



# 数字媒体技术

## 第1章 数字媒体技术概述

刘绍辉 刘贤明

计算机科学与技术学院 哈尔滨工业大学

shliu@hit.edu.cn

2021年春季



# 第1章 多媒体技术概述



## 1.1 多媒体的概念

- 1.1.1 多媒体是什么
- 1.1.2 超文本的概念
- 1.1.3 超媒体的概念
- 1.1.4 多媒体系统的结构
- 1.1.5 示例

## 1.2 多媒体的发展历程

- 1.2.1 多媒体与计算机
- 1.2.2 发展历程
- 1.2.3 最初的多媒体系统
- 1.2.4 多媒体系统关键技术

## 1.3 多媒体数据压缩与编码

- 1.3.1 为什么要压缩
- 1.3.2 两种类型的压缩
- 1.3.3 三种类型的编码
- 1.3.4 压缩与编码

## 1.4 多媒体与光盘

## 1.5 多媒体与网络

- 1.5.1 因特网是什么
- 1.5.2 万维网是什么

## 1.6 多媒体国际标准

- 1.6.1 国际电信联盟(ITU)标准
- 1.6.2 ISO/IEC标准
- 1.6.3 图像, 声音和视频编码标准
- 1.6.4 因特网技术标准
- 1.6.5 万维网协会(W3C)

## 1.7 多媒体内容处理

- 1.7.1 内容是什么
- 1.7.2 内容处理
- 1.7.3 内容标记语言
- 1.7.4 多媒体内容检索

## 1.8 迈向信息时代

## 参考文献和站点

# 1.1 多媒体的概念

## ◆多媒体(multimedia)

- **Multi:** 多
- **Media:** 媒体, 为Medium一词的复数
  - ✓ 一种传播和表达信息的方法
  - ✓ 水, 空气?
- 通过使用不同的准则在感知、表达、陈述、存储、传输和信息交换媒体中加以区分
  - ✓ 感知媒体(The Perception Medium)
  - ✓ 表示媒体(The Representation Medium)
  - ✓ 显示媒体(The Presentation/Display Medium)
  - ✓ 存储媒体(The Storage Medium)
  - ✓ 传输媒体(The Transmission Medium)
  - ✓ 信息交换媒体(The Info. Exchange Medium)



# 1.1 多媒体的概念(续1)

## ◆多媒体(multimedia)

➤ 融合两种或两种以上媒体的人-机互动的信息交流和传播媒体

### ● 多媒体定义

多媒体技术是利用计算机对文字、图像、图形、动画、音频、视频等多种数字信息进行综合处理、建立逻辑关系和人机交互作用的产物。



# 1.1 多媒体的概念(续2)

## ◆超文本(hypertext)

- 包含指向其他文档或文档元素的指针的电子文档。传统文本以线性方式组织，超文本以非线性方式组织

## ◆超媒体(hypermedia)

- 由文字、声音、图形、图像或电视等媒体元素组成相互关联的超文本系统，用户可方便地浏览与主题相关的内容；提供一种符合人类思维习惯的工作和学习环境
- 是以Web系统为基础的通过超链接组织在一起的全球多媒体信息系统。

## ◆社会媒体(Social Media)

- the social interaction among people in which they create, share or exchange information and ideas in virtual communities and networks--Wikipedia





# 1.1 社会媒体



## ◆ 社会媒体(Social Media)



# 1.1 社交媒体(续1)



passively receive textual data

- ◆ Portal websites ;
- ◆ Information is edited by professional editors, and for passive users.
- ◆ Constrained by the network conditions and capturing devices, **textual information dominates.**



actively obtain textual and little visual data

- ◆ Search engine, forums ;
- ◆ Users mainly contribute to textual data;
- ◆ Constrained by the network conditions and capturing devices, **textual information dominates.**



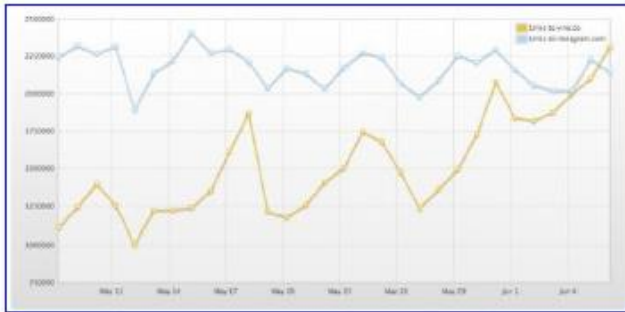
actively contribute, obtain and propagate rich multimedia data

- ◆ Micro-blogging, multimedia sharing websites, SNS ;
- ◆ User-contributed becomes the main mechanism for data generation;
- ◆ Development in capturing devices and network transmission, **huge multimedia data are produced and consumed.**



Internet media roadmap

# 1.1 社交媒体(续1)



Twitter increases its support on multimedia content, and releases the 6-second video sharing app: Vine.



WeChat has attracted more than 300 million users in two years, which is tending to replace SMS.



actively contribute, obtain and propagate rich multimedia data

- ◆ Micro-blogging, multimedia sharing websites, SNS ;
- ◆ User-contributed becomes the main mechanism for data generation;
- ◆ Development in capturing devices and network transmission, huge multimedia data are produced and consumed.

People are getting used to get access to information in the form of multimedia data.

capturing devices, textual information dominates.



Internet media roadmap



# 1.1 社交媒体(续2)



350 million photos are uploaded **daily** in November 2013 on **facebook**



image tweet



1.4 million minutes of chats are produced **every minute** on **skype**



audio photo



100 hour videos are uploaded **every minute**, resulting in 2 billion videos totally by the end of 2013 on **You Tube**



geo-tagged video



Internet Protocol

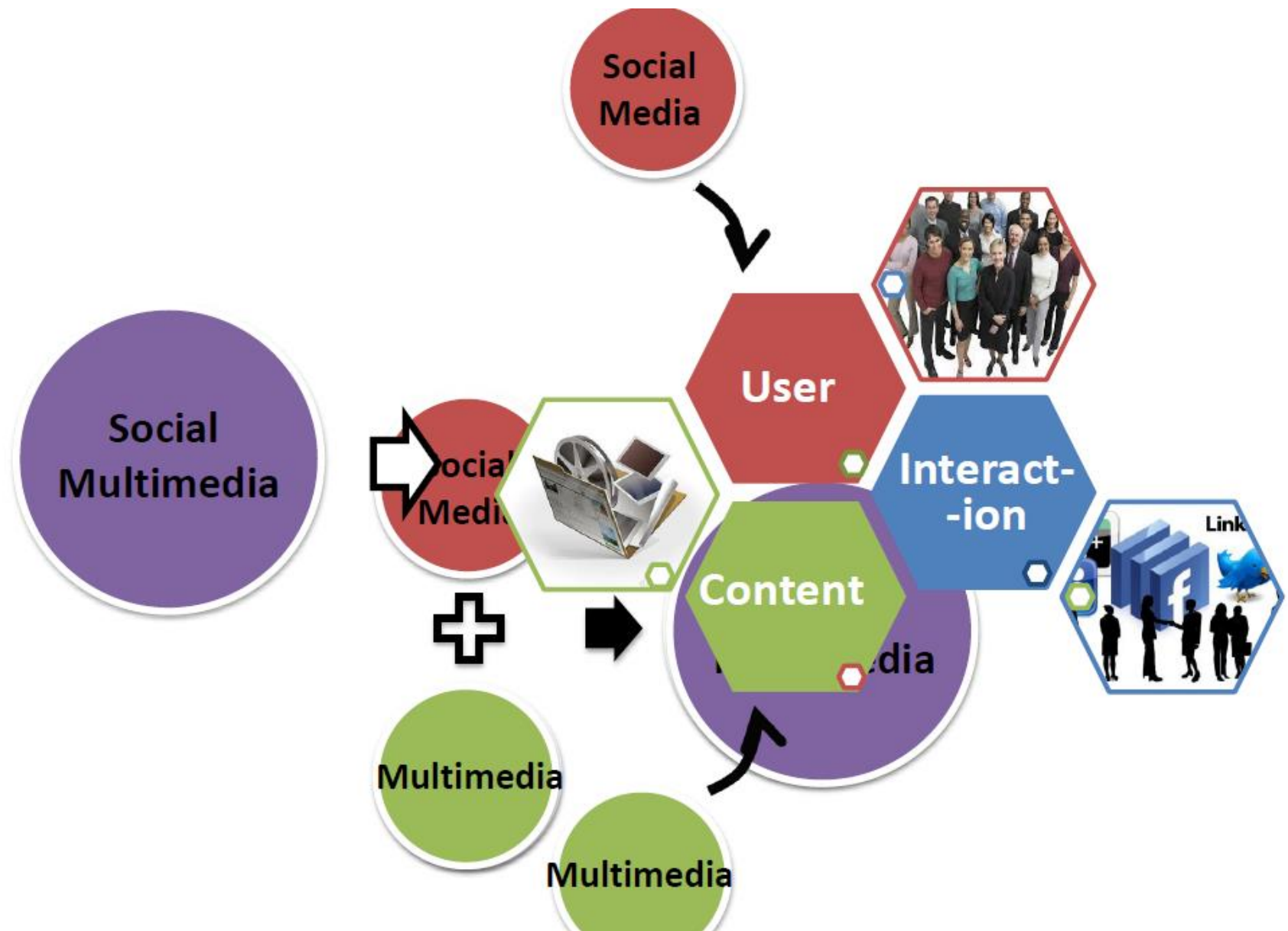
# 1.1 社交媒体(续3)

## ◆ 社交媒体(Social Media)

- the **social interaction** among people in which they **create**, **share** or **exchange** information and ideas in virtual communities and networks—Wikipedia
- “An online source of multimedia resources that fosters an environment of significant **individual participation** and that promotes **community curation, discussion** and **re-use** of content.” -----*Mor Naaman*

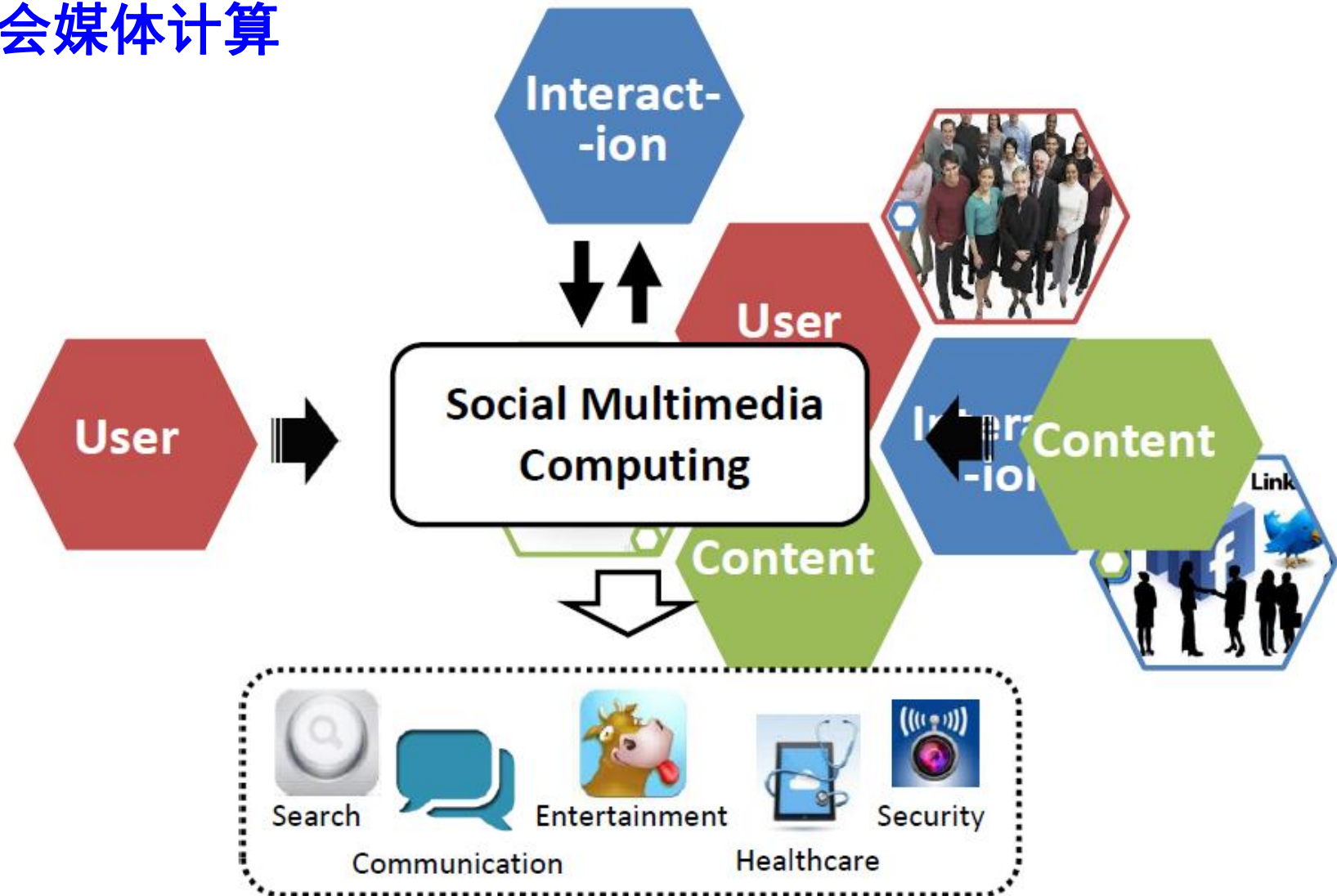


# 1.1 社会多媒体(续4)



# 1.1 社交媒体(续5)

## ◆ 社交媒体计算

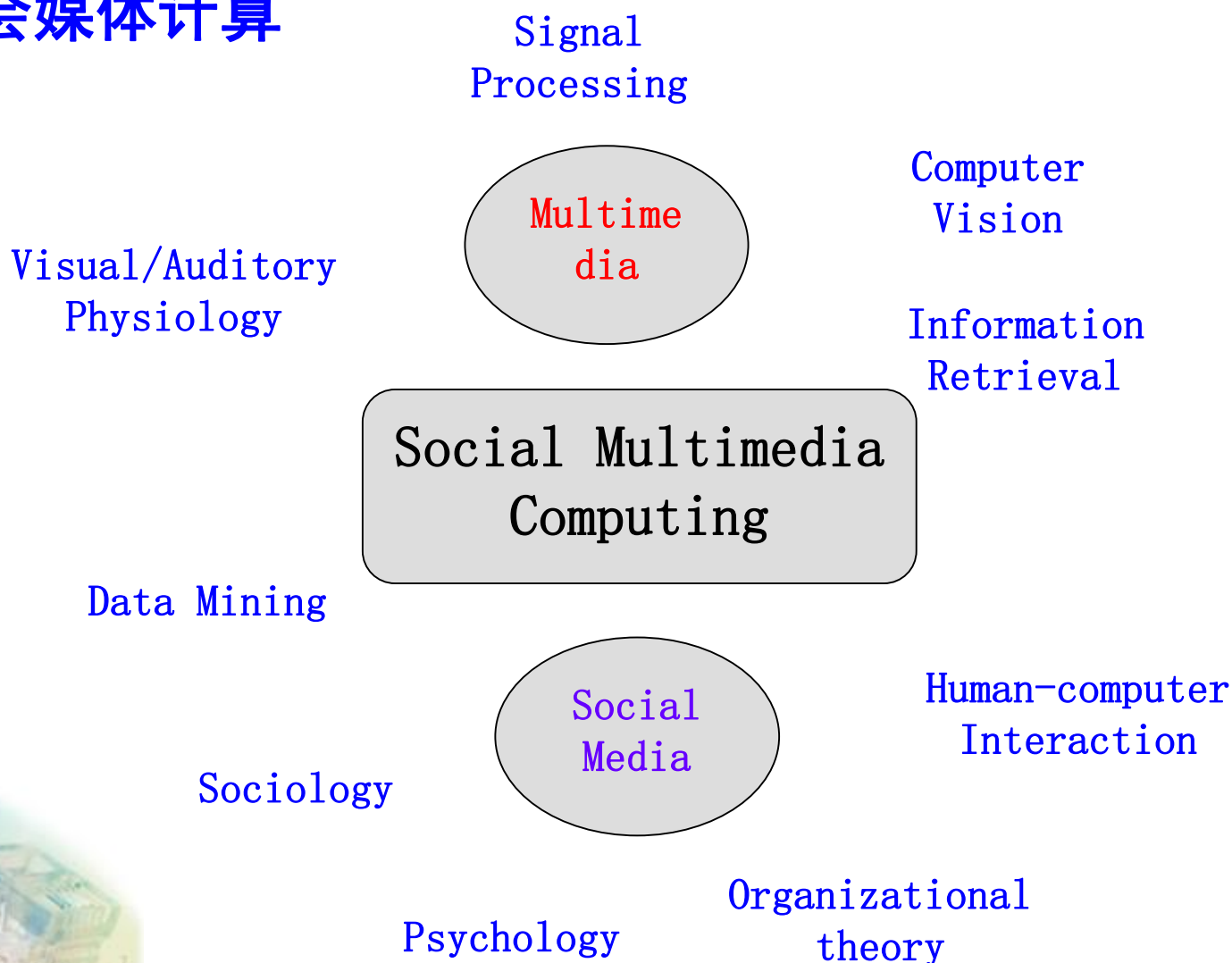




# 1.1 社交媒体(续6)



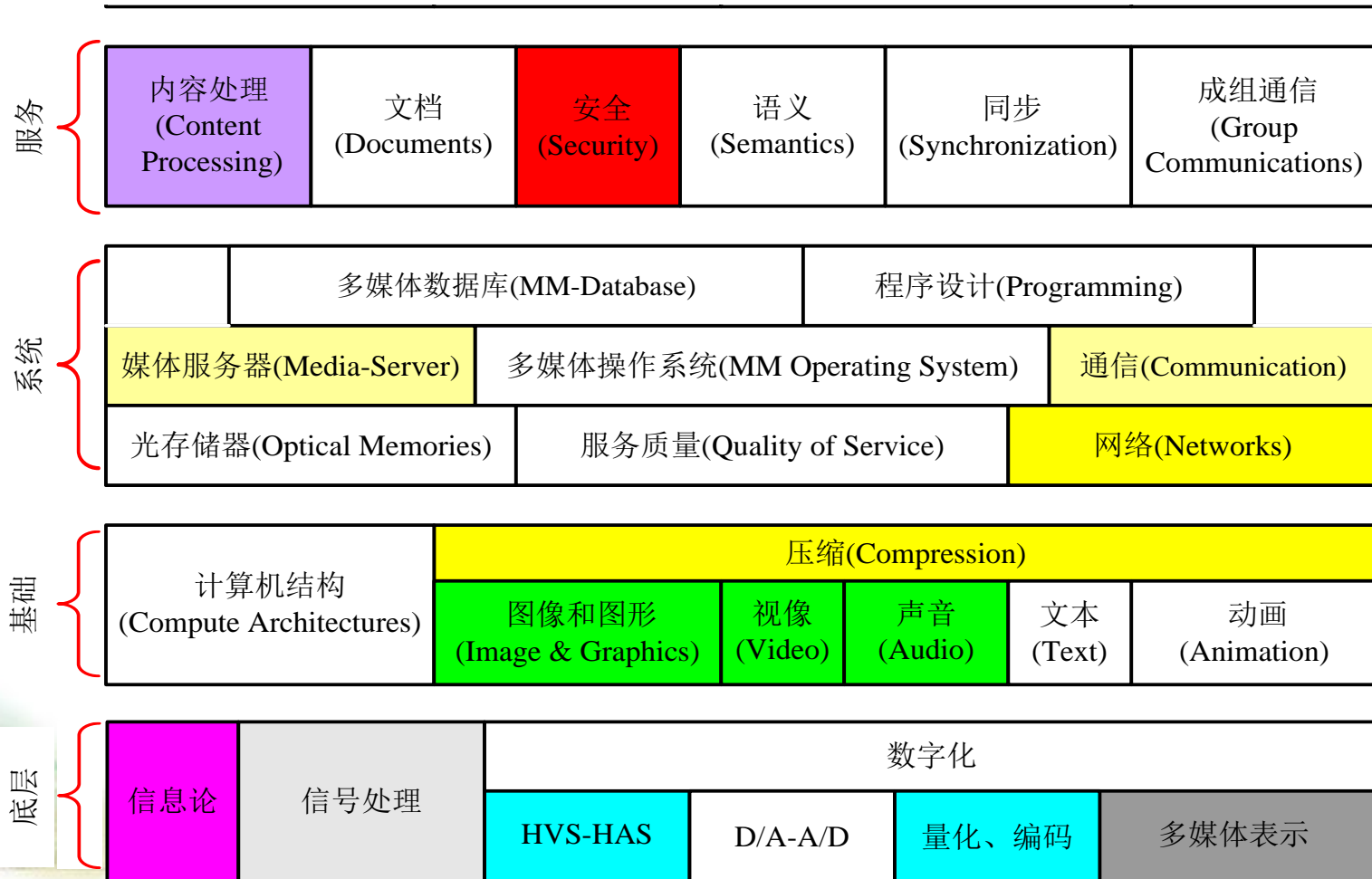
## ◆ 社交媒体计算





# 1.1 多媒体的概念(续7)

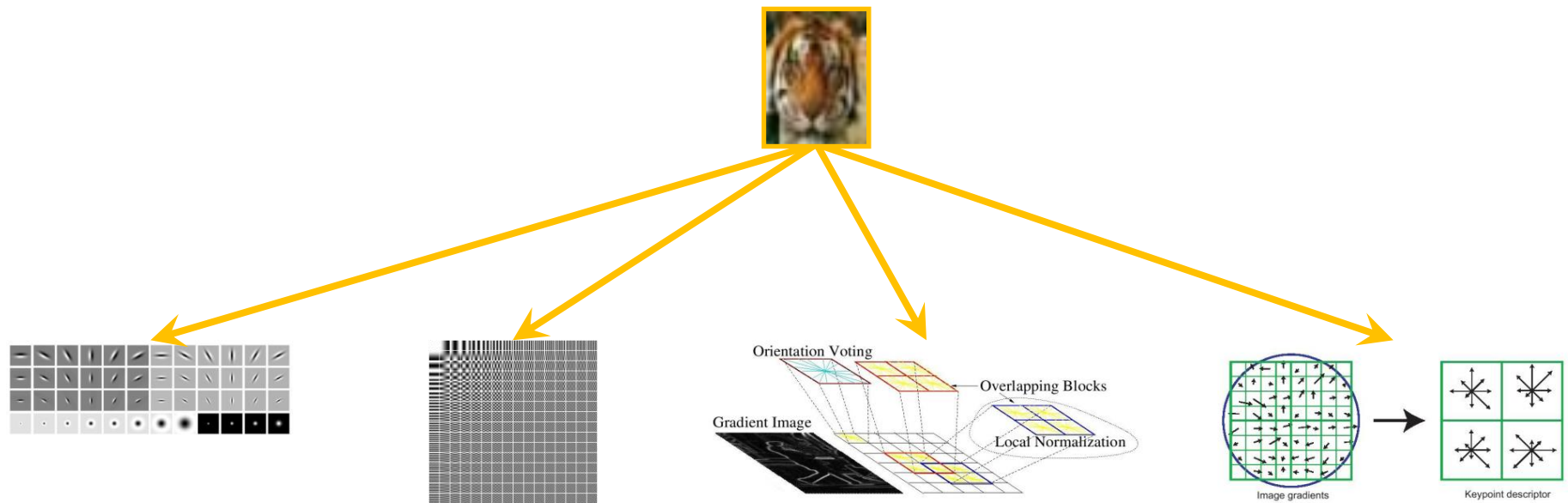
- ◆ 多媒体系统的结构(彩色部分为本课程涉及的内容)  
(引自<http://www.kom.tu-darmstadt.de/mm-book/>)



# 1.1 多媒体的概念-图像视频数据表达的核心问题(注1)

◆问题一：消除冗余

◆问题二：面向多态应用的层次化特征协同表达



通过学习获得的特征  
面向**分类**任务

离散余弦变换特征  
面向**编码**任务

方向梯度特征  
面向**检测**任务

尺度不变性特征  
面向**检索**任务

# 1.1 多媒体的概念(续8)

## ◆信息论

## ◆信号处理

## ◆音视频基础

➤ Human visual/auditory system and perception (HV/AS)

➤ 视频表示 (analog and digital)

➤ 声音表示 (analog and digital)

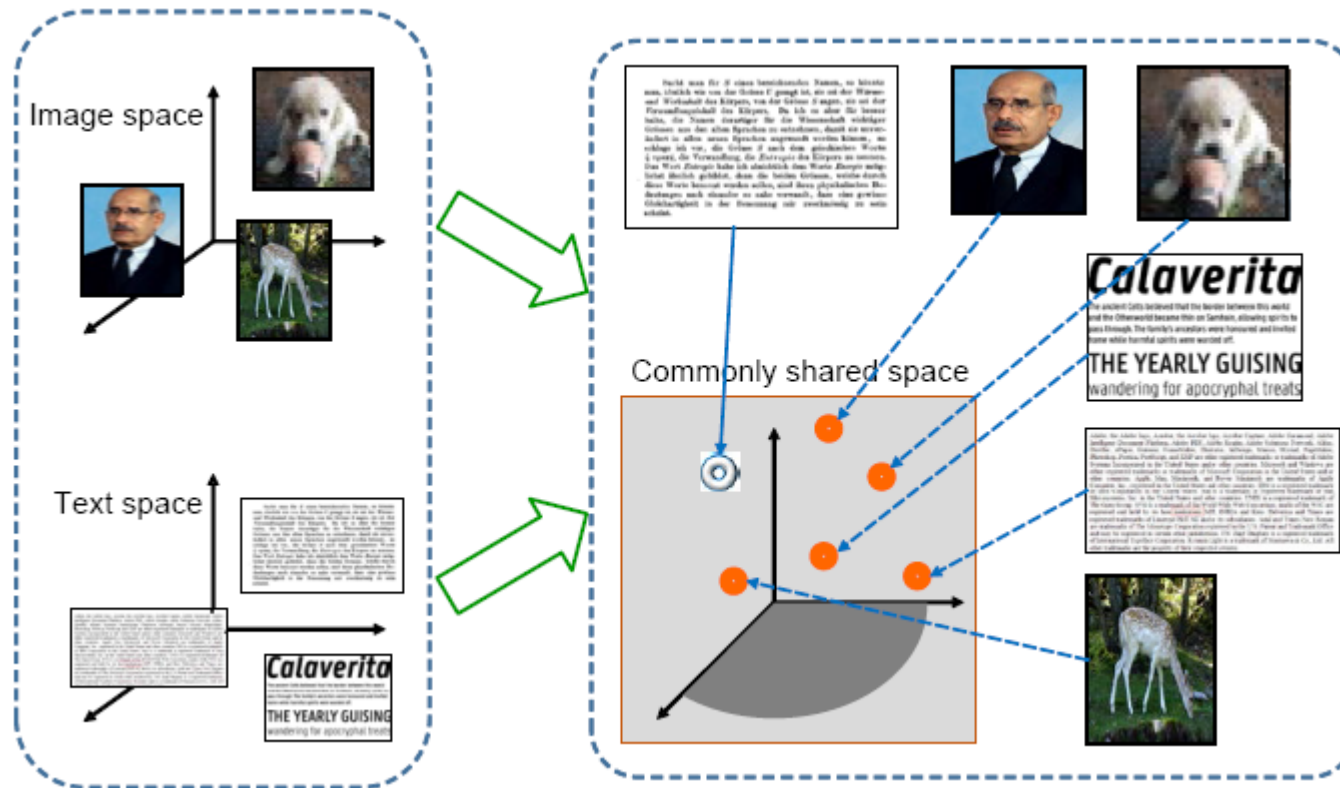
## ◆媒体处理算法

## ◆多媒体无所不含



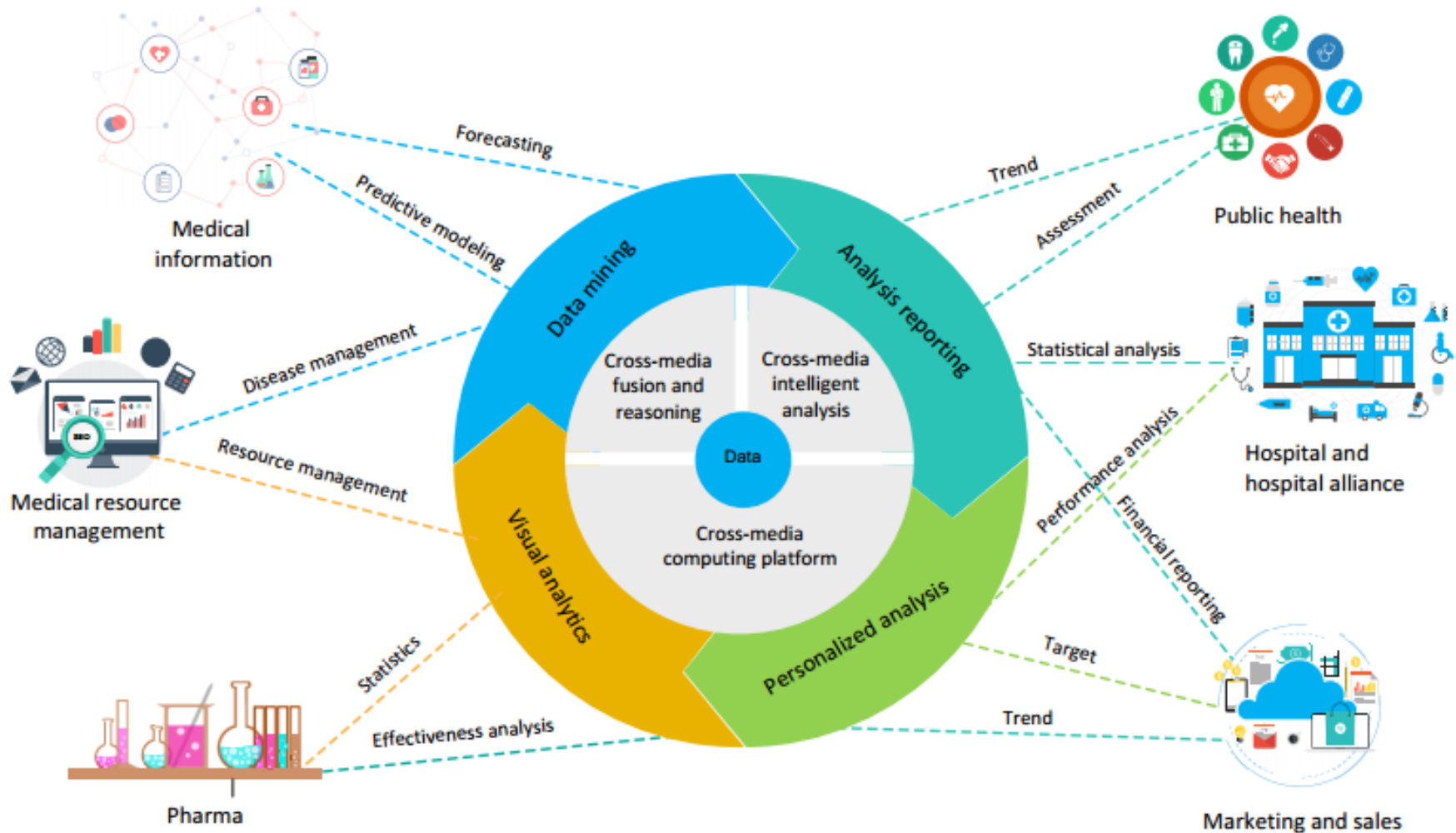
# 1.1 多媒体的概念(续9)

## ◆ 跨媒体(Cross-Media)



# 1.1 多媒体的概念(续9)

## ◆ 跨媒体(Cross-Media)

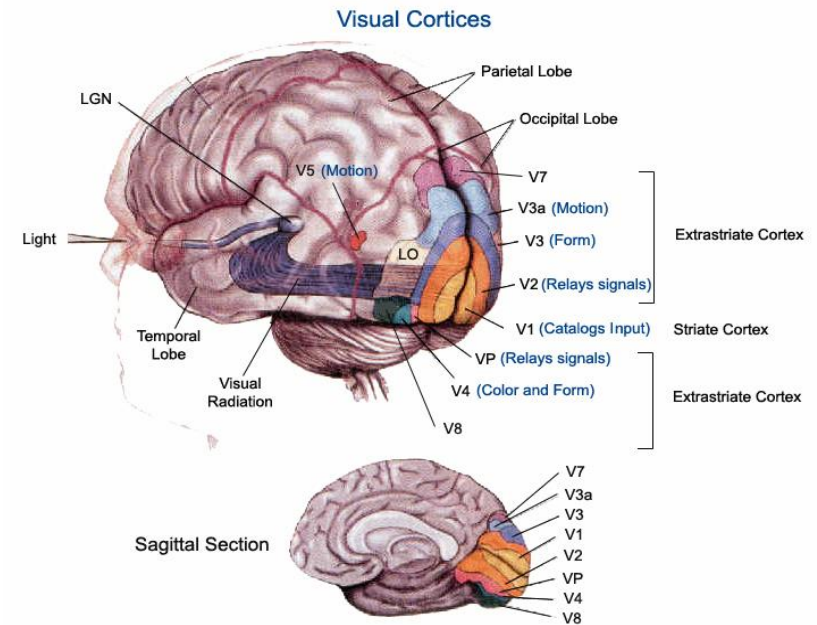
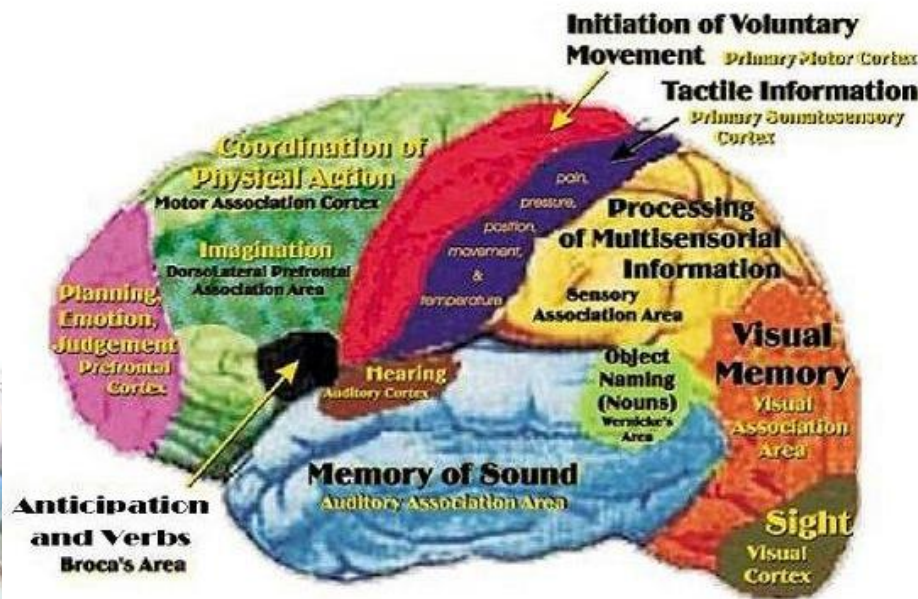




# 1.1多媒体概念-示例-HVS:人类视觉感知机制



- ◆ 人类视觉系统呈现明显的分层结构，通过人眼输入的视觉信号依次经过视觉皮层的V1，V2等区域来触发我们对视觉信息的感知。
- ◆ 大脑皮层中的视觉记忆（Visual Memory）区域会存储学习得到的经验信息，并辅助我们对视觉信号的感知结果进行判断。



# 1.1 多媒体概念-应用事例-无参考质量评价(续1)

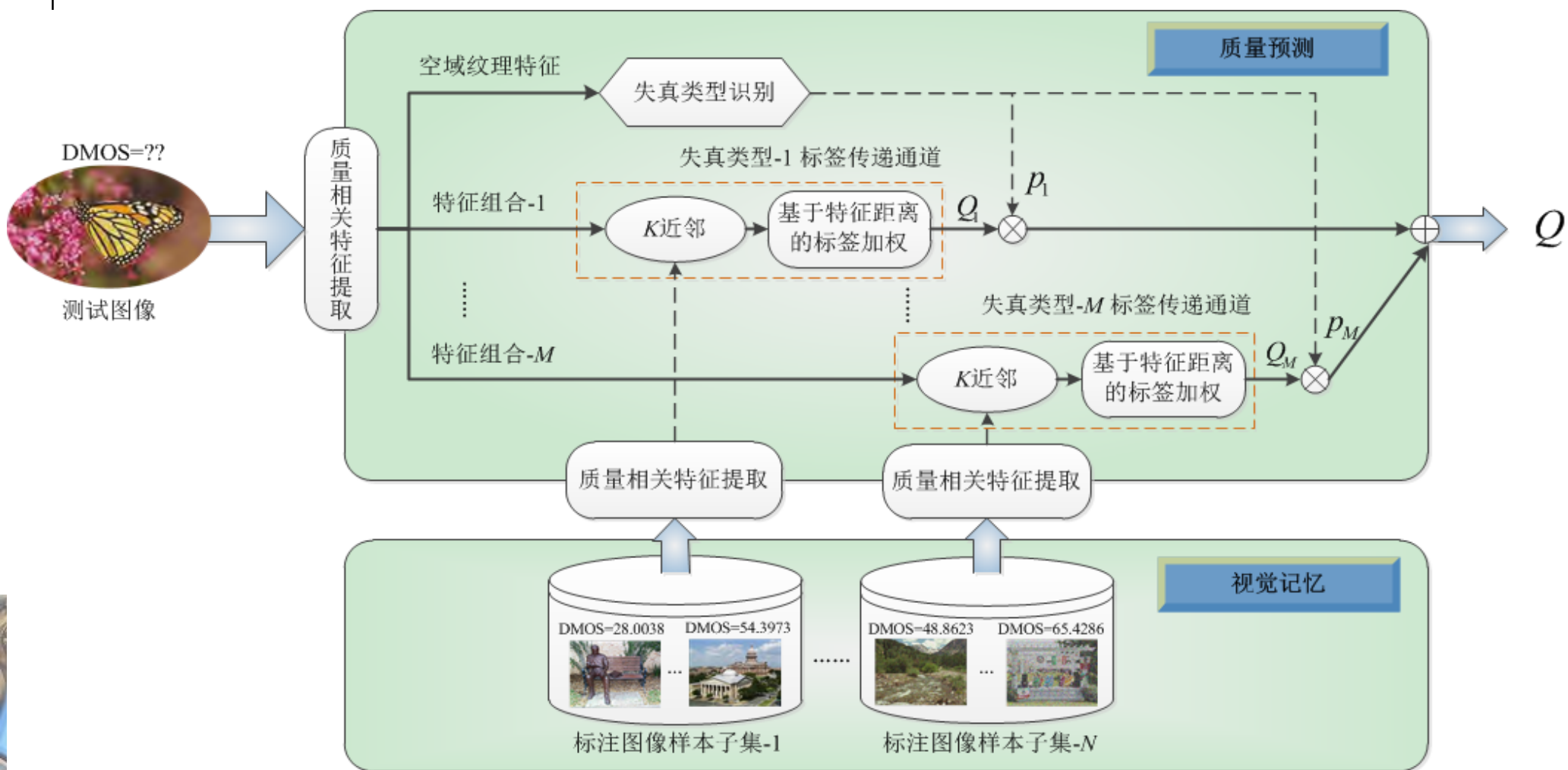
- ◆ 无参考图像质量评价的目标是开发有效的计算模型，在不借助失真图像原始参考信息的情况下预测人眼对失真图像的感知质量。



- ◆ 基于多通道特征融合与标签传递的方法

- ◆ Q. B. Wu, H. L. Li, F. M. Meng, K. N. Ngan, B. Luo, C. Huang, and B. Zeng, "Blind image quality assessment based on multi-channel feature fusion and label transfer," *IEEE Transactions on Circuits and Systems for Video Technology*, 2015

# 1.1 多媒体概念-应用事例-无参考质量评价(续2)



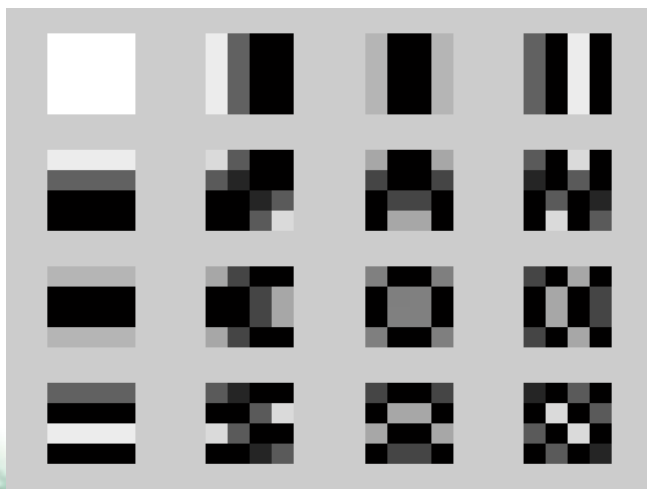
# 1.1 多媒体概念-应用事例-无参考质量评价模拟V1区域 ： DCT特征(续3)



- ◆ V1是人脑中第一层视觉感知皮质区域，并对简单的局部特征变化敏感（如边缘、光栅、空间频率的变化等）。
- ◆ 首先用DCT域上归一化的频带系数 $c_i$  的分布来捕捉图像局部块对不同空间频率的响应。

$$c_i = \sqrt{\frac{1}{N_i} \sum_{x,y \in U_i} f_i^2(x,y)}$$

$$U_i = \{x,y | x+y=i, 0 \leq x < W, 0 \leq y < W\}$$



二维DCT变换基示意图

DC	$f_1(0,1)$	$f_2(0,2)$	$f_3(0,3)$
$f_1(1,0)$	$f_2(1,1)$	$f_3(1,2)$	$f_4(1,3)$
$f_2(2,0)$	$f_3(2,1)$	$f_4(2,2)$	$f_5(2,3)$
$f_3(3,0)$	$f_4(3,1)$	$f_5(3,2)$	$f_6(3,3)$

二维DCT频带分布



# 1.1 多媒体概念-应用事例-无参考质量评价模拟V1区域

## : DCT特征(续4)



- ◆ 提取 3 个 DCT 域统计信息来刻画图像局部空频响应的变化。 $c_i^j$  表示第  $j$  个块内频带为  $I$  的 DCT 系数,  $g_i^j = c_i^j - c_{i+1}^j$  表示近邻频带系数差:

DCT系数偏度分布 $P(S)$ :

$$s^j = \frac{E(C^j - \mu(C^j))^3}{\sigma^3(C^j)} \quad C^j = \{c_1^j, \dots, c_{N_c}^j\}$$
$$P(S) = \text{norm}(\text{hist}(S)) \quad S = \{s^1, \dots, s^{N_b}\}$$

DCT系数频带熵 $E$ :

$$P(C_i) = \text{norm}(\text{hist}(C_i)) \quad C_i = \{c_i^1, \dots, c_i^{N_b}\}$$
$$E = [e_1, e_2, \dots, e_{N_c}] \quad e_i = E[-\log_2(P(C_i))]$$

DCT系数频带差熵 $D$ :

$$P(G_i) = \text{norm}(\text{hist}(G_i)) \quad G_i = \{g_i^1, \dots, g_i^{N_b}\}$$
$$D = [d_1, d_2, \dots, d_{N_c-1}] \quad d_i = E[-\log_2(P(G_i))]$$

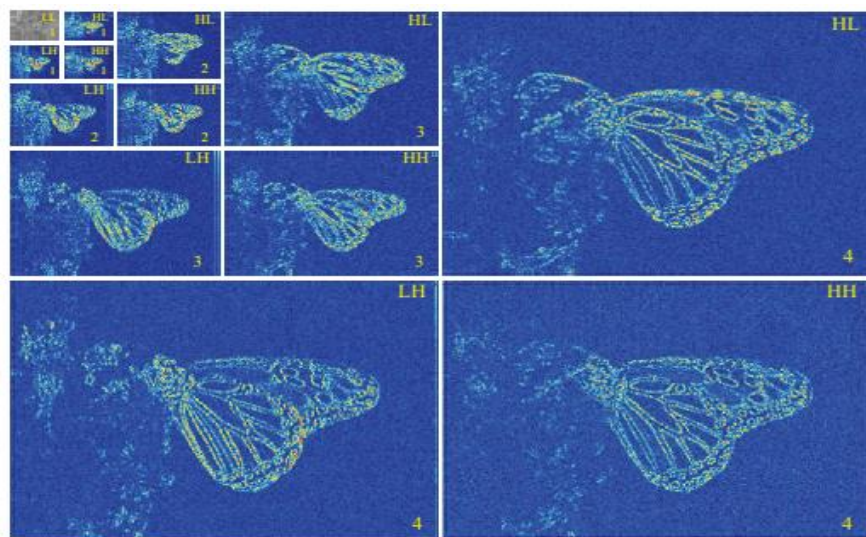




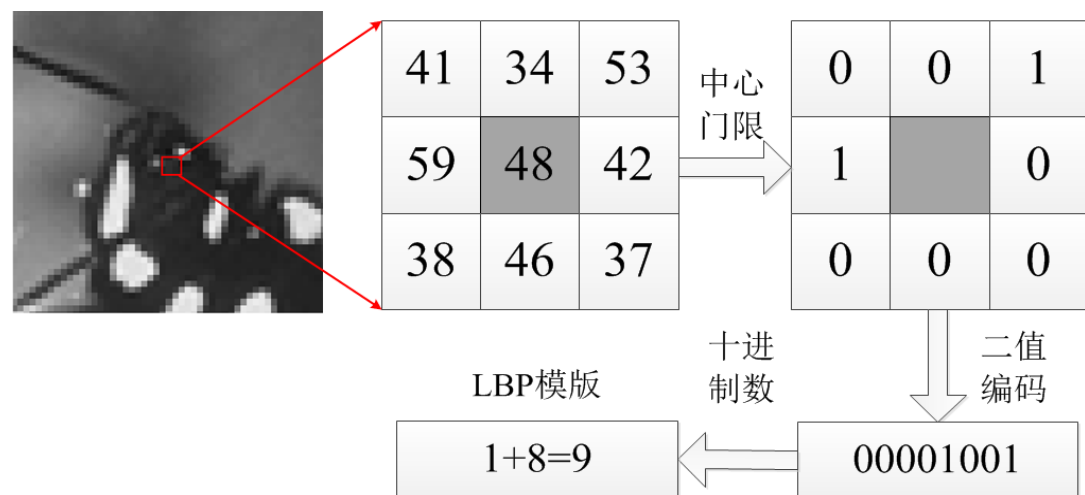
# 1.1 多媒体概念-应用事例-无参考质量评价模拟V2区域 ：小波域和空域特征 (续5)



- ◆ V2位于视觉皮质中的更高层，并在图像表示方面呈现两种特性：（1）V2 区域会捕捉视觉信号中的局部方向性和带通信息；（2）V2区域会对图像中更复杂的位置、尺度不变的形状及纹理信息变化产生敏感反应。
- ◆ 我们从图像的小波域来提取视觉信号的带通和方向性信息；并从二值化的局部空域结构来提取图像纹理信息。



图像小波域结构



图像空域局部二值模版 (LBP)

# 1.1 多媒体概念-应用事例-无参考质量评价模拟V2区域

## : 小波域和空域特征 (续6)



- ◆ 提取 2 个小波域统计特征来刻画图像的带通和方向性信息。让  $\lambda_{k,l}$  表示图像在第  $k$  个方向、第  $l$  个尺度小波子带的系数集合, 其中  $1 \leq l \leq L$  并且  $k = \{1, 2, 3\}$  分别对应 HL, LH 和 HH 三个方向。

小波子带熵 $\hat{E}$ :

$$\hat{e}_{k,l} = E[-\log_2(P(\lambda_{k,l}))] \quad P(\lambda_{k,l}) = \text{norm}(\text{hist}(\lambda_{k,l}))$$
$$\hat{E} = [\hat{e}_{1,1}, \dots, \hat{e}_{1,L}, \dots, \hat{e}_{3,L}]$$

小波子带间KL散度 $\hat{D}$ :

$$\hat{d}_{k,l} = E_{P(\lambda_{k,l+1})} \left[ \log_2 \left( \frac{P(\lambda_{k,l+1})}{P(\lambda_{k,l})} \right) \right]$$
$$\hat{D} = [\hat{d}_{1,1}, \dots, \hat{d}_{1,L-1}, \dots, \hat{d}_{3,L-1}]$$

- ◆ 利用图像空域的局部二值模版的分布来刻画其纹理信息变化。让  $\lambda$  表示图像中所有局部块LBP模版的集合, 则有

空域纹理结构分布LBP:

$$LBP = \text{norm}(\text{hist}(\lambda))$$



# 1.1 多媒体概念-应用事例-无参考质量评价-不同失真类型下特征组合(续7)



- ◆ 人类视觉系统拥有内在的三色属性。因此，从图像的亮度通道  $Y$  和色度通道  $Cb/Cr$  上共同提取感知相关特征，则有

DCT域多通道特征：

$$\mathcal{P} = [P^Y(S), P^{Cb}(S), P^{Cr}(S)]$$

$$\mathcal{E} = [E^Y, E^{Cb}, E^{Cr}]$$

$$\mathcal{D} = [D^Y, D^{Cb}, D^{Cr}]$$

小波域多通道特征：

$$\hat{\mathcal{E}} = [\hat{E}^Y, \hat{E}^{Cb}, \hat{E}^{Cr}]$$

$$\hat{\mathcal{D}} = [\hat{D}^Y, \hat{D}^{Cb}, \hat{D}^{Cr}]$$

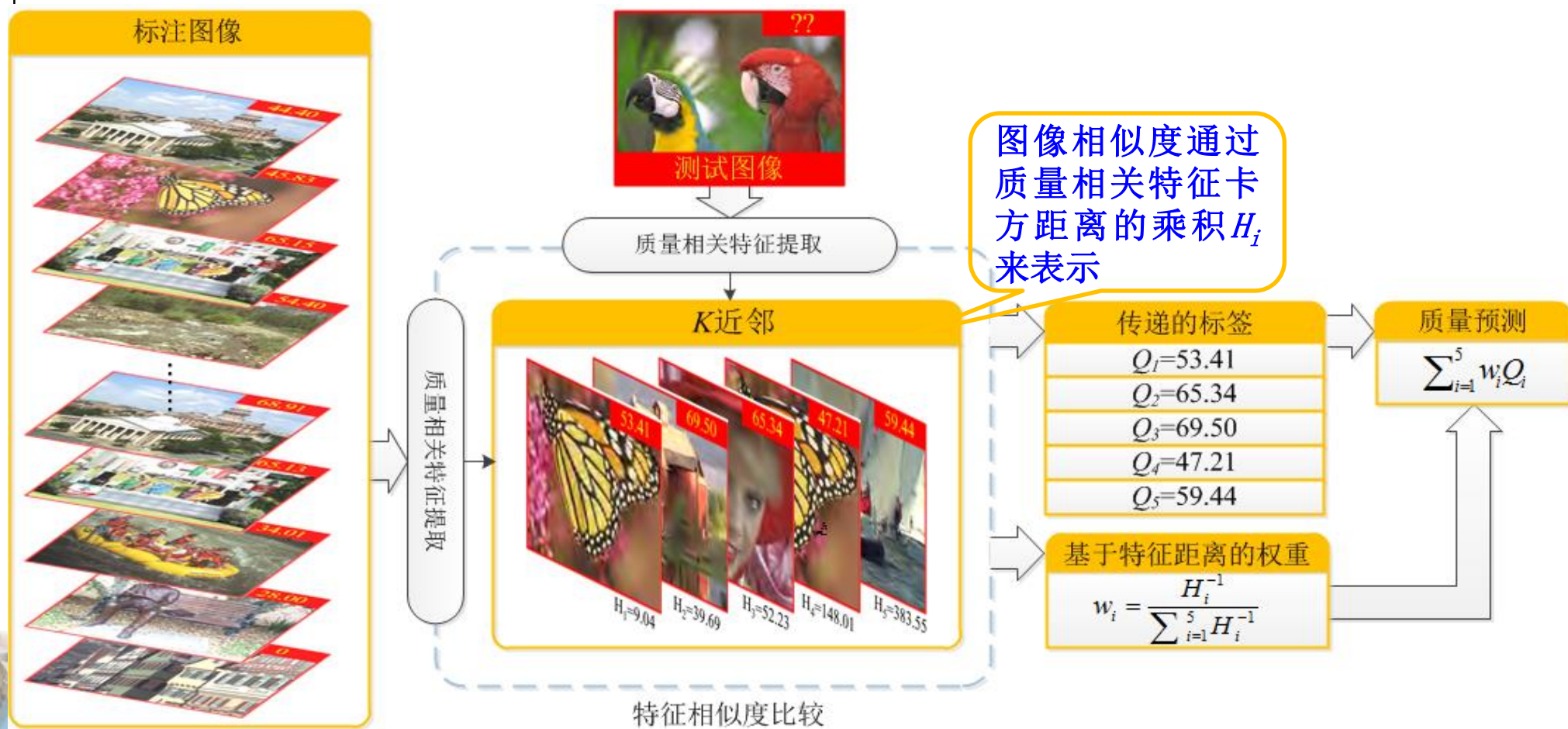
- ◆ 质量相关特征在不同失真类型下表现不同。通过交叉验证来选取每种失真类型下最优的特征组合。

Feature combination ID	Distortion type	Selected features
1	JP2K	$\mathcal{P}, \hat{\mathcal{E}}, \hat{\mathcal{D}}, \text{LBP}$
2	JPEG	$\mathcal{P}, \mathcal{D}, \hat{\mathcal{E}}, \hat{\mathcal{D}}, \text{LBP}$
3	WN	$\mathcal{E}, \hat{\mathcal{E}}, \hat{\mathcal{D}}, \text{LBP}$
4	Blur	$\mathcal{D}, \hat{\mathcal{E}}, \hat{\mathcal{D}}, \text{LBP}$
5	FF	$\mathcal{P}, \hat{\mathcal{E}}, \hat{\mathcal{D}}, \text{LBP}$





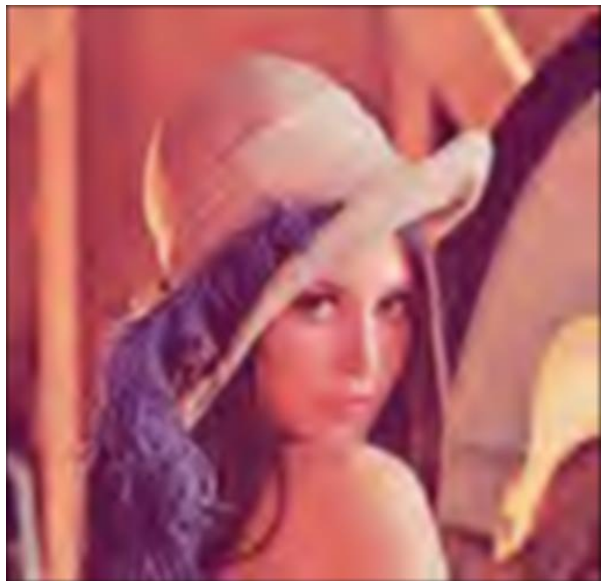
# 1.1 多媒体概念-应用事例-无参考质量评价-基于标签传递的预测模型(续8)



# 1.1 多媒体概念-应用事例-无参考质量评价-实验结果(续9)



- ◆ 我们首先展示了两幅不同类型失真图像预测得到的主观质量打分。其中，预测值越接近0表示感知质量越好，越接近100表示感知质量越差



TCLT-YCbCr预测打分60.4233,  
感知质量较差



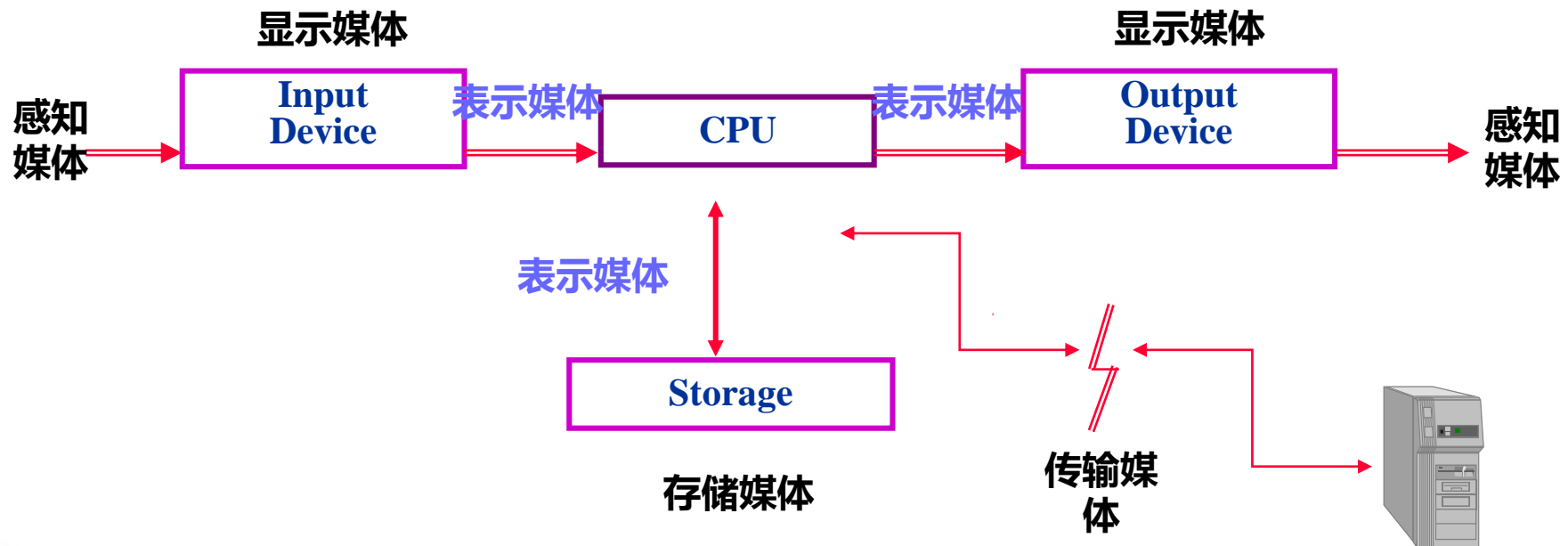
TCLT-YCbCr预测打分33.0116,  
感知质量较好





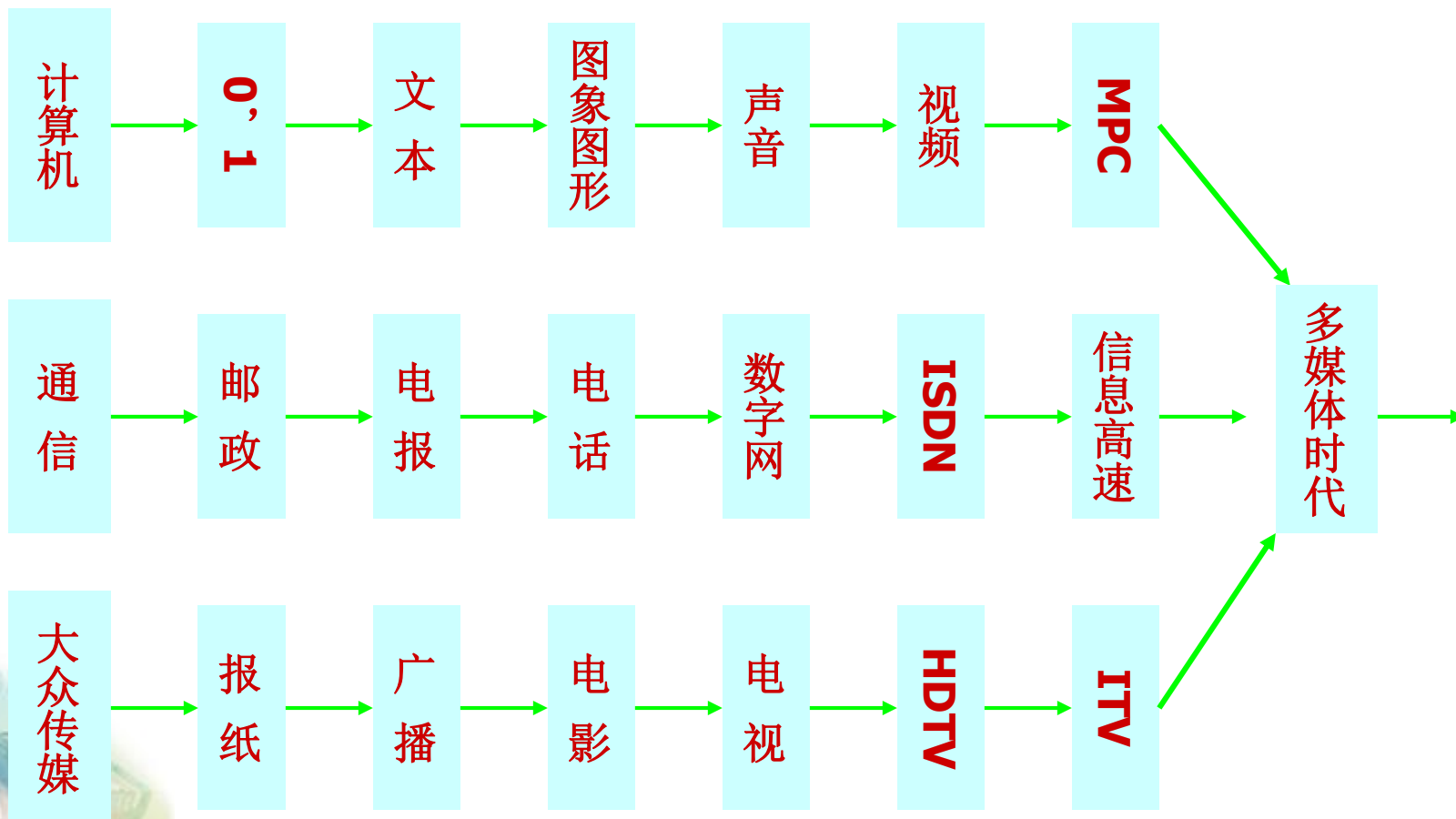
# 1.2 多媒体发展历程

## ◆ 多媒体与计算机



## 1.2 多媒体发展历程(续1)

### ◆ 发展历程



## 1.2 多媒体发展历程(续2)

### ◆最初的多媒体系统

- 最早研究和提出多媒体系统的分别是计算机工业的代表 IBM, Intel, Apple和Commodore公司, 家用电器工业的代表Philips、Sony等公司
- IBM和Intel联合推出的DVI (Digital Video Interactive) , 可使计算机能够处理影像视频信息, 这就使得计算机跨入了传统的电视领域(Compuvision)
- Philips和Sony公司为代表的家用电器工业, 将电视技术进行改进, 使其向智能化、有交互能力方向发展, CD-I (Compact Disc-Interactive) 是他们最早的尝试(Teleputer)
- MPC

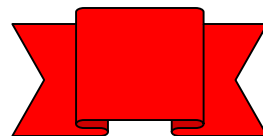
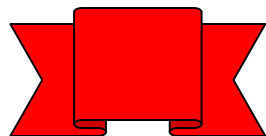


## 1.2 多媒体发展历程(续3)

### ◆MPC

微软总裁比尔·盖茨说：

Window MPC 标准的首要任务是让每个PC用户在硬件和软件的投入与积累得到肯定和连续的支持，把PC带入家庭，使PC连通每家用户的电视、电话与立体声音响，并成为家庭管理和娱乐中心，是MPC全部努力的最终目标。



## 1.2 多媒体发展历程(续3)

### ◆多媒体系统关键技术

- 编码与压缩（音频及视频）
- 硬件支持（海量存储及各种多媒体设备）
- 软件支持（操作系统、数据库、制作工具）
- 信息组织和管理（超文本和超媒体）
- 人机交互（自然性，实时性，立体感）
- 网络支持（带宽、服务质量QoS）
- 系统集成（媒体内、媒体间同步，三电一体化）
- 安全（系统，信息，内容，隐私，版权...）





## 1.2 多媒体发展历程(续4)

### ◆多媒体系统若干热点

- 人机交互与多通道用户界面
- 语音输入, 手写体输入, 自然语言理解
- 虚拟现实及应用, 数字地球, 分布式虚拟现实技术, 基于图像绘制技术
- 科学计算可视化及信息可视化
- 多媒体处理, 如图像、视频检索等 (如点播、视频库、CAI)
- 智能电网、智能电视
- 多媒体云
- 社交网络
- 移动搜索





# 1.3 多媒体数据压缩与编码

## ◆ 压缩和编码有何区别？

- 压缩不一定都用编码，而编码可以压缩数据

## ◆ 为什么要压缩

- 降低多媒体数据对存储器容量的要求
- 降低多媒体数据对传输带宽的要求

1280\*960的彩色照片：3.68M  
1秒钟音乐：1.4112M；30分钟的  
16bit未压缩WAV音乐数据量大约317M；  
1分钟320\*240的彩色监控视频数据量为330M，1小时19G，1天463G

## ◆ 两种类型的数据压缩

- 无损压缩：经过压缩和解压缩后的数据与压缩前的原始数据完全一样的数据压缩技术。文本、代码和数字数据文件必须用无损压缩方法压缩。可把普通文件大小压缩到原来的25~50%；常用的无损压缩算法包括哈夫曼编码和LZW等算法。
- 有损压缩：经过压缩和解压缩后的数据与压缩前的原数据不完全一样，但但不影响人对原始资料表达的信息造成误解的数据压缩技术。电视和声音文件经常包括多于呈现给观众和听众的信息，因此可用有损压缩方法去掉这些信息，使压缩后的数据减少到原来的5%



有什么用处？

## 1.3 多媒体数据压缩与编码(续1)

### ◆ 三种类型的编码

- 熵编码：不考虑数据源的无损数据压缩技术。其核心思想是按照符号出现的概率大小给符号分配长度合适的代码，对常用的符号给它分配长度较短(即位数较少)的代码，对不常用的符号给它分配长度较长(即位数较多)的代码。最常见的熵编码技术是霍夫曼编码和算术编码
- 源编码：考虑数据源特性的数据压缩技术。编码时考虑信号源的特性和信号的内容，因此也称“基于语义的编码(semantic-based coding)”。例如，图像编码考虑相邻像素的值可能完全相同或相近，视像相邻帧之间的变化不大，也可能完全相同。为获得比较大的压缩比，通常采用有损数据编码技术。
- 混合编码：组合源编码和熵编码的数据有损压缩技术。影视、图像和声音媒体几乎都采用这种编码方式，如JPEG，MPEG-Video和MPEG-Audio。





# 1.3 多媒体数据压缩与编码(续2)

## ◆ 压缩与编码

表 1-1 压缩与编码技术

压缩类型	编码类型	编码技术	
无损压缩 (lossless compression)	熵编码 (entropy coding)	行程长度编码(run-length coding)	
		统计编码 (statistical coding)	霍夫曼(Huffman coding)
			算术编码(arithmetic coding)
有损压缩 (lossy compression)	源编码 (source coding)	预测(prediction)	差分脉冲编码调制(DPCM)
			增量调制(delta modulation)
		变换(transformation)	快速傅里叶变换(FFT)
			离散余弦变换(DCT)
			离散小波变换(DWT)
		(按重要性的)分层编码(layered coding)	二进制位的位置(bit position)
			子采样(subsampling)
			子带编码(subband coding)
	混合编码 (hybrid coding)	矢量量化*(vector quantization, VQ)	
		JPEG, JPEG 2000	
		MPEG-1, -2, -4, H.261~H.264	
		其他专有的编码方法	

\*矢量量化是一种数据有损压缩技术, 它用待编码的数据块匹配预先定义的有限数目的码字

# 1.4 多媒体与光盘

- ◆光盘存储器在多媒体发展史上起了重要作用
- ◆下列盘统称为光盘存储器或光盘
  - CD(compact disc)
  - DVD(Digital Versatile Disc)
  - HD DVD(High Definition DVD)
  - BD(Blu-ray Disc)
  - MO disk(magneto-optical disk)——磁光盘
- ◆光盘的存储容量

表 1-2 光盘的存储容量

名称		Blu-ray Disc	HD DVD	DVD	CD-ROM
激光波长(nm)		405(蓝紫)		650(红光)	780(红光)
数值孔径(NA)		0.85	0.65	0.6	0.45
存储容量 (单面)	单层	25 GB	15 GB	4.7 GB	650 MB
	双层	50 GB	30 GB	8.5 GB	-



# 1.4 多媒体与光盘(续)

## ◆ 每种光盘都有很多成员，包括

- -ROM,
- -Audio,
- -Video, -R(Recordable),
- -RAM
- .....

## ◆ 不同的成员用于存储不同的数据，见表1-3

表 1-3 DVD 与 CD 的主要成员

Blu-ray Disc*	HD DVD*	DVD	CD	主要用途
BD-ROM	HD DVD-ROM	DVD-ROM	CD-ROM	存储数据等
BD-Video	HD DVD-Video	DVD-Video	Video-CD	存储影视节目
BD-Audio	HD DVD-Audio	DVD-Audio	CD-Audio	存储音乐节目
BD-R**	HD DVD-R	DVD-R	CD-R	存储档案等
BD-RAM	HD DVD-RAM	DVD-RAM	CD-MO	随机存储器

\*有些规范正在制定。

# 1.5 多媒体与网络

## ◆ 计算机网络

- 从20世纪90年代开始，计算机网络对多媒体的发展起着巨大的推动作用，多媒体的传输越来越多地依靠网络
- 因特网(Internet)
  - ✓ 通过网络设备把世界各国使用TCP/IP协议的计算机相互连接在一起的计算机网络
  - ✓ 世界上规模最大、用户最多的计算机网络
  - ✓ 20世纪全球发展最迅速、影响最深远和冲击最大的信息存取和处理工具
- 互联网(internet, internetwork)
  - ✓ 泛指由多种网络通过网络互连设备互连而成的网络
  - ✓ 在世界上，因特网只有一个，而互联网却数不胜数
- 万维网(World Wide Web, WWW, Web, W3)
  - ✓ 全球性分布式多媒体信息系统
  - ✓ 因特网上的最典型应用





# 1.5 多媒体与网络(续)

## ◆万维网的核心技术

- 统一资源地址(URL): 指定网上信息资源地址的统一命名方法
  - ✓ 通过它可指定访问资源时所用的协议(如HTTP, FTP)、资源所在服务机的名称、资源的路径和资源名称
- 超文本传输协议(HTTP): 在服务机和客户机之间传送超文本文档的通信协议
  - ✓ 用于建立与Web服务器的连接和给客户浏览器传送HTML网页
- 超链接(HyperLink): 用于在资源之间建立定向逻辑链接, 实际上是一个单元指向另一个单元的指针
  - ✓ 例如, 文档中的某个元素(单词、短语、符号或图像)指向同一个文档或不同文档中的另一元素、同一个Web站点或不同Web站点上的另一个文档、程序或其他对象的指针



## 1.6 多媒体国际标准

### ◆标准

- 由公认的非商业化组织推荐、政府组织推荐或既成事实的硬件或软件技术准则
- 由专家组或委员会对现有方法、步骤和技术经过反复深入细致研究后编写的详细说明

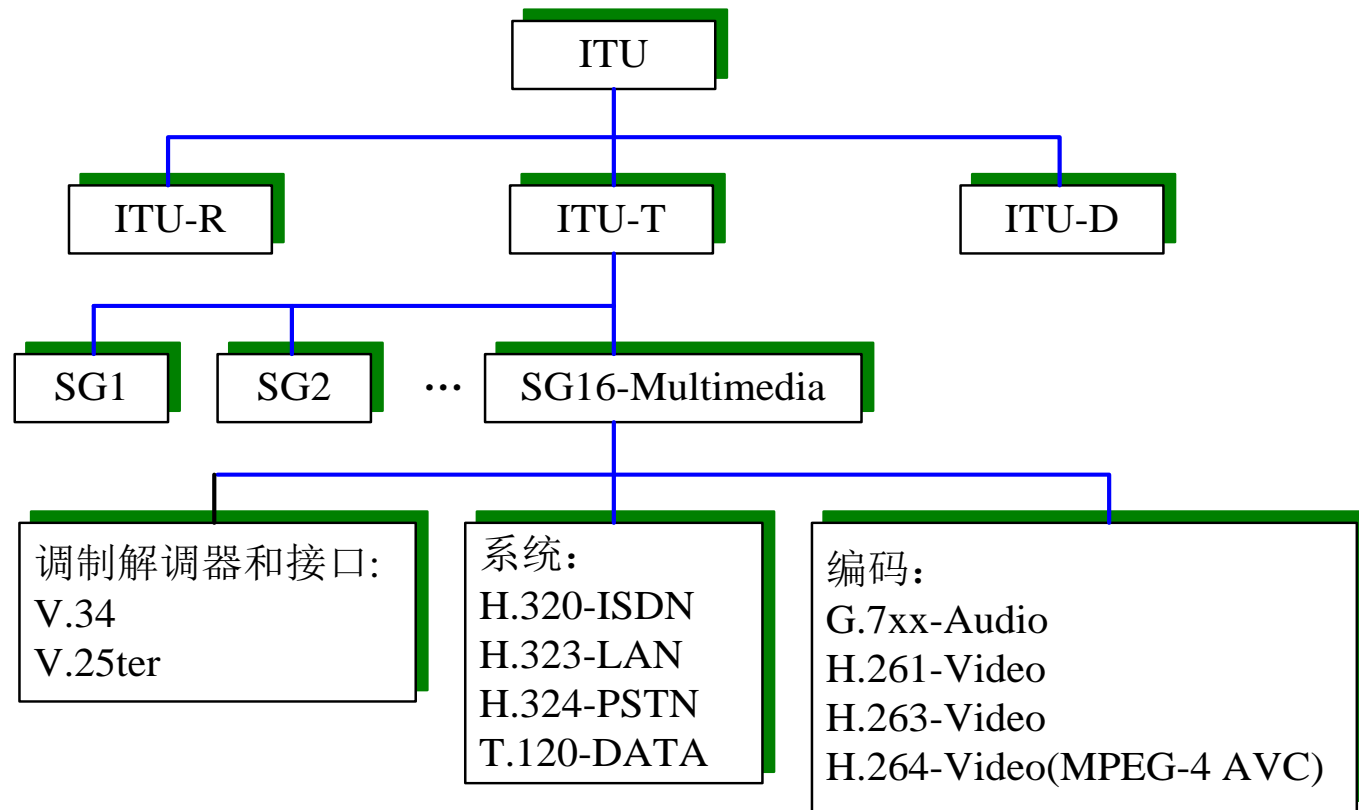
### ◆四种国际标准：

- ITU标准
- ISO/IEC标准
- 因特网技术标准
- 万维网协会(W3C)标准



# 1.6 多媒体国际标准(续1)

## ◆ 国际电信联盟(ITU)标准



VCEG开发的标准





# 1.6 多媒体国际标准(续2)

## 应用设计

SGs 7, 9, 16,  
ITU-R

高级应用系统: 电子商务、远程学习、远程医疗和媒体广播等  
基本应用系统: E-mail, WWW, FTP, 虚拟线路(VPS/VCS)等

## 中间件设计

SGs 7, 9, 16, SSG,  
ITU-R  
JTC1, IETF, W3C,  
3GPP, 3GPP2

会话(Conversation), 消息传递(Messaging), 检索(Retrieval), 广播(Broadcasting)等

媒体编码&会话  
QoS  
文档结构

结构  
汇聚  
通信协议的执行

目录  
传真  
安全

可访问性(Accessibility)  
人机接口

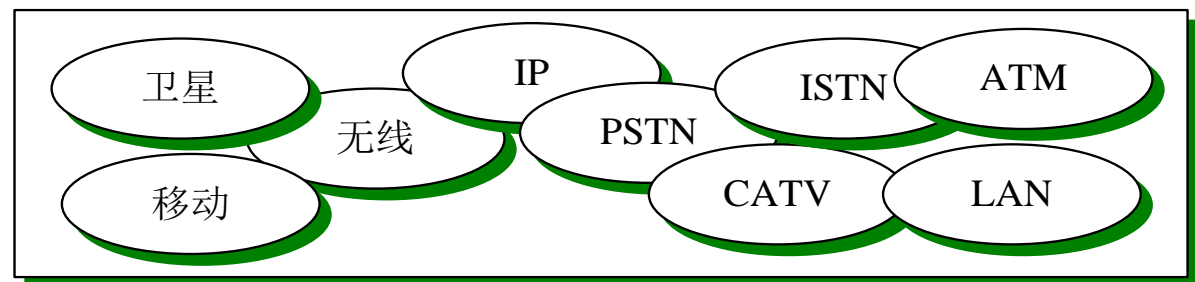
因特网&无线接入

互操作性  
媒体通信  
工作协调

## 网络设计

SGs 2, 4, 11,13, 15, SSG,  
ITU-R,

网关控制(Gateway control), 各种数字订户线路(XDSL), 调制解调器(Modem)等



ITU多媒体通信标准化框架的研究范围和研究组



## 1.6多媒体国际标准(续3)

### ◆ISO/IEC标准

#### ➤ 国际标准化组织(ISO)

- ✓ 1946年成立的自愿参加和无条件约束的国际组织,总部设在瑞士日内瓦,其成员目前有130个国家的国家标准化组织
- ✓ 负责制定包括计算机、通信等众多领域的国际标准,以便于国际间信息、科学、技术、经济等活动领域的相互交流合作
- ✓ 突出成就: 定义了被广泛接受的计算机网络互联的ISO/OSI模型
- ✓ 注意: ISO (International Organization for Standardization)不是该组织英文全称的缩写,而是取自希腊词isos, 表示“平等”的意思



## 1.6 多媒体国际标准(续4)

### ➤ 国际电工技术委员会(IEC)

- ✓ IEC (International Electrotechnical Commission)是成立于1906年的国际性电工标准化机构，现有60多个国家参加
- ✓ 负责有关电气工程和电子工程领域中的国际标准化工作。目前已扩展到包括电子、电力、电磁、微电子、通信、机器人、多媒体、能源、仪器仪表等领域的技术规范以及相关的通用规范，如术语和符号、度量方法和性能、设计和开发、安全与环境等。我国已于1957年参加该委员会
- ✓ IEC于1993年将多媒体作为未来标准化工作的最重要范围之一，并成立了多媒体研究委员会(Multimedia Research Committee)，把多媒体标准化模型系统(Multimedia Standardization Model Systems)、多媒体用户接口(Multimedia User Interface)和多媒体中的颜色管理(Color Management in Multimedia)等内容纳入IEC的研究范围
- ✓ IEC的工作促进了IEC、ISO和ITU三大国际标准化组织联合开发多媒体标准的进程



## 1.6 多媒体国际标准(续5)

### ➤ ISO/IEC JTC 1

- ✓ ISO/IEC Joint Technical Committee 1的缩写
- ✓ ISO和IEC在1987年联合成立的技术委员会。由于JTC 1的研究范围是信息技术领域的标准化，因此常把ISO/IEC JTC 1称为“ISO/IEC联合信息技术委员会”
- ✓ ISO/IEC JTC 1下设许多技术分会，分别负责制定各自专业范围的标准。现有18个分会(subcommittee, SC)
  - 例如，SC 29 - Coding of Audio, Picture, and Multimedia and Hypermedia Information(声音、图像和多媒体与超媒体信息的编码)
    - SC下面有工作组(WG)
    - ISO/IEC JTC 001/SC 29/WG 01 "Coding of still pictures"
    - ISO/IEC JTC 001/SC 29/WG 11 "Coding of moving pictures and audio"
- ✓ ISO/IEC JTC 1成立以来制定了许多非常著名的标准，如MPEG-1, -2, -4，对推动信息技术的发展和应用做出了卓越贡献







## 1.6 多媒体国际标准(续6)

### ◆ 图像、声音和视频编码标准

#### ◆ 图像编码标准

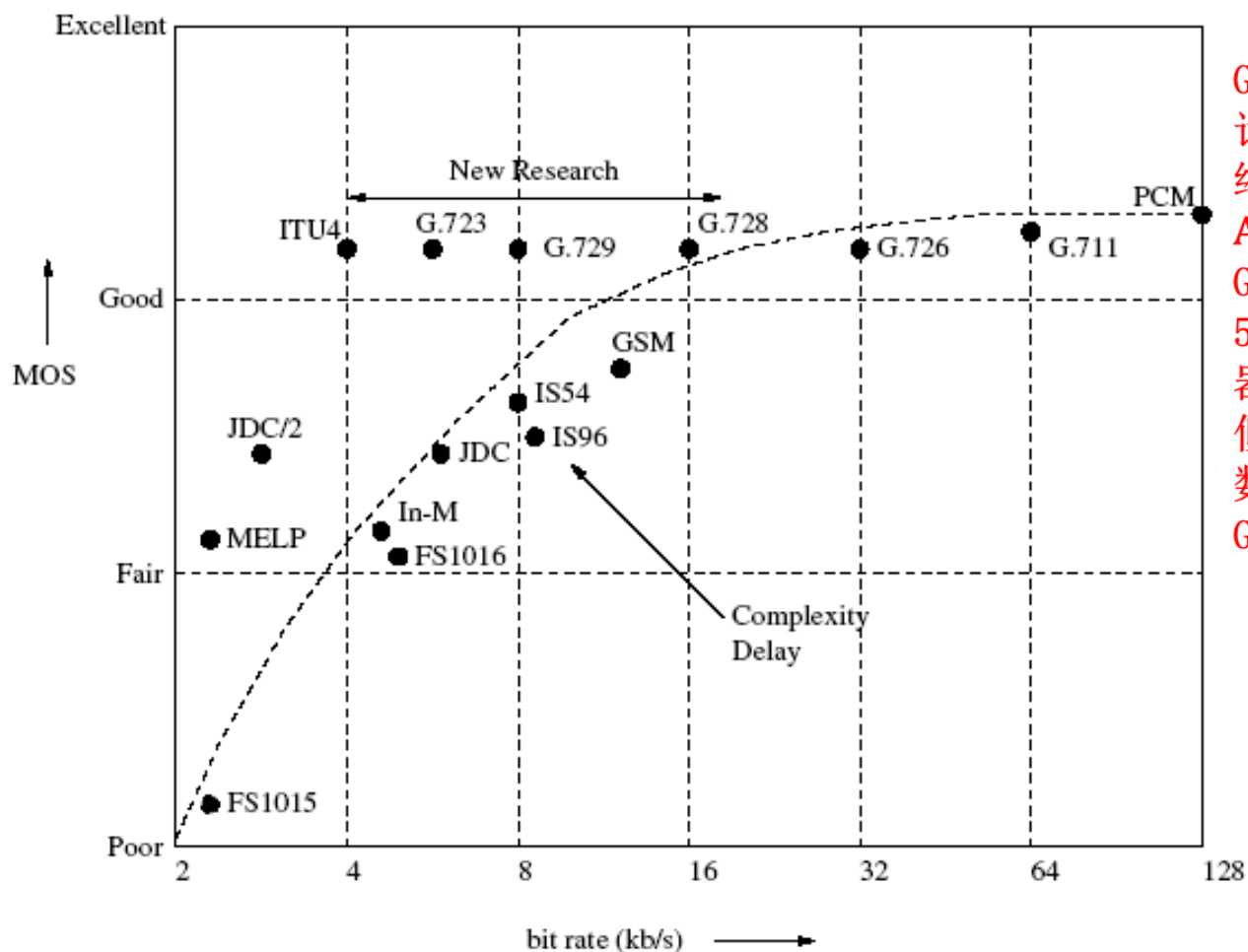
- 传真图像压缩, CCITT, Group1-4, RLE, Huffman, 当前使用的Group3
- 1986-1994, JPEG
- 1987-88, GIF (LZW)
- 1993, CCITT联合ISO, JBIG
- 1987, GIF
- 1995, PNG(LZ77的改进版本zlib)
- 1996-2001, JPEG2000
- 2015-2020, JPEG PLENO

✓ <https://jpeg.org/>

# 1.6 多媒体国际标准(续7)

## ◆声音、图像和视频编码标准

## ◆语音编码标准



G.711: 波形编码, 主要用于电话, 采样率8k每秒, 16bit, 压缩成8bit, 64kbps, 分为

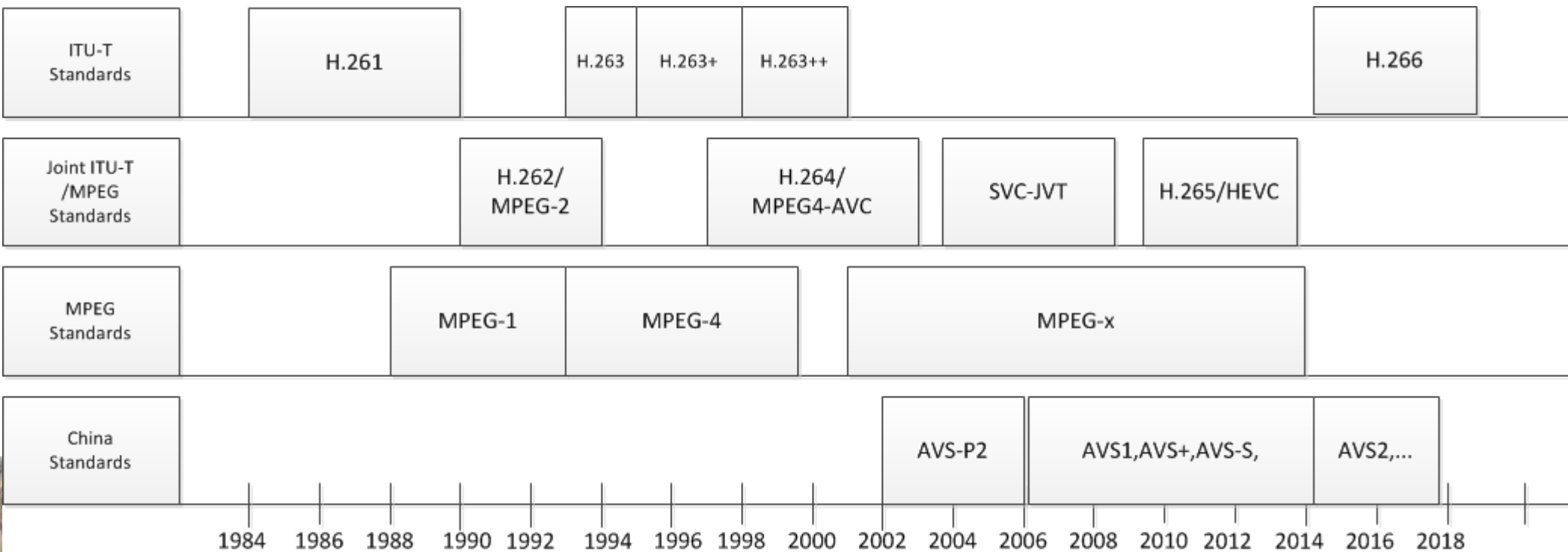
**A-law,  $\mu$ -law**

G.723: 采用LPC, 分为6.3kbps, 5.3kbps两种码率, 高速率编码器的激励信号采用多脉冲最大似然量化(MP-MLQ), 低速率采用代数激励线性预测(ACELP)

G.729: 波形参数混合编码

# 1.6 多媒体国际标准(续8)

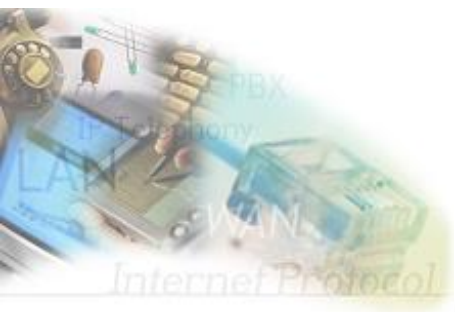
- ◆声音、图像和视频编码标准
- ◆视频编码标准



## 1.6 多媒体国际标准(续9)

### ◆ 因特网标准化组织

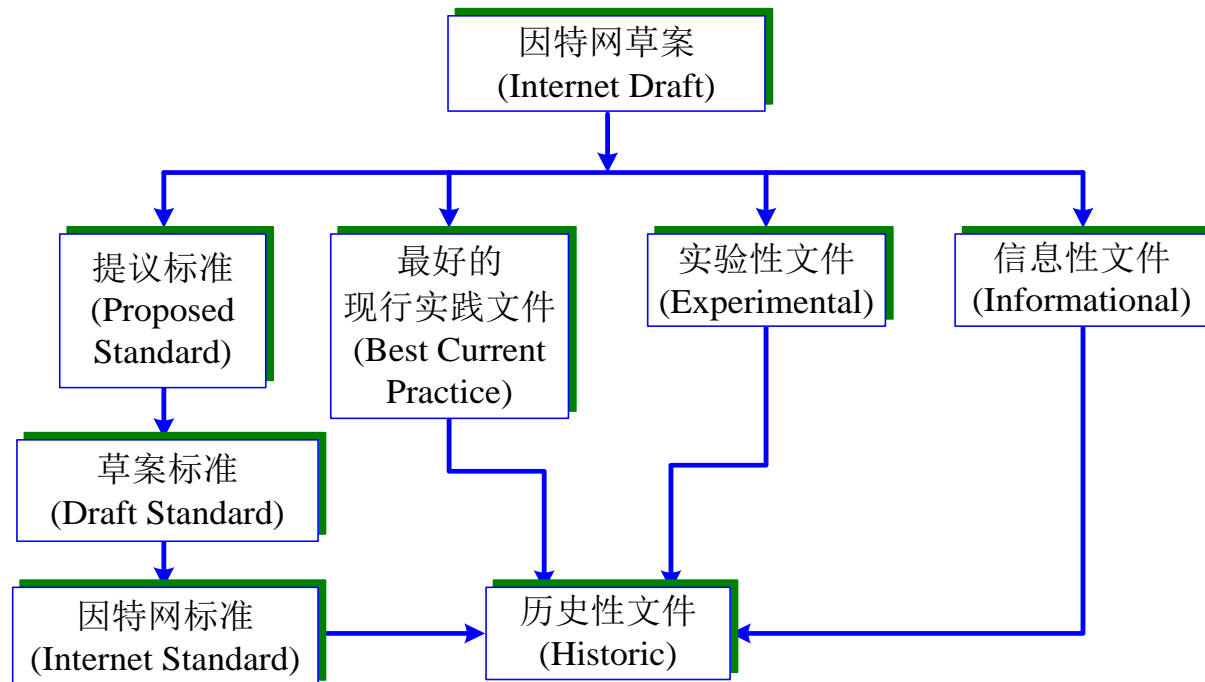
- 因特网标准由1992年成立的因特网协会或称互联网协会 (Internet Society, ISOC)协调的四个组织负责
  - ✓ 因特网工程特别工作组(Internet Engineering Task Force, IETF): 成立于1986年, 从事研究与因特网有关的技术问题并给因特网体系结构研究部(IAB)提出解决方案
  - ✓ 因特网体系结构研究部(Internet Architecture Board, IAB): 由一组定期举行会议讨论因特网有关事宜的研究人员组成的团体, 他们对各类问题做出决定和安排, 制定因特网的发展规划。成立于1983年, 原名为Internet Activities Board
  - ✓ 因特网工程指导组(Internet Engineering Steering Group, IESG): 由各地区的主管(Area Director)和因特网工程特别工作组(IETF)的主席组成, 实际是IETF的执行委员会, 与因特网体系结构研究部(IAB)一起评审IETF提出的标准
  - ✓ 因特网研究特别工作组(Internet Research Task Force, IRTF): 从理论上考虑长期的因特网问题的志愿组织, 并将其建议提交给因特网体系结构研究部(IAB)



## 1.6 多媒体国际标准(续10)

### ➤ 因特网技术标准

- ✓ 主要由IETF承担，因此也称IETF标准。从1969年开始，开发一个因特网标准时通常先公布提案，并命名为“RFC (Request for Comments)”文件
- ✓ 例如，2008年1月23日-2013年2月，公布的因特网标准共有67个-74个，公布的RFC文件共有5235个-6879个



因特网标准的制定过程

<http://www.rfc-editor.org/rfcxx00.html>  
<http://www.ietf.org/download/rfc-index.txt>





## 1.6 多媒体国际标准(续11)

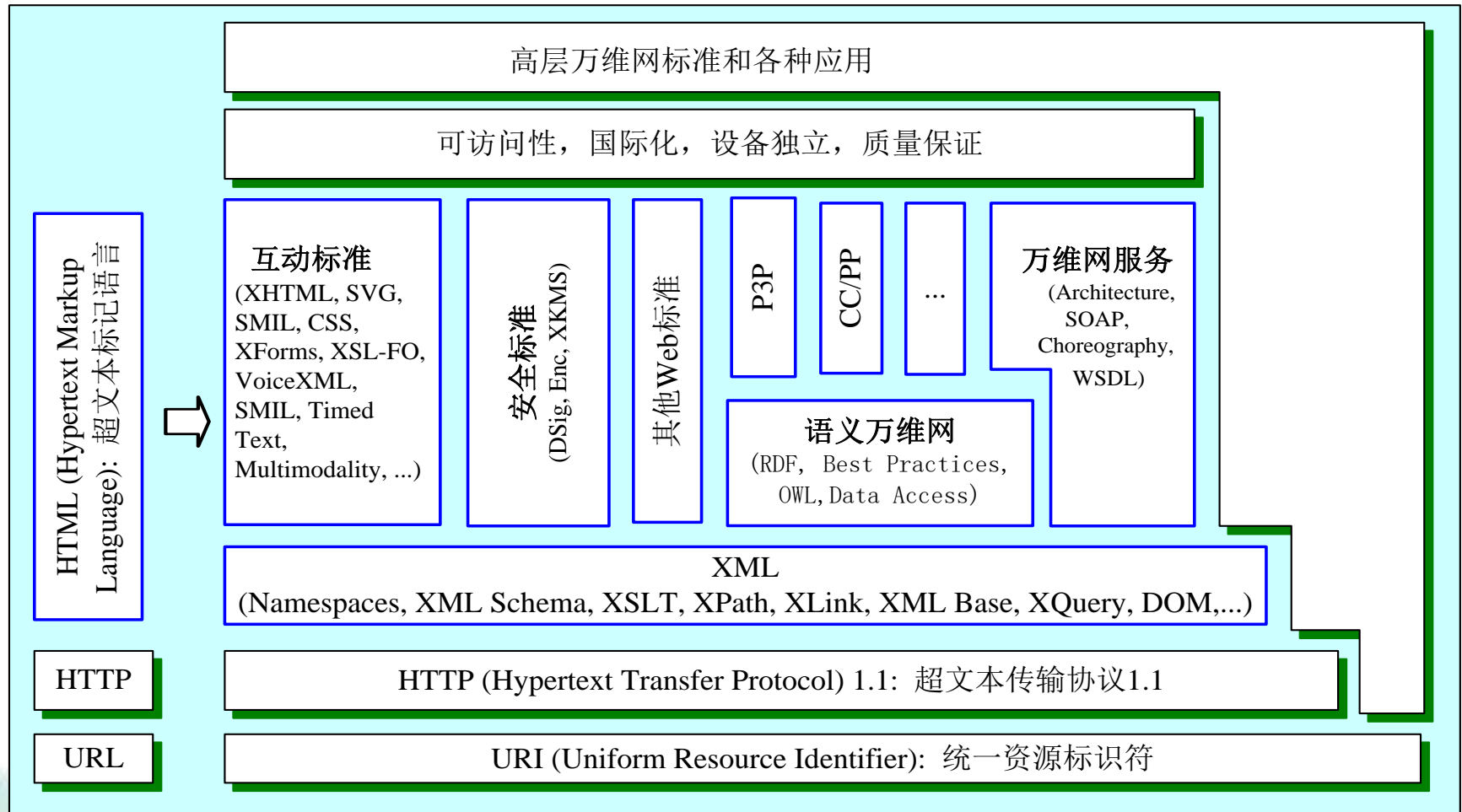
### ◆万维网协会(W3C)

- World Wide Web Consortium的缩写
- 1994年10月在美国麻省理工学院计算机科学实验室成立的开发万维网技术的国际性组织
- 致力于分析万维网的研究状况，促进行业标准(如HTML，XML等语言标准)的制定，鼓励开发可协同工作的产品(如Web浏览器)
- 目前由发明万维网的Tim Berners-Lee领导，美国的麻省理工学院计算机科学与人工智能实验室(MIT CSAIL)、法国的欧洲信息与数学研究论坛(ERCIM)和日本的应庆大学共同管理
- 1994年以来，开发了90多种适合当今和将来万维网模式的设计规范，称为“W3C推荐标准(W3C Recommendations)”

2012年12月17日，万维网联盟（W3C）正式宣布凝结了大量网络工作者心血的HTML5规范已经正式定稿。根据W3C的发言稿称：“HTML5是开放的Web网络平台的奠基石。”

2013年3月4日，HTML5编程语言的一个最新漏洞被发现，它允许网站利用数GB垃圾数据对用户展开轰炸，甚至会在短时间内将硬盘塞满。多款主流浏览器均会受此影响。

# 1.6 多媒体国际标准(续12)

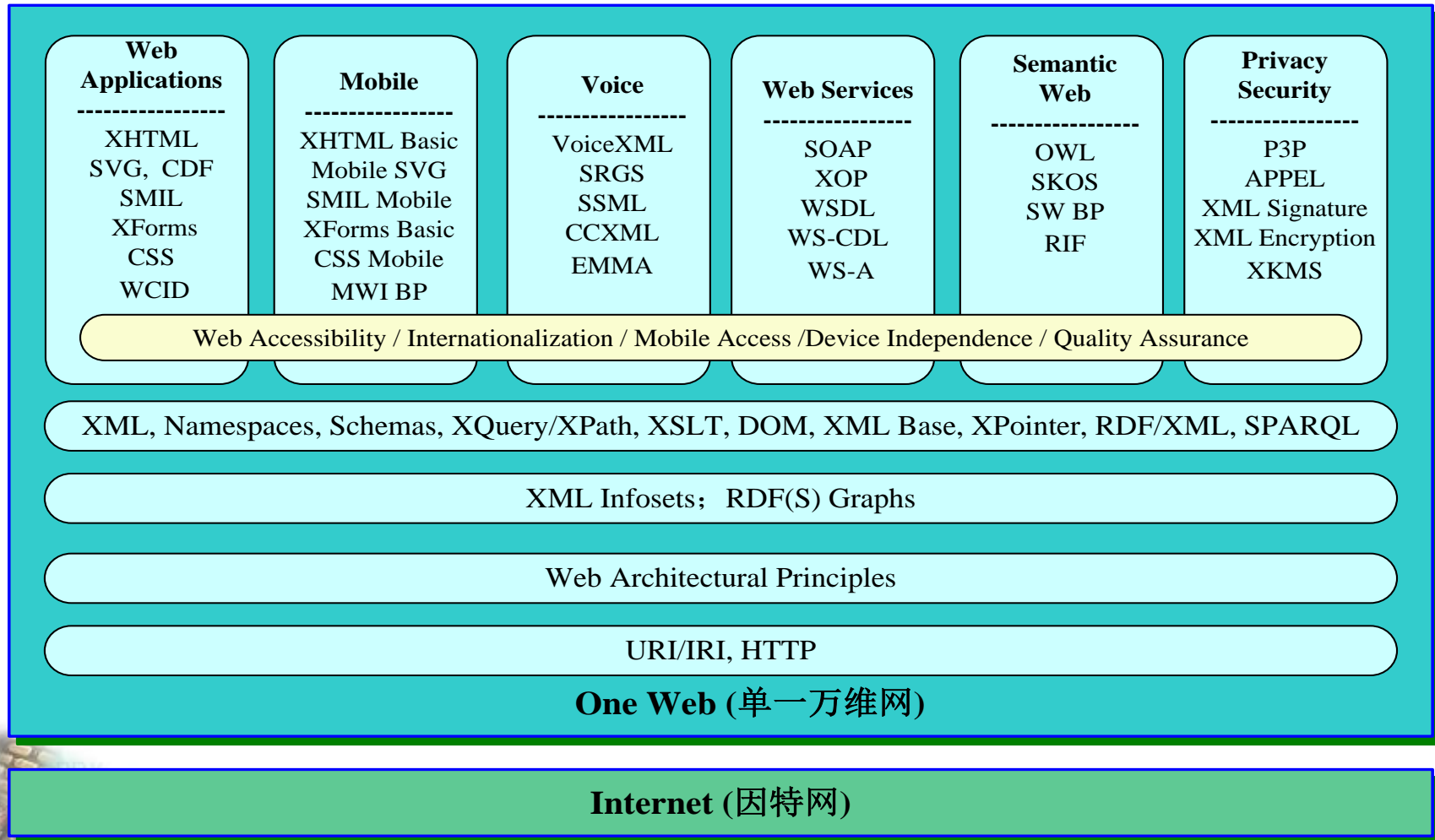


构建万维网的基本规范

(引自 <http://www.w3.org/2004/04/w3c-flir-v1.6.3A4.pdf>)



# 1.6 多媒体国际标准(续13)



## W3C技术框架

(引自 <http://www.w3.org/Consortium/technology>)



# 1.7 多媒体内容处理

## ◆ DCIKW是什么

### ➤ data(数据)

- ✓ 以数字、字符或图像等可读语言或其他记录方法表示的事实、概念或计算机指令
- ✓ 适用于人或自动装置进行通信、解释或处理
- ✓ 数据本身没有意义，通常需要在一定的语义环境中才有意义

### ➤ Content(内容)

- ✓ 内容是对数据的描述

### ➤ information(信息)

- ✓ 信息是对内容的解释，也就是数据的含义
- ✓ 出现在上下文中并能给人传达某种含义时，数据就成为信息
- ✓ 从广义上说，信息就是消息

### ➤ knowledge(知识)

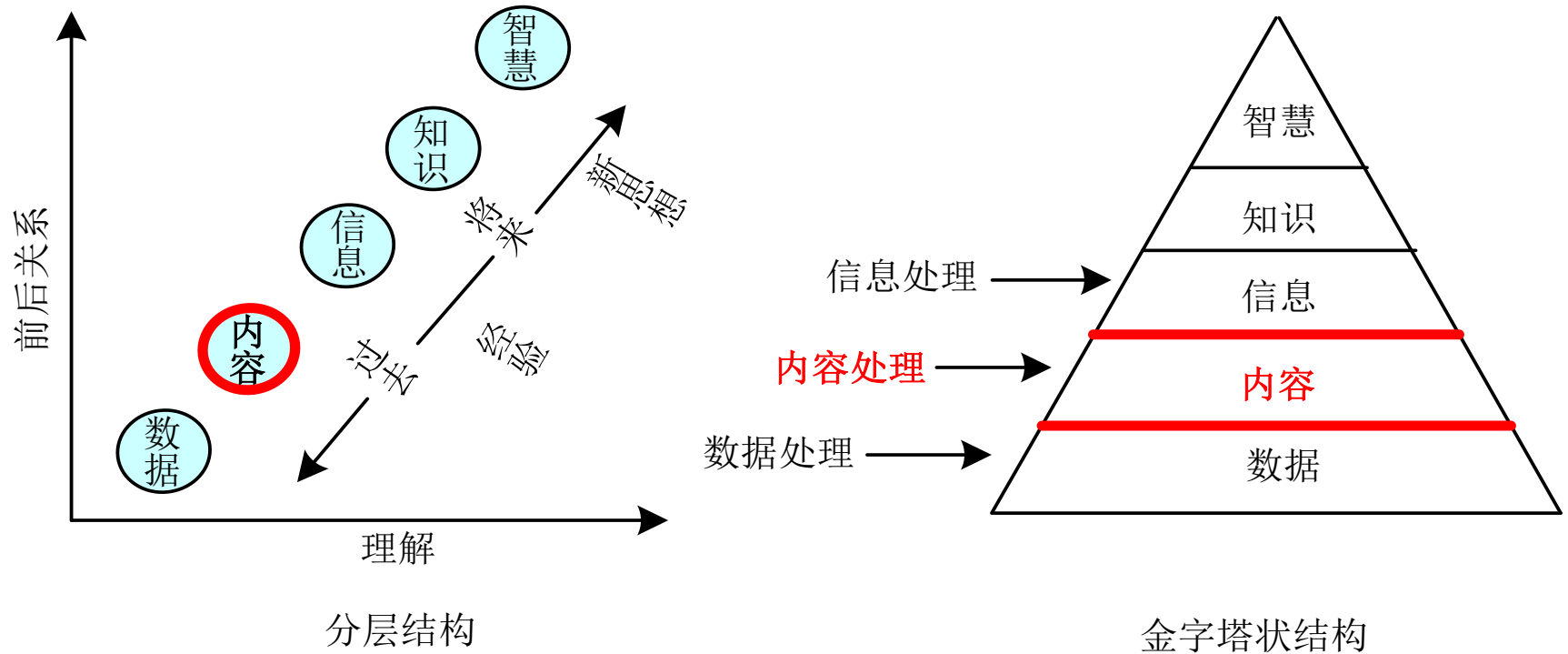
- ✓ 在某个感兴趣领域中的事实、概念和关系

### ➤ wisdom(智慧)

- ✓ 知识累积后产生的洞察力、判断力和发明创造能力



# 1.7 多媒体内容处理(续1)



DCIKW的结构







# 1.7 多媒体内容处理(续2)

## ◆ 多媒体内容是什么

- “内容是对数据的描述，信息是对内容的描述，内容不都是信息，信息不一定包含全部内容”
- “多媒体内容是对多媒体数据的描述，多媒体信息是对多媒体内容的描述，多媒体内容不都是多媒体信息，多媒体信息不一定包含全部多媒体内容”

## ◆ 多媒体内容处理(multimedia content processing)

- 本教材将“内容处理”定位在“数据处理”和“信息处理”之间的处理，因此可把内容处理理解为对内容执行的操作，包括内容的分析、分类、管理、搜索、检索、浏览、传输及其所需的数据处理
- 对内容处理的解释同样可扩展到对多媒体内容处理的解释
- 多媒体内容处理可理解为用计算机自动或半自动地将多媒体数据表示成内容以及对内容进行操作的技术
- 多媒体信息处理可理解为用计算机自动或半自动地将多媒体内容表示成信息以及对信息进行操作的技术



# 1.7 多媒体内容处理(续3)

## ◆ 多媒体内容处理包含

- 多媒体内容分析(multimedia content analysis): 解决的问题是用什么方法和特征来描述没有文字或难以用文字描述的媒体数据。目前比较成熟但不太有效的描述方法是用它们的低层特性(如颜色直方图、形状和轮廓)来表示图像、视像或音乐
- 多媒体内容分类(multimedia content classification): 对给定的多媒体内容判断它属于哪种类型的技术
- 多媒体内容管理(multimedia content management): 在多媒体内容处理系统中控制多媒体内容的获取、保存、检索和发布等的技术
- 多媒体内容搜索(multimedia content search): 查找指定的多媒体内容或多媒体数据的技术, 使用类似性度量法, 通过计算和比较以判断是否匹配或者满足给定的搜索属性或条件
- 多媒体内容检索(multimedia content retrieval): 从存储媒体、数据库或网络服务机中获取多媒体内容的技术
- 多媒体内容浏览(multimedia content browsing): 粗看文字、图片、视像或粗听声音以寻找特定内容的技术





# 1.7 多媒体内容处理(续4)

## ◆ 内容标记语言

- 标记(markup)是组织和标注文档内容的一套字符集
- 标记语言(markup language, ML)是用于组织和表示数据的一套规则, 在文件中体现为一系列语句或代码

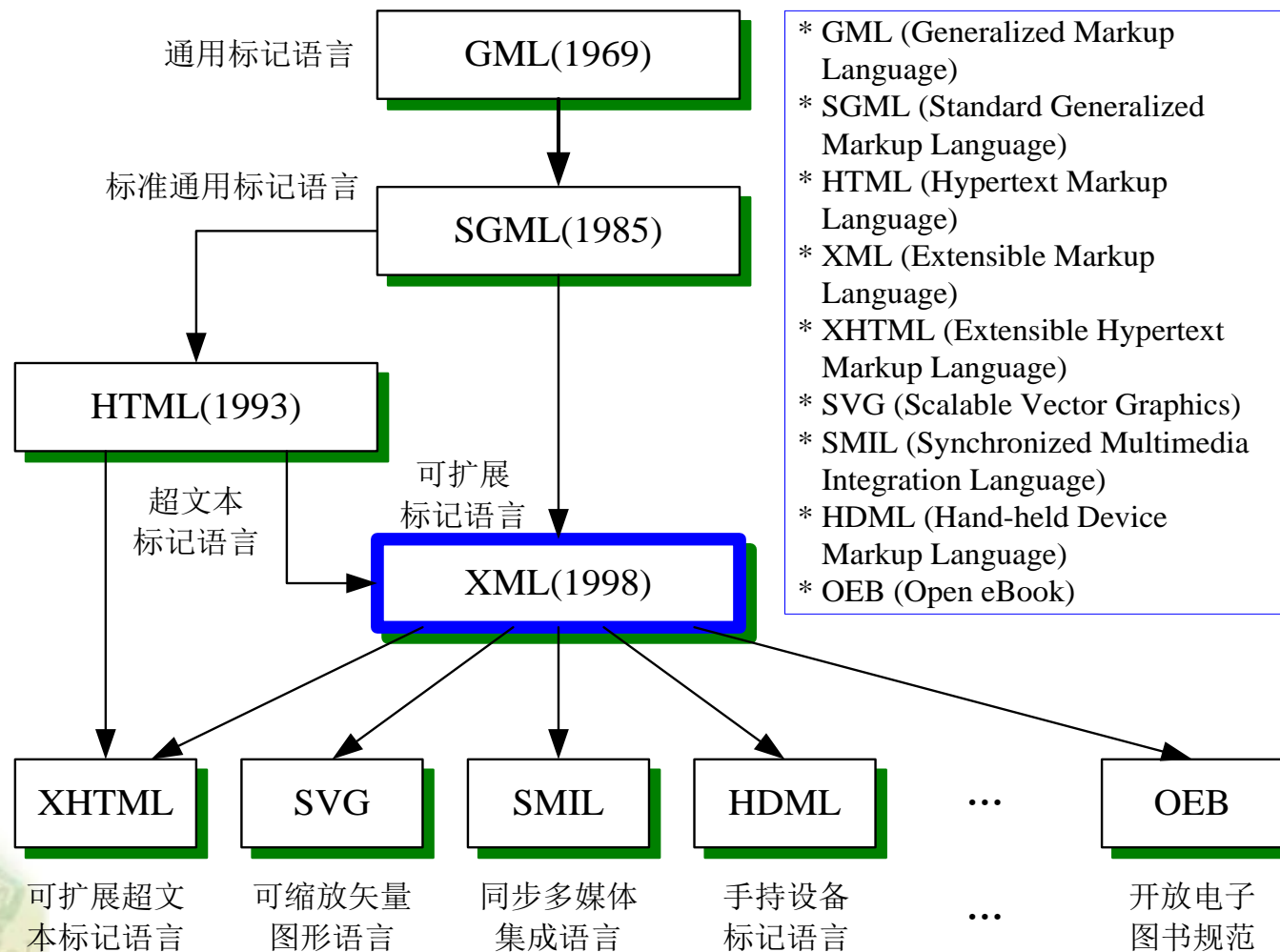
## ◆ 标准通用标记语言(SGML)

- Standard General Markup Language的缩写
- markup的含义是可插入到文档(document)中的标记。两种标记
  - ✓ 程序标记(procedural markup), 用来描述文档显示的样式(style), 如字体的大小、黑体、斜体和颜色等, 现在市场上出售的大多数的字处理软件都内嵌有标记, 而且几乎都是针对自己的软件产品而制定的标记
  - ✓ 描述标记(descriptive markup), 也称为普通标记(generic markup), 用来描述文档中的文句是什么, 如篇、章、节、段落、列表或表格, 而不是描述文句所显示的样式。
- SGML的精华是把文档的内容与样式分开处理
  - ✓ 可支持无数的文档结构类型, 如布告、技术手册、章节目录、设计规范、各种报告、信函和备忘录等
  - ✓ 可创建与特定的软硬件无关的文档, 因此很容易与使用不同计算机系统的用户交换文档



# 1.7 多媒体内容处理(续5)

## ◆ 标记语言系列



# 1.7 多媒体内容处理(续6)

## ◆多媒体内容检索

### ➤ 内容分析是核心

- ✓ 内容分析是对视听材料进行自动或半自动标注, 通过从视听材料中抽取的文字和视听材料的特征, 使用机器学习和其他技术对视听材料中的语义概念(semantic concept)建立模型, 模仿视听材料中出现的语义概念, 如人群(crowd)、道路(road)、动物(animal)、小汽车(car)等
- ✓ 语义概念实际上是一组实例或事件特征的抽象, 并用一套规则表示, 以确定概念与特征之间的隶属关系。

### ➤ 开发多媒体内容描述辞典(Thesaurus with Multimedia Content Description)

- ✓ 左边的猫熊在网上通常没有文字描述, 而在辞典中通常只有文字解释, 如果使用既有文字描述又有猫熊低层特性(如颜色直方图)描述的“多媒体内容描述辞典”, 这将有助于提高检索“猫熊”词条的查全率(recall)和查找精度(precision)





## 1.7 多媒体内容处理(续7)



### 猫熊(panda)

哺乳动物，体长4~5尺，形状像熊，尾短，通常头、胸、腹、背、臀白色，四肢、两耳、眼圈黑色，褐色，毛粗而厚，性耐寒。生活在我国西南地区高山中，吃竹叶、竹笋。是我国特产的一种珍贵动物。……，also 熊猫，大熊猫，大猫熊。——引自中国社会科学院语言研究所词典编辑室，《汉英双语现代汉语词典》，(2002年增补本)

图1-12 多媒体内容描述辞典中的“猫熊”词条

# 1.8 迈向信息时代

## ◆ 信息技术工业的发展可分成4个时代

- 1964~1981年以大、中型机为中心(Systems-centric)
- 1981~1994年以个人计算机为中心(PC-centric)
- 1994~2005年以互联网络为中心(Network-centric)
- 2005~2015年以内容为中心(Content – centric)

——引自“计算机世界(Computerworld)”专栏作家David C. Moschella在1997年出版的书《Waves of Power: Dynamics of Global Technology Leadership 1964-2010》

## ◆ 以内容为中心的时代

- 从20世纪90年代开始，随着互联网的普及，信息技术已逐步渗透到人类活动的各个方面
- 与电视网络系统类似，当技术嵌入到系统变成不可见之后，人们的注意力将会从网络软硬件本身的研究和开发逐步转向到软硬件的使用和服务，以网络为中心的时代将逐步过渡到以内容为中心的时代；我们感受到这个时代已经来到





# 1.8 迈向信息时代(续1)

## ◆迈向信息时代

- 信息时代需要信息科学指导
- 信息科学研究进程中有过三个好主意
  - ✓ 20世纪50年代出现的在形式逻辑基础上处理信息的“信息检索(information retrieval)”
  - ✓ “相关(relevance)”概念
  - ✓ 20世纪80年代出现的人和系统之间的“互动(interaction)”概念。相关和互动相结合就构成了“相关反馈(relevance feedback)”，实际上就是把“正在查找和处理信息的人”加到信息处理系统，构成由人参与调节的自动控制系统。
- 进入20世纪90年代，引入机器学习理论和技术也应该算是一个好主意，这个主意在多媒体内容处理中得到了充分利用，并且取得了相当大的进展





## 1.8 迈向信息时代(续2)

### ◆迈向多媒体信息时代

- 以多媒体内容为中心过渡到以多媒体信息为中心需要相当长的时间，因为指导多媒体内容和信息处理的多媒体信息科学[1]才刚刚起步

[1] 多媒体信息科学是以多媒体信息为主要研究对象的科学。过去有些学者认为，信息科学要研究的内容包括：阐明信息的概念和本质(哲学信息论)，发掘信息的处理机制(智能理论与计算理论)，寻找信息的传递规律(通信理论)，揭示信息的调节原则(控制理论)，探索信息的度量和变换(基本信息论)，研究信息的提取方法(识别信息论)，开发信息的再生理论(决策理论)。多媒体信息科学的研究也不应该有例外



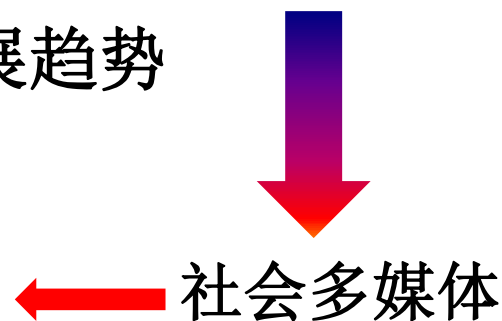
## 1.8 迈向信息时代(续3)

### ◆多媒体信息科学的核心-媒体计算

- 媒体计算，特别是网络环境下的海量媒体计算已经成为新型信息技术战略性新兴产业国家重大需求的共性基础
- 多媒体信息处理技术在新时代上的特征
  - ✓ 海量化：以互联网为代表的海量媒体信息基础设施
  - ✓ 语义化：要求计算机能够理解媒体信息内容，精确获取和推送
  - ✓ 多样化：媒体资源经过信息化手段后各种非结构化信息资源
  - ✓ 社会化：Web媒体应用的发展中，用户参与其中，既是信息的消费者，又是信息的提供者，社会网络与媒体内容相互融合和渗透
- 展望一下未来几年多媒体技术的发展趋势



社会多媒体计算  
多媒体社会网络







## 1.8 迈向信息时代(续4)

### ◆多媒体技术的未来发展趋势

- 与认知科学相结合
- 基于机器学习的媒体信息高效计算
- 扩展信息论的基础理论是媒体信息表示与传输的有效途径
- 媒体内容安全是媒体信息处理的重要内容，在我国信息安全战略中占有十分重要的地位
- 新型硬件体系结构与计算平台将成为媒体信息高效计算的基础
- 跨媒体智能处理

### ◆研究热点

- 上下文情景化理解
- 多媒体交互动力学建模
- 盗版追踪及版权保护
- 社交媒体社区分析
- 隐私保护
- 跨媒体智能

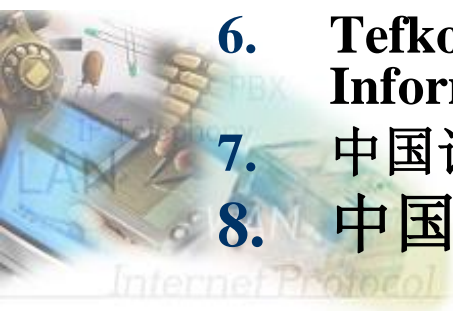


# 第1章 多媒体技术概要



## ◆ 参考文献和站点

1. ITU-T Study Groups (Study Period 2005 - 2011), <http://www.itu.int/ITU-T/studygroups/areas-domain.html>.
2. MEDIACOM 2004, *A Framework for Multimedia Standardization, Project Description – Version 3.0* March 2002. <http://www.itu.int/itudoc/itu-t/com16/mediacom/projdesc.html>.
3. KOM–Multimedia Communications Lab, <http://www.kom.tu-darmstadt.de/mm-book/> (username: mm\_book, psw: mm\_docs).
4. Nikhil Sharma, Updated: December 1, 2005, *The Origin of the “Data Information Knowledge Wisdom” Hierarchy*, [http://www-personal.si.umich.edu/~nsharma/dikw\\_origin.htm](http://www-personal.si.umich.edu/~nsharma/dikw_origin.htm)
5. Geoff Nunberg, Paul Duguid, *The Data, Information, Knowledge, Wisdom Chain: The Metaphorical link*, [http://ioc.unesco.org/Oceanteacher/OceanTeacher2/02\\_InfTchSciCmm/DI\\_KWchain.pdf](http://ioc.unesco.org/Oceanteacher/OceanTeacher2/02_InfTchSciCmm/DI_KWchain.pdf)
6. Tefko Saracevic, *Information Science*, Journal of the American Society for Information Science. 50(12):1051–1063, 1999
7. 中国计算机学会通讯, 7(2):1-96, 2011.2
8. 中国计算机学会通讯, 8(4), 2012.4, 8(5), 2012.5, 8(12), 2012.12



# 编程作业



- ◆ 回去了解BMP图像文件的结构
- ◆ 编程实现BMP图像的显示，以及像素值的获取；对其进行任意大小的分块并置乱，然后显示置乱后的结果（类似马赛克处理）
- ◆ 编程：掌握BMP和JPEG图像的格式，以及JPEG压缩基本流程，并编写界面友好的程序来阅读和显示BMP以及JPEG图像，提交源代码和可执行程序，并撰写实验总结报告，包括文件格式，压缩流程，界面编写原理





END

## 第1章 多媒体技术概述

