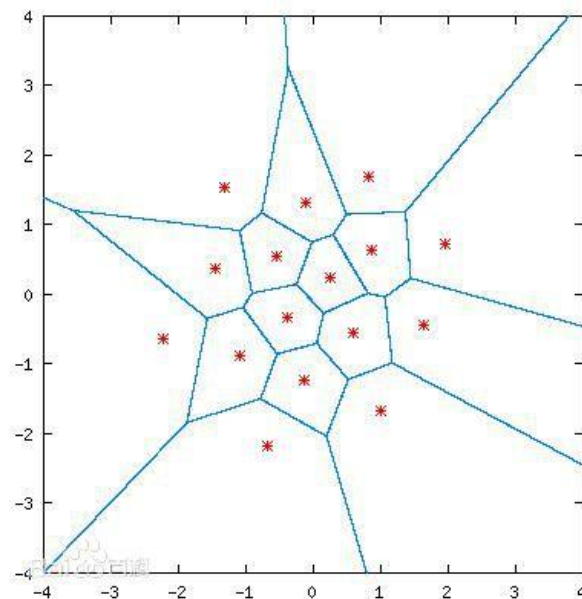


- ◆ 矢量量化编码是近年来图像、语音信号编码技术中颇为流行的一种新型量化编码方法。
- ◆ 矢量量化编码方法一般是有失真编码方法。矢量量化的名字是相对于标量量化而提出的。对于PCM数据，一个数一个数地进行量化叫**标量量化**。
- ◆ 若对这些数据分组，每组 K 个数构成一个 K 维矢量，然后以矢量为单元，逐个矢量进行量化，**称矢量量化**。

MC 矢量量化

◆ 矢量量化是70年代后期发展起来的一种数据压缩技术基本思想：将若干个标量数据组构成一个矢量，然后在矢量空间给以整体量化，从而压缩了数据而不损失多少信息。从 N 维实空间 \mathbf{R}^N 到 \mathbf{R}^N 中 L 个离散矢量的映射，也可称为分组量化，标量量化是矢量量化在维数为1时的特例。

在传统编码中，首先将信号经某种映射变换变成一个数的序列，然后对其一个一个地进行标量量化编码。而在矢量量化编码中，则是把输入数据几个一组地分成许多组，成组地量化编码，即将这些数看成一个 k 维矢量，然后以矢量为单位逐个矢量进行量化。





第2章 多媒体数据压缩国际标准

- ◆ 2.1 多媒体数据压缩编码的重要性和分类
- ◆ 2.2 常见数据压缩方法分类与基本原理
 - 2.2.1 量化与向量量化原理
 - 2.2.2 统计编码原理
 - 2.2.3 Huffman编码
 - 2.2.4 算术编码
 - 2.2.5 预测编码
 - 2.2.6 变换编码
- ◆ 2.3 音频压缩标准
- ◆ 2.4 静态图像压缩编码的国际标准
- ◆ 2.5 视频压缩的国际标准
- ◆ 2.6 可伸缩性编码和分布式编码



Morse Code: e和t最短

Samuel Morse, telegraph key circa 1860



An 1845 version of the finger key, which replaced type and portrule.

- A{-}滴滴
- B{---}答滴滴滴
- C{--}答滴滴滴
- D{-}答滴滴
- E{-}滴**
- F{--}滴滴答滴
- G{-}答答滴
- H{----}滴滴滴滴
- I{-}滴
- J{----}滴滴答答
- K{-}答滴滴
- L{-}滴滴滴滴
- M{-}答答
- N{-}答滴
- O{---}答答答
- P{---}滴滴答滴
- Q{---}答答滴答
- R{-}滴滴答滴
- S{---}滴滴滴
- T{-}答**
- U{-}滴滴答
- V{---}滴滴滴滴
- W{-}滴滴答
- X{---}答滴滴答
- Y{---}答滴滴答
- Z{---}答答滴滴
- 1{-----}
- 2{-----}
- 3{-----}
- 4{-----}
- 5{-----}
- 6{-----}
- 7{-----}
- 8{-----}
- 9{-----}
- 0{-----}

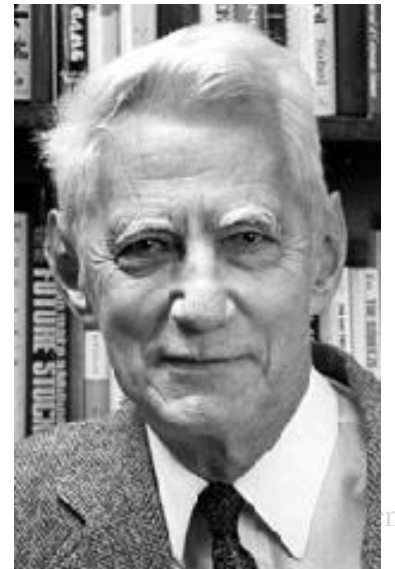


Samuel Morse photo by student Mathew Brady, 1845

MC 信息的科学定义

◆ 信息论有关**信息熵的概念**对数据压缩有着重要的指导意义，它一方面给出了数据压缩的理论极限，另一方面又指明了数据压缩的技术途径。现代信息论的创始人香农（C.E Shannon, 1916-2001）在他发表的著名论文《通信的数学理论，1948》中，从研究通信系统传输的实质出发，对信息作了科学的定义，并进行了定性和定量的描述。

香农认为：**信息是有秩序的量度，是人们对事物了解的不确定性的消除或减少。信息是对组织程度的一种测度，信息能使物质系统有序性增强，减少破坏、混乱和噪音。**



- ◆ 香农提出：信息的传播过程是“信源”（信息的发送者）把要提供的信息经过“信道”传递给“信宿”（信息的接收者），信宿接收这些经过“译码”（即解释符号）的信息符号的过程。并由此建立了通信系统模型。我们从一则消息中获得了信息，那么我们获得的信息有多少呢？**度量信息多少的测度就是信息量。**信息的度量反映了人们对于信息的定量认识。
- ◆ 根据香农的有关信息的定义，信息如何测度呢？显然，信息量与不确定性消除程度有关。消除多少不确定性，就获得多少信息量。用数学的语言来讲，不确定就是随机性。不确定性的大小可以直观地看成是事先猜测某随机事件是否发生的难易程度。



The Entropy of the English Language

◆ If we assume equal probabilities for all characters, a separate code for each character, and that there are 96 printable characters (the number on a standard keyboard) then each character would take $\lceil \log 96 \rceil = 7$ bits.

	<i>bits/char</i>
bits $\lceil \log(96) \rceil$	7
entropy	4.5
Huffman Code (avg.)	4.7
Entropy (Groups of 8)	2.4
Asymptotically approaches:	1.3
Compress	3.7
Gzip	2.7
BOA	2.0

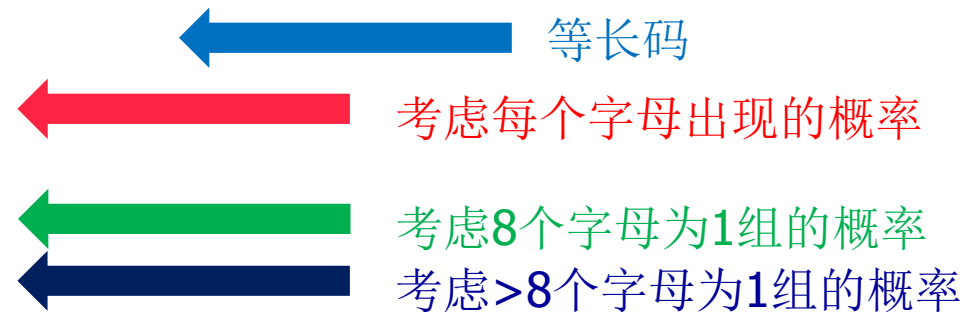


Table 1: Information Content of the English Language



中文词频规律示例

表1 文学作品中的词频

排序	《红楼梦》		《毛泽东选集》		《邓小平文选》		《莎士比亚全集》	
	词	词频	词	词频	词	词频	词	词频
1	了	0 040 73	的	0 085 48	的	0 079 46	the	0 031 83
2	的	0 027 15	和	0 016 42	了	0 015 70	and	0 029 37
3	我	0 010 88	在	0 015 04	我们	0 015 12	I	0 023 54
4	不	0 010 66	了	0 014 50	是	0 012 66	to	0 021 55
5	贾	0 010 10	是	0 011 57	在	0 011 49	of	0 019 51
6	道	0 008 69	一	0 005 82	和	0 010 40	a	0 016 56
7	你	0 008 50	我们	0 005 78	要	0 008 55	you	0 015 69
8	也	0 007 81	中国	0 005 44	问题	0 007 79	my	0 013 92
9	宝玉	0 007 75	为	0 004 88	有	0 007 43	in	0 012 27
10	来	0 007 67	有	0 004 48	不	0 007 29	that	0 012 17
11	又	0 007 11	不	0 004 39	党	0 006 43	is	0 010 35
12	便	0 006 65	地	0 004 17	就	0 005 45	not	0 009 59
13	说	0 006 30	革命	0 004 08	工作	0 005 22	me	0 008 78
14	这	0 006 09	而	0 003 92	也	0 005 05	it	0 008 75
15	去	0 005 55	国民党	0 003 91	一个	0 004 36	with	0 008 70

MC 等长信源编码定理

◆ 等长编码定理给出了信源进行等长编码所需码长的理论极限值。一个熵为 $H(S)$ 的离散无记忆信源，若对长为 N 的符号序列进行等长编码，设码字是从 r 个字母的码符号集中，选取 L 个码元组成。对于任意 $\varepsilon_N > 0$ ，只要满足

$$L \log r \geq N [H(S) + \varepsilon_N]$$

则当 N 足够大时，可实现几乎无失真编码，即译码错误概率能为任意小。反之，则不可能实现无失真编码。

MC 变长信源编码定理

◆ 香农第一定理，这个定理是香农信息论中非常重要的一个定理，它指出，要做到无失真的信源编码，信源每个符号所需要的**平均码元数就是信源的熵值**，如果小于这个值，则唯一可译码不存在，可见，熵是无失真信源编码的极限值。定理还指出，通过对扩展信源进行编码，当 N 趋向于无穷时，平均码长可以趋进该极限值。

MC 变长信源编码

◆在变字长编码中，对于出现概率大的信息符号，编以短字长的码，对于出现概率小的信息符号编以长字长的码，如果码字长度严格按照符号概率的大小的相反顺序排列，则平均码字长一定小于按任何其他符号顺序排列方式得到的码字长度。

◆设信源有 q 个符号，第 i 个消息以长为 l_i 的 r -元码表示，当信源为无记忆时，平均每个信源号所需的码长为平均码长：

$$\bar{L} = \sum_{i=1}^q P(s_i) l_i$$



变长码与定长码对比

◆变长码与定长码的优缺点：同步、译码延时、效率方面。

- 码A与信源字母集不是一一对应的；
- 码B虽然与信源字母集一一对应，但仍不是唯一可译的；
- 码C和码D是唯一可译的。

信源符号集	概率	码A	码B	码C	码D
a1	0.5	0	0	0	0
a2	0.25	0	1	10	01
a3	0.125	1	00	110	011
a4	0.125	10	11	111	0111



第2章 多媒体数据压缩国际标准

- ◆ 2.1 多媒体数据压缩编码的重要性和分类
- ◆ 2.2 常见数据压缩方法分类与基本原理
 - 2.2.1 量化与向量量化原理
 - 2.2.2 统计编码原理
 - 2.2.3 Huffman编码
 - 2.2.4 算术编码
 - 2.2.5 预测编码
 - 2.2.6 变换编码
- ◆ 2.3 音频压缩标准
- ◆ 2.4 静态图像压缩编码的国际标准
- ◆ 2.5 视频压缩的国际标准
- ◆ 2.6 可伸缩性编码和分布式编码



Huffman 编码基本思路

◆ Huffman 编码就是利用变字长最佳编码实现信源符号按概率大小顺序排列。

- (1) 先将信源符号按概率大小顺序排列
- (2) 出现概率最小的两个符号概率相加合成一个概率。
- (3) 将合成概率看成一个新组合符号概率，重复上述做法，直到最后只剩下两个符号概率为止。
- (4) 反过来逐步向前编码，每一步有两个分支各赋予一个二进制码，可以对概率大的编码为0。



Huffman 编码举例

◆ 例如，信源有四个符号：

X	a1	a2	a3	a4
P(X)	1/2	1/4	1/8	1/8

◆ 信息熵：

- $H(x) = -1/2 \log_2(1/2) - 1/4 \log_2(1/4) - (1/8 \log_2(1/8))^*2$
- = 1.75 bit/字符

◆ 采用三元（二进制编码）

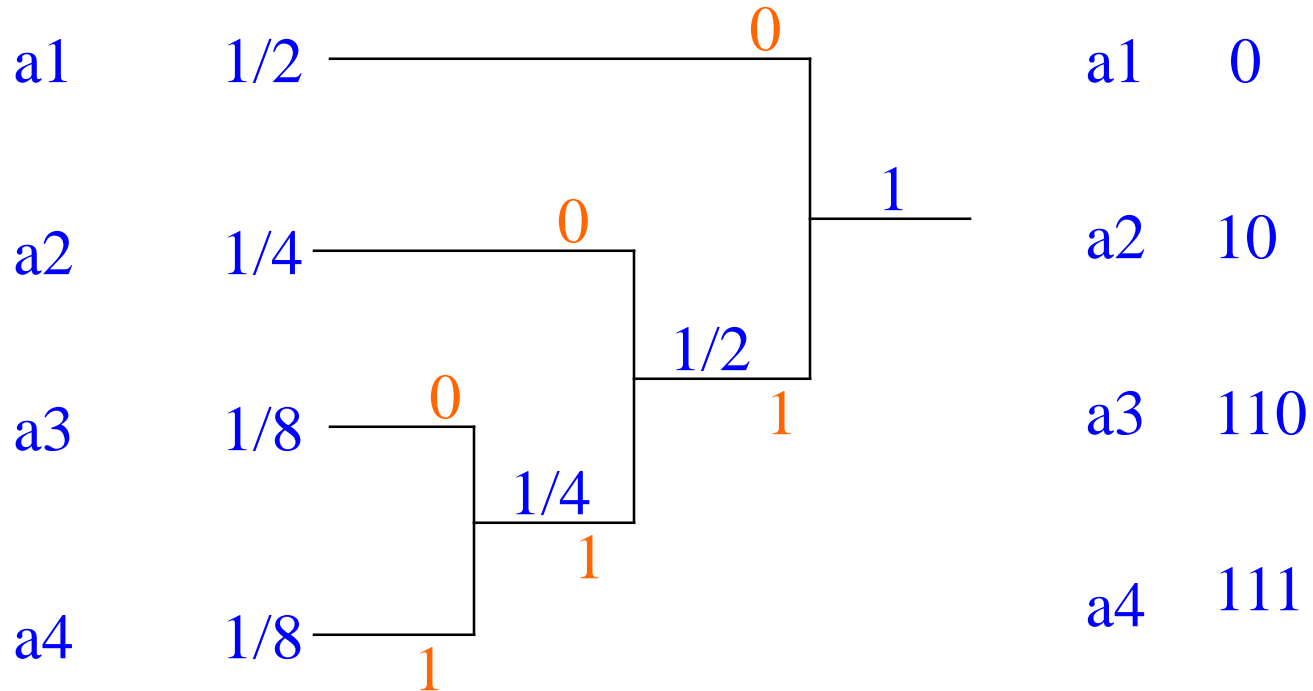
a1	a2	a3	a4
0	10	110	111

◆ 平均码长：

- $L = (1/2) * 1 + (1/4) * 2 + (1/8) * (3+3) = 1.75 \text{ bit/字符}$



Huffman 编码树



$$L(a_1)=1 \quad L(a_2)=2 \quad L(a_3)=3 \quad L(a_4)=3$$

MC 编码效率

◆ Huffman编码效率: $=1.75/1.75=100\%$

◆ 4个符号 PCM编码:

a1 a2 a3 a4

00 01 10 11

$R = \log_2 4 = 2 \text{ bit}$

$L = 2$, 编码效率: $=H(x)/L = 1.75/2 = 87.5\%$

◆ Huffman 编码字长参差不齐 ;

◆ Huffman编码在信源编码概率分布不均匀时效率高



第2章 多媒体数据压缩国际标准

- ◆ 2.1 多媒体数据压缩编码的重要性和分类
- ◆ 2.2 常见数据压缩方法分类与基本原理
 - 2.2.1 量化与向量量化原理
 - 2.2.2 统计编码原理
 - 2.2.3 Huffman编码
 - 2.2.4 算术编码
 - 2.2.5 预测编码
 - 2.2.6 变换编码
- ◆ 2.3 音频压缩标准
- ◆ 2.4 静态图像压缩编码的国际标准
- ◆ 2.5 视频压缩的国际标准
- ◆ 2.6 可伸缩性编码和分布式编码



Huffman 编码处理不好的情况

- ◆ Huffman 编码使用整数个二进制位对符号进行编码，这种方法在许多情况下无法得到最优的压缩效果。假设某个字符的出现概率为 80%，该字符事实上只需要
- ◆ $-\log_2(0.8) = 0.322$ 位编码，但 Huffman 编码一定会为其分配一位 0 或一位 1 的编码。可以想象，整个信息的 80% 在压缩后都几乎相当于理想长度的 3 倍左右，压缩效果可想而知。

MC 算数编码的原理

◆ 算术编码方法是将被编码的信息表示成实数**0和1之间的一个间隔**。信息越长编码表示它的间隔就越小，表示这一间隔所需二进制位就越多，大概率符号出现的概率越大对应于区间愈宽，可用长度较短的码字表示；小概率符号出现概率越小区间愈窄，需要较长码字表示。

◆ 信息源中连续的符号根据**某一模式生成概率的大小来减少间隔**。可能出现的符号要比不太可能出现的符号减少范围少，因此只增加了较少的比特位。

MC 算术编码示例

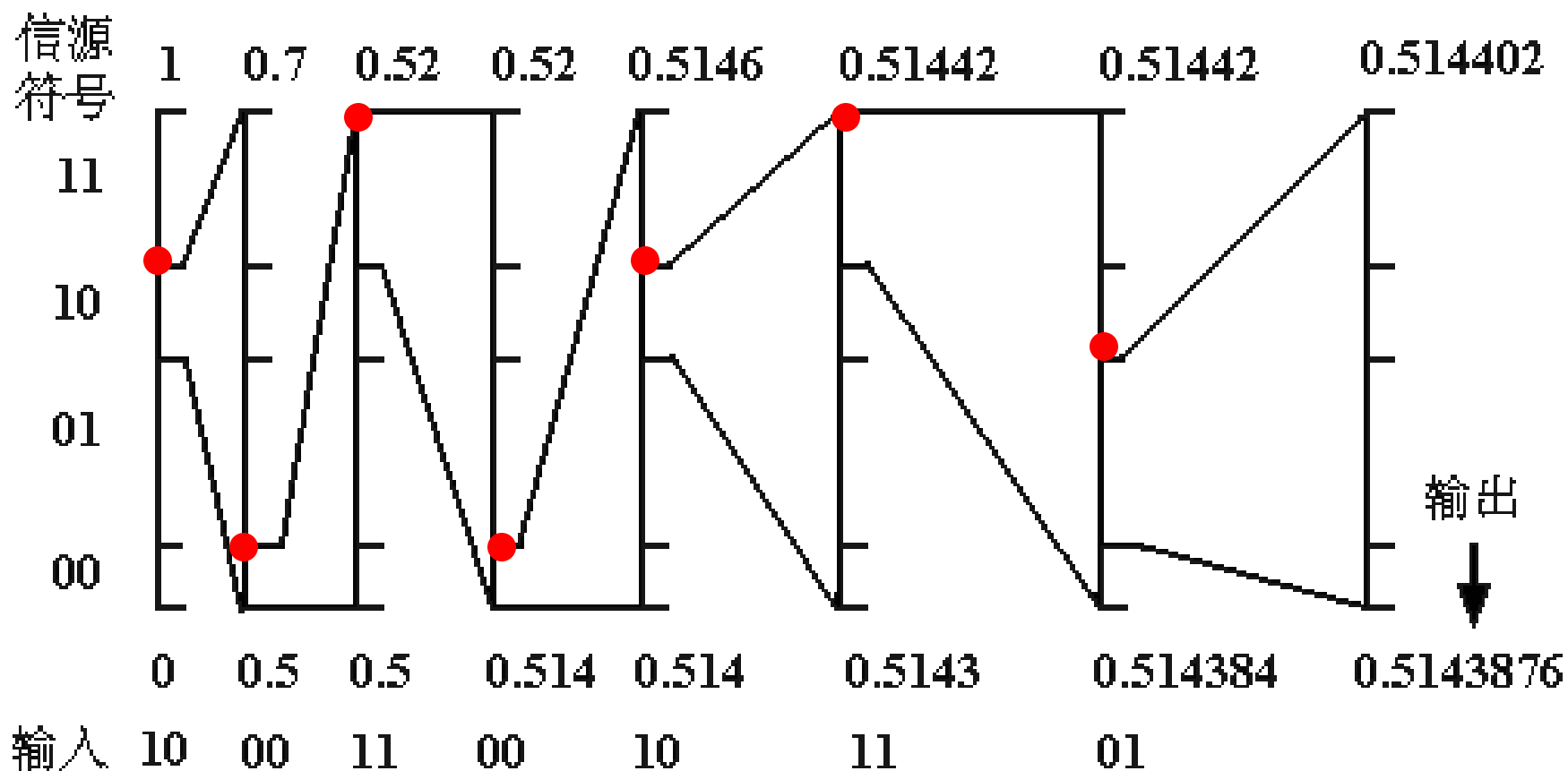
假设信源符号为 {00, 01, 10, 11}，这些符号的概率分别为 { 0.1, 0.4, 0.2, 0.3 }，根据这些概率可把间隔 [0, 1) 分成 4 个子间隔： [0, 0.1), [0.1, 0.5), [0.5, 0.7), [0.7, 1)，二进制消息序列的输入为： 10 00 11 00 10 11 01

符号	00	01	10	11
概率	0.1	0.4	0.2	0.3
初始编码间隔	[0, 0.1)	[0.1, 0.5)	[0.5, 0.7)	[0.7, 1)



算术编码过程

编码间隔



MC 算数编码优缺点

- ◆ 算术编码对整条信息（无论信息有多么长），其输出仅仅是一个数，而且是一个介于 0 和 1 之间的二进制小数。
- ◆ 例如算术编码对某条信息的输出为 1010001111，那么它表示小数 0.1010001111，也即十进制数 0.64（0.6396484375）。
- ◆ 若符号数为 2 的整数次幂时，信源符号序列直接表示为定点小数！
- ◆ 问题：计算机的有限字长



第2章 多媒体数据压缩国际标准

- ◆ 2.1 多媒体数据压缩编码的重要性和分类
- ◆ 2.2 常见数据压缩方法分类与基本原理
 - 2.2.1 量化与向量量化原理
 - 2.2.2 统计编码原理
 - 2.2.3 Huffman编码
 - 2.2.4 算术编码
 - 2.2.5 预测编码
 - 2.2.6 变换编码
- ◆ 2.3 音频压缩标准
- ◆ 2.4 静态图像压缩编码的国际标准
- ◆ 2.5 视频压缩的国际标准
- ◆ 2.6 可伸缩性编码和分布式编码

MC 预测编码基本原理

◆ 预测编码是数据压缩理论的一个重要分支。它根据离散信号之间存在一定相关性的特点，利用前面的一个或多个信号对下一个信号进行预测，然后对实际值和预测值的差(预测误差)进行编码。如果预测比较准确，那么误差信号就会很小，就可以用较少的码位进行编码，以达到数据压缩的目的。

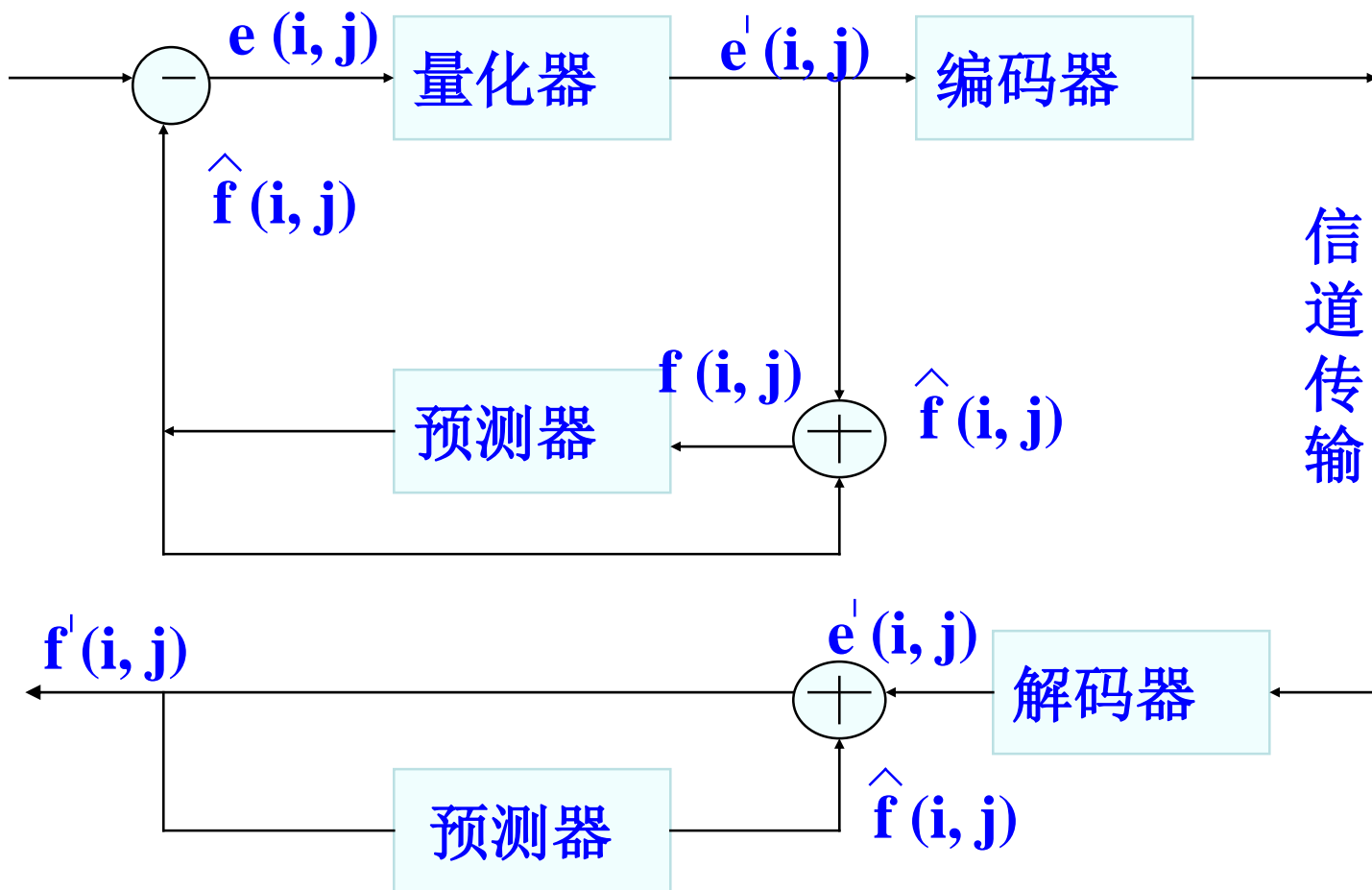
◆ 第n个符号 X_n 的熵满足：

$$H(x_n) \geq H(x_n | x_{n-1}) \geq H(x_n | x_{n-1}x_{n-2}) \geq \dots \geq H(x_n | x_{n-1}x_{n-2}\dots x_1)$$

◆ 所以参与预测的符号越多，预测就越准确，该信源的不确定性就越小，数据码率就可以降低。



DPCM 差分脉冲编码调制编解码



MC 预测方程式

◆ 线性预测:

$$x'_k = f(x_1, x_2, x_3, \dots, x_{k-1}, k)$$

◆ 如果 a_i 是常数, 则为时不变线性预测 (DPCM), 否则为自适应线性预测 (ADPCM)

$$x'_k = \sum_{i=1}^{k-1} a_i(k) x_i$$

◆ 最简单的预测方程:

$$x'_k = x_{k-1}$$

MC 最佳线性预测

- ◆ 使误差函数 $mse = E[(x_n - x'_n)^2]$
- ◆ 达到最小值的预测方程式叫做最佳线性预测。
- ◆ 求最佳线性预测的各个参数 a_i ，列方程组：

$$\frac{\partial E[(x_n - x'_n)^2]}{\partial a_i} = 0, (i = 1, 2, \dots, n - 1)$$

- ◆ 代入 $x'_n = \sum_{i=1}^{n-1} a_i x_i$ 得到联立方程组：

$$E[x_n x_i] = \sum_{l=1}^{n-1} a_l E[x_l x_i], (i = 1, 2, \dots, n - 1)$$

- ◆ 如果为一阶线性预测，则可求得：

$$a_1 = \frac{E[x_n x_{n-1}]}{E[x_{n-1}^2]} \quad x'_n = a_1 x_{n-1}$$

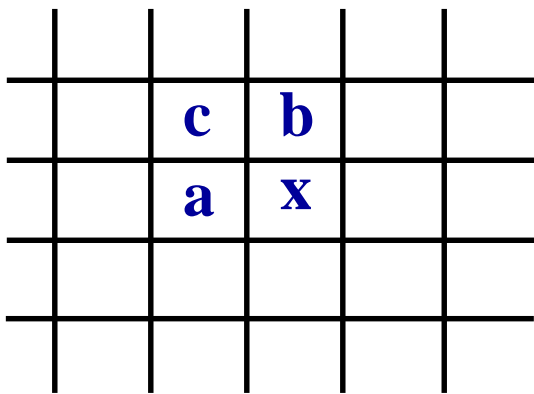


图像信号的预测编码

- ◆ 一幅数字图像可以看成是一个空间点阵，图像信号不仅在水平方向是相关的，在垂直方向也是相关的。根据已知样值与待预测样值间的位置关系，可以分为：
 - 一维预测（**行内预测**）：利用同一行上相邻的样值进行预测。
 - 二维预测（**帧内预测**）：利用同一行和前面几行的数据进行预测。
 - 三维预测（**帧间预测**）：利用相邻几帧（或不同波段）上的取样值进行预测



静止图像的二维预测编码



三邻域预测法

选择值	预测值
0	非预测
1	a
2	b
3	c
4	$a+b-c$
5	$a+(b-c)/2$
6	$b+(a-c)/2$
7	$(a+b)/2$

◆ 这种压缩算法被应用到JPEG标准的无损压缩模式之中，中等复杂程度的图像压缩比可达到2:1。



活动图像的帧间预测编码

- ◆ 视频信号的冗余度主要体现在空间相关性（帧内）、时间相关性（帧间）和色度空间表示上的相关性。
- ◆ 对于每秒25帧（30）的电视信号，其相继帧之间存在极强的相关性。**据统计256级灰度的黑白图像序列，帧间差值超过3的像素数不超过4%**。所以在活动图像序列中可以利用前面的帧来预测后面的帧，以实现数据压缩。
- ◆ 帧间预测编码技术被广泛应用到H.261、H.263、MPEG-1和MPEG-2等视频压缩标准之中



具有运动补偿的帧间预测：原理

- ◆活动图像序列中的一个画面可以大致划分为3个区域：
 - (1) **背景区**：相邻两个画面的背景区基本相同。
 - (2) **运动物体区**：可以视为由前一个画面的某一区域的像素平移而成。（位移矢量）
 - (3) **暴露区**：指物体运动后而显露出来的曾被遮盖的背景区域。
- ◆运动补偿预测就是将**前一个画面的背景区+平移后的运动物体区**作为后一个画面的预测值。

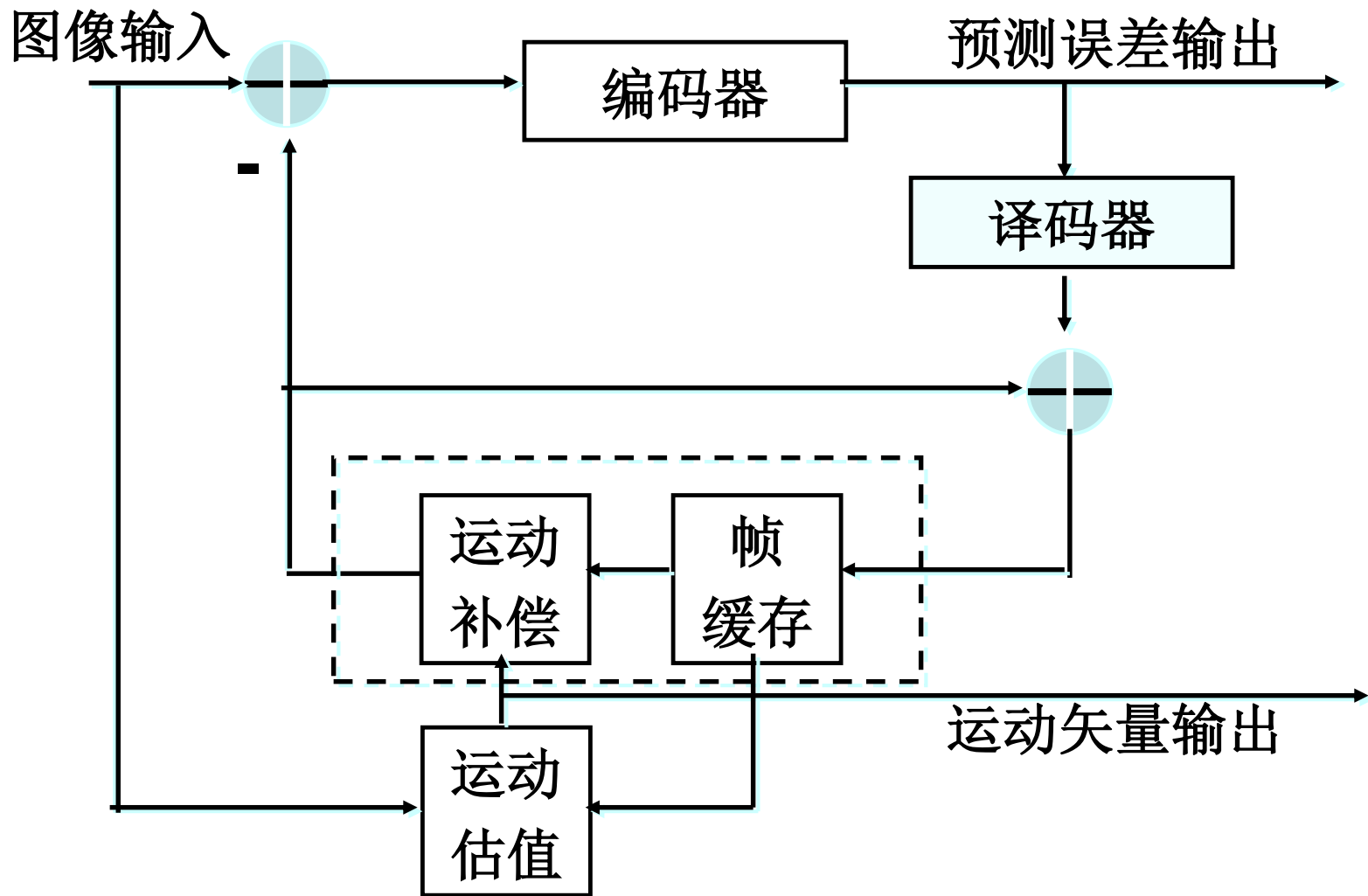


运动补偿帧间预测编码：步骤

- ◆ 运动补偿帧间预测从原理上包括如下几个基本步骤：
 - (1) **图像分割**：把图像划分为静止的背景和若干运动的物体，各个物体可能有不同的位移，但构成同一物体的所有像素的位移相同。
 - (2) **运动矢量估值**：考察前后两个画面，利用运动估值算法得到每个物体的位移矢量。
 - (3) **运动补偿**：用运动矢量补偿物体的运动效果再进行预测。
 - (4) **编码**：除了对实测值与预测值之间的差值进行编码传送外，还要传送位移矢量以及区域分割信息。



运动补偿帧间预测编码：实现结构





运动补偿帧间预测编码：运动估计

◆运动估计有下述三种方法：

(1)块匹配法（Block Matching Algorithm BMA）以像素块为准进行运动估计。

(2)像素递归法（Pixel Recursive Algorithm PRA）以像素为准进行递归的运动估计。

(3)傅立叶变换法，其原理基于傅立叶变换可以分别处理平移、旋转和尺寸变化的特性。



第2章 多媒体数据压缩国际标准

- ◆ 2.1 多媒体数据压缩编码的重要性和分类
- ◆ 2.2 常见数据压缩方法分类与基本原理
 - 2.2.1 量化与向量量化原理
 - 2.2.2 统计编码原理
 - 2.2.3 Huffman编码
 - 2.2.4 算术编码
 - 2.2.5 预测编码
 - 2.2.6 变换编码
- ◆ 2.3 音频压缩标准
- ◆ 2.4 静态图像压缩编码的国际标准
- ◆ 2.5 视频压缩的国际标准
- ◆ 2.6 可伸缩性编码和分布式编码



变换编码原理：正交变换

- ◆ 设：原始图像为 X ，变换矩阵为 T ，均为方阵
- ◆ 则： $Y = T * X$ ， Y 就是变换后的图像
- ◆ 若变换矩阵 T 满足： $T' * T = I$ ， $T^{-1} * T = I$ ，则变换 T 称为正交变换。
- ◆ 由 Y 还原出原始图像 X 的方法是： $X = T' * Y$

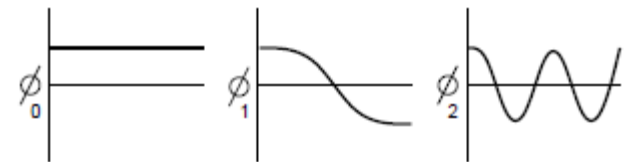
MC 变换编码：基函数

◆ DCT变换的变换矩阵T

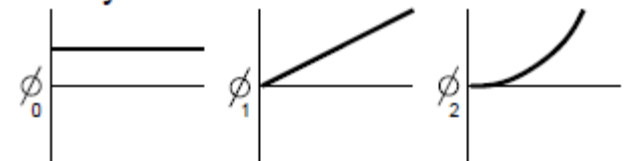
$$T_{ij} = \begin{cases} \sqrt{1/n} \cos \frac{(2j+1)i\pi}{2n} & i = 0, 0 \leq j < n \\ \sqrt{2/n} \cos \frac{(2j+1)i\pi}{2n} & 0 < i < n, 0 \leq j < n \end{cases}$$

the first three basis functions
for discrete cosine, polynomial, and
wavelet transformations

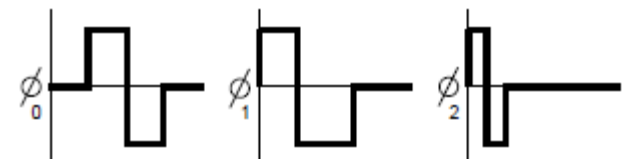
Cosine



Polynomial



Wavelet





最佳的正交变换：K-L变换

◆ K-L变换实质上是作坐标系的转换，尽量让向量落在最少的坐标轴上或其周围，从而只用较少的变换系数就可以恢复出质量不错的图像，压缩效率比较高，均方误差小。

◆ 缺点：图像变化后变换核矩阵也要相应变换，求解过程比较复杂，没有快捷方法。

Michel Loève



Born January 22, 1907
Jaffa, Palestine, Ottoman Syria

Died February 17, 1979
(aged 72)
Berkeley, California

Nationality French American

Fields Mathematics

Institutions University of California, Berkeley

Alma mater École Polytechnique

Doctoral advisor Paul Pierre Lévy

Doctoral students Julius Blum
Leo Breiman
Robert Cogburn
Stanley Nash
Emanuel Parzen

Known for Karhunen–Loève theorem

Kari Karhunen

Born	1915
Died	1992
Nationality	Finnish
Fields	Mathematics
Institutions	University of Helsinki
Doctoral advisor	Rolf Nevanlinna
Known for	Karhunen–Loève theorem

MC 离散余弦变换DCT

- ◆ 离散余弦变换(Discrete cosine Transform)简称DCT。任何连续的实对称函数的傅里叶变换中只含余弦项，因此余弦变换与傅里叶变换一样有明确的物理量意义。
- ◆ DCT是先将整体图像分成 $N \times N$ 像素块,然后对 $N \times N$ 像素块逐一进行DCT变换。由于大多数图像的高频分量较小，相应于图像高频成分的系数经常为零，加上人眼对高频成分的失真不太敏感，所以可用更粗的量化，因此传送变换系数所用的数码率要大大小于传送图像像素所用的数码率。到达接收端后再通过反离散余弦变换回到样值，虽然会有一定的失真，但人眼是可以接受的。



正向离散余弦变换(FDCT)

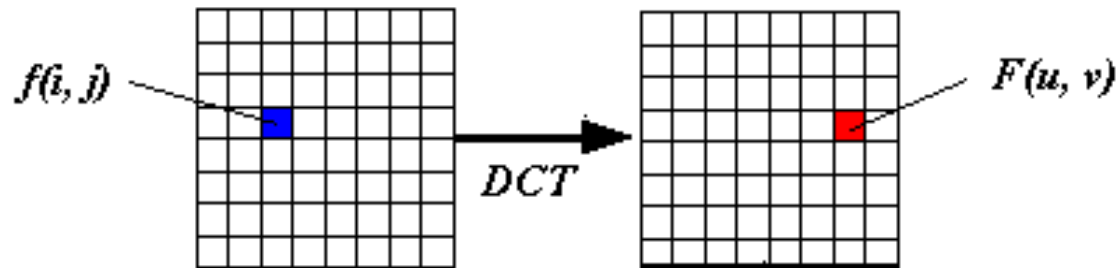
◆对每个单独的彩色图像分量，把整个分量图像分成 8×8 的图像块，作为两维离散余弦变换DCT的输入。通过DCT变换，把能量集中在少数几个系数上。

$$F(u, v) = \frac{1}{4} C(u) C(v) \left[\sum_{i=0}^7 \sum_{j=0}^7 f(i, j) \cos \frac{(2i+1)u\pi}{16} \cos \frac{(2j+1)v\pi}{16} \right]$$

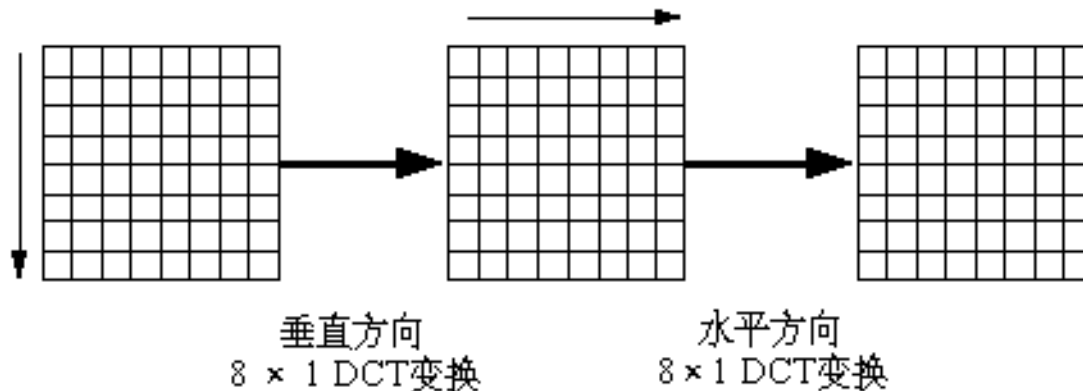
$$f(i, j) = \frac{1}{4} C(u) C(v) \left[\sum_{u=0}^7 \sum_{v=0}^7 F(u, v) \cos \frac{(2i+1)u\pi}{16} \cos \frac{(2j+1)v\pi}{16} \right]$$



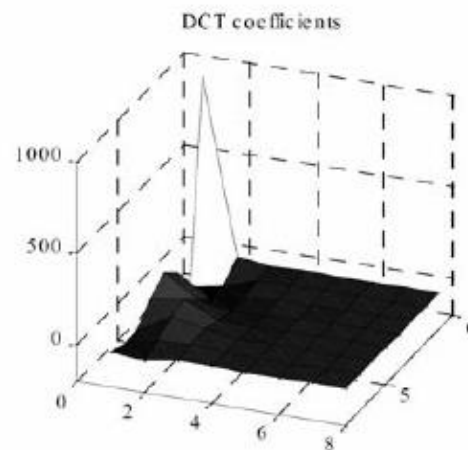
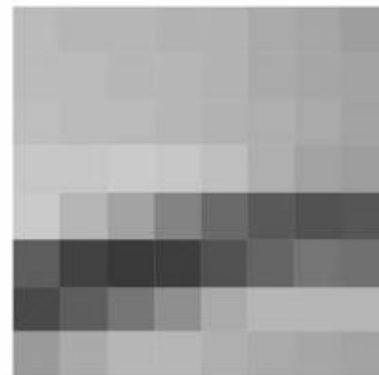
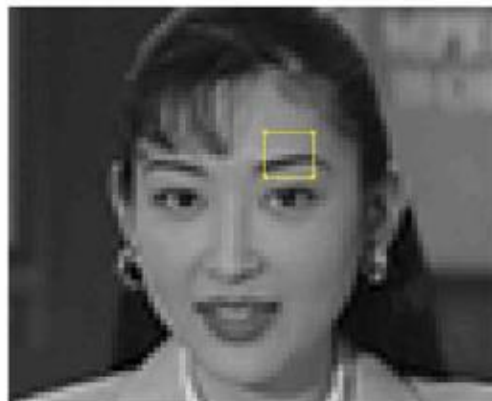
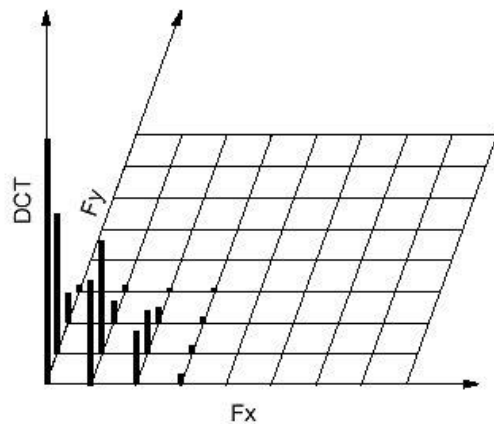
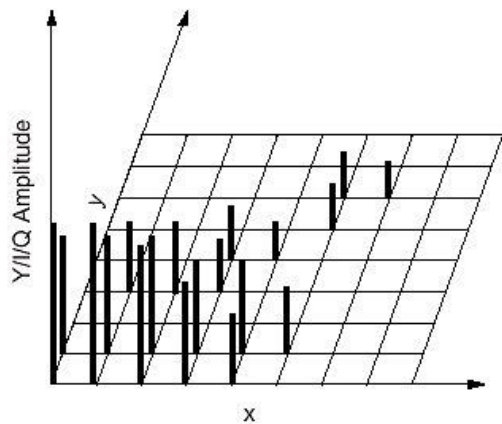
DCT计算



◆ 在计算两维的DCT变换时，可把两维的DCT变换变成一维的DCT变换



DCT结果示例



MC 小结：变换编码

- ◆ 不同的基函数 → 不同的变换编码
- ◆ 正交变换：若变换矩阵 T 满足： $T^T * T = I$ ， $T^{-1} * T = I$ ，则变换 T 称为正交变换。
- ◆ 最佳的正交变换：K-L变换
 - 卡洛南-洛伊变换
 - Karhunen-Loève Transform
 - 无快速算法，且变换矩阵随不同的信号样值集合而不同
- ◆ 多媒体处理中最常见的变换
 - 离散余弦变换(Discrete cosine Transform)简称DCT



小结：数据压缩方法分类与原理

- ◆ 量化与向量量化原理
- ◆ 统计编码原理：熵
 - Huffman编码
- ◆ 算术编码：数字
- ◆ 预测编码：去除相关
- ◆ 变换编码
 - 正交变换、最佳变换、DCT

