



# 数字媒体技术基础

**Meng Yang**

[www.smartllv.com](http://www.smartllv.com)

**SUN YAT-SEN University**



**机器智能与先进计算  
教育部重点实验室**



**智能视觉语言  
学习研究组**

# Course Materials



- **Prescribed textbook:**

- 讲义: 智能新媒体——人工智能下的数字媒体技术 (在编)

- **Other references:**

- Ze-Nian Li, Mark S. Drew, Jiangchuan Liu, Fundamentals of Multimedia (Second Edition), Springer, 2014
- 数字媒体基础及应用技术, 姬秀娟、周彦鹏、张晓媛、杨艳竹, 清华大学出版社, 2014年3月1日

- **Lecture notes will be available online**

- 见教学大纲
- 1.1 多媒体基础
  - 1.1.1 媒体与多媒体
  - 1.1.2 人的多媒体感知
  - 1.1.3 计算机多媒体
- 1.2 数字化多媒体
  - 1.2.1 数字化媒体系统
  - 1.2.2 数字媒体表示

## 第一部分

### 1.1 多媒体基础

# 信息技术发展的三个阶段

## 信息、信息技术以及3C技术

史前时代

语言、表情  
感觉器官



- 洞窟和摩崖壁画是现在发现最多的史前绘画，最早的距今已有约2万年的历史。

# 信息技术发展的三个阶段

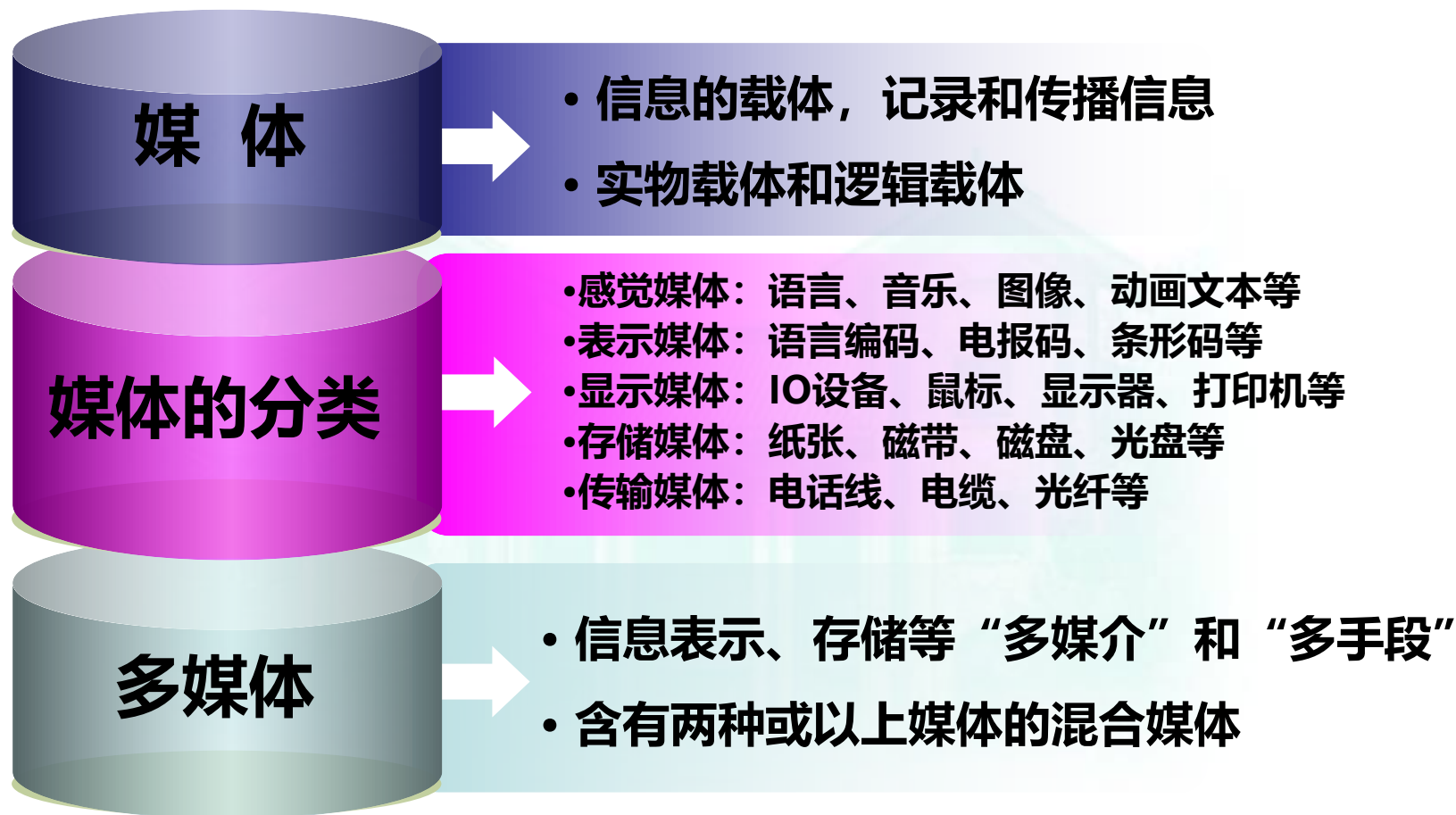
信息、信息技术以及3C技术



□ 1946年世界上第一台计算机ENIAC在美国宾州大学诞生，数字计算机的诞生标志着数字化信息时代的开始。



# 1.1.1 媒体与多媒体



## 1.1.2 人的多媒体感知

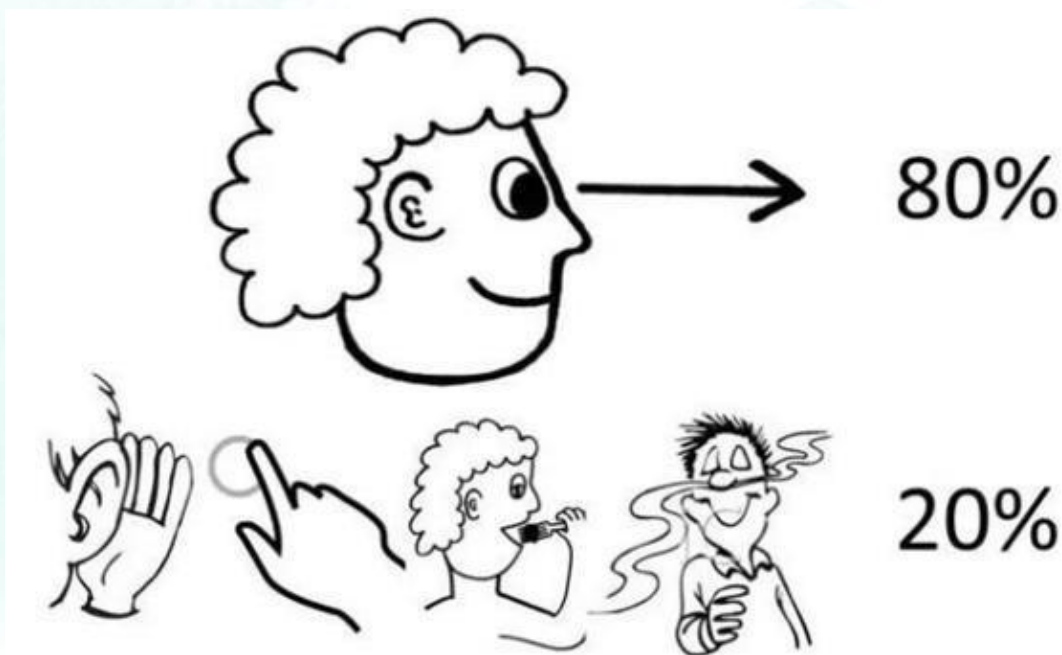
- 五感：视觉、听觉、味觉、嗅觉、触觉





## 1.1.2 人的多媒体感知

- 五感：视觉、听觉、味觉、嗅觉、触觉

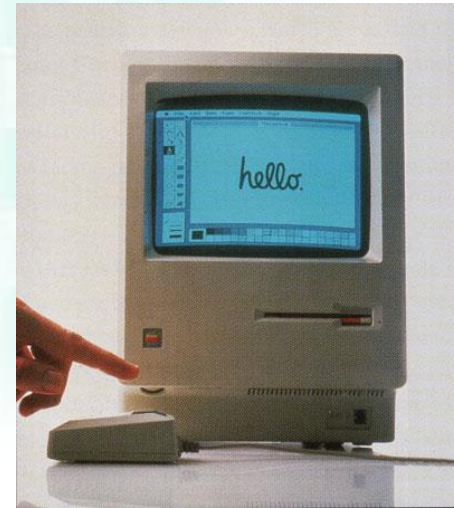


## 1.1.2 人的多媒体感知

- ❑ 最近有科学家在《自然》（Nature）发表论文指出：“多媒体工作”与“失忆”密切相关——喜欢同时对着手机和电脑双线操作的人往往出现注意力问题和记忆困难。大脑的记忆抑制与媒体多任务处理之间存在密切关联；在电脑和手机之间多线作业会加剧舌尖现象（指的是记忆的暂时性抑制，更通俗地说，就是话到嘴边了，可你死活就是想不起自己原本要介绍的某人的名字，某道试题的答案，或是某个故事的结局。）

# 1.1.3 计算机多媒体

- ❑ “Multimedia” has no strict definition.
- ❑ In our context, multimedia indicates the computer technology (multimedia computing) for more efficient communication by using different media types:
  - Text
  - Audio and speech
  - Images
  - Graphics
  - Video

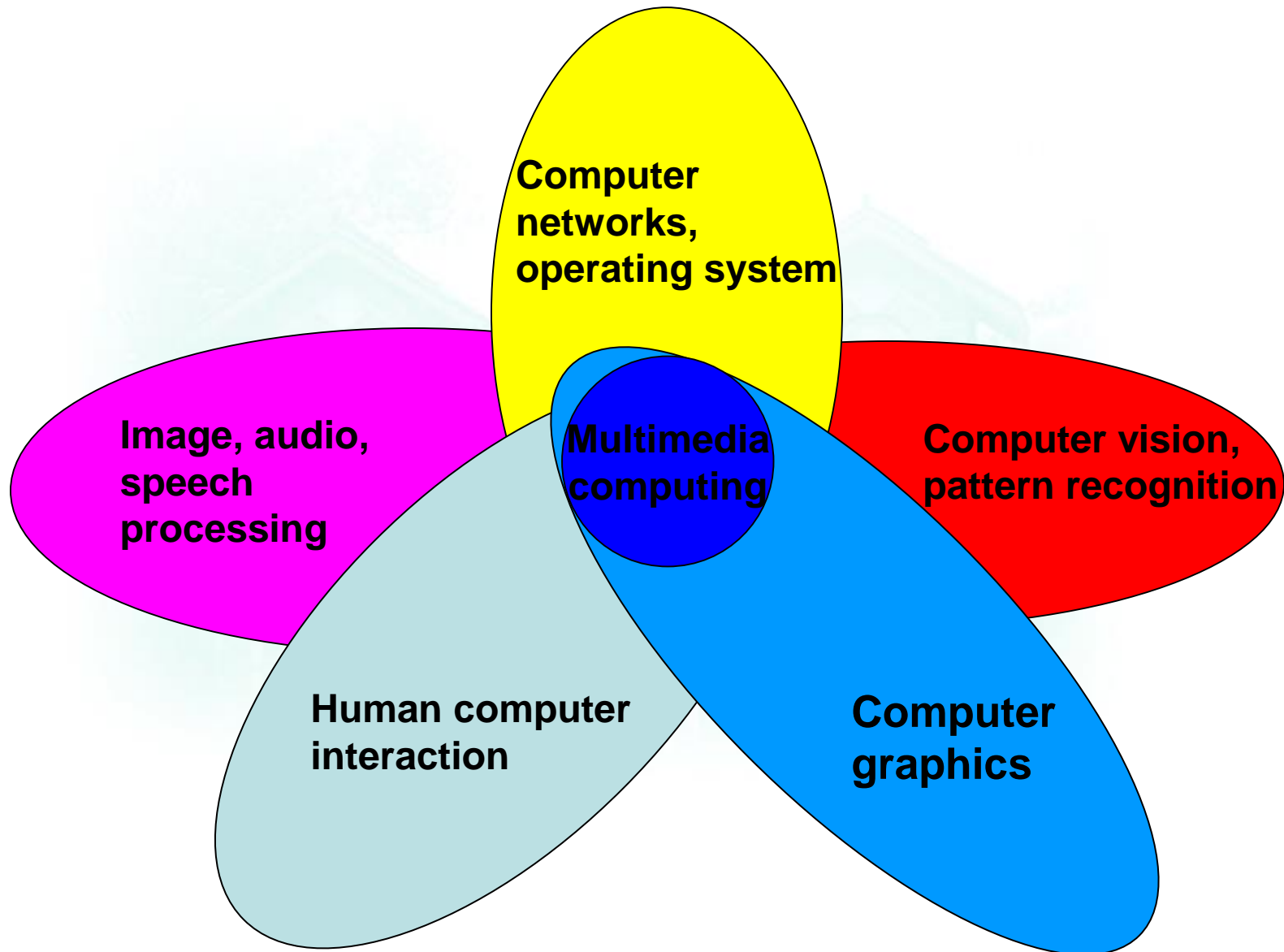


- ❑ Multimedia involves more than simple addition of new data types.
- ❑ It integrates a wide range of symbol modes simultaneously into a coherent framework.
- ❑ The framework is usually denoted as a multimedia system.

## ❑ **Developing a successful multimedia system is non-trivial.**

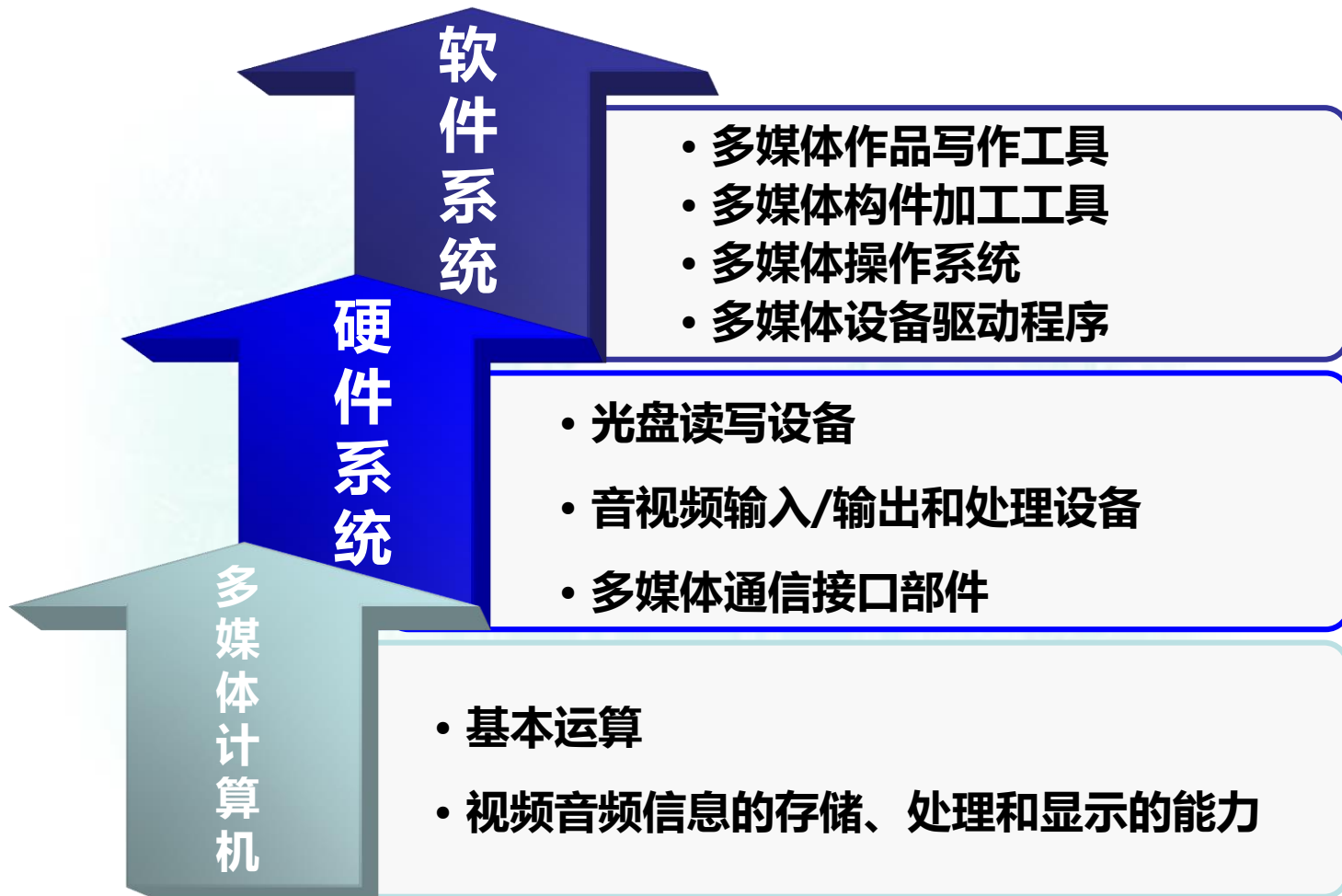
- Continuous media types such as video need a lot of space to store and very high bandwidth to transmit.
- They also have tight timing constraints.
- Automatically analyzing, indexing and organizing information in audio, image and video is much harder than from text.
- Multimedia involves many different research areas and needs more complex and more efficient algorithms and hardware platforms.

# Multimedia is Multidisciplinary





# 1.1.3 多媒体计算机



# 1.1.3 多媒体计算机

多媒体软件系统各种软件间的相互关系如下图所示：

多媒体素材采集工具软件	多媒体构件加工工具软件	多媒体作品写作工具软件	多媒体作品软件	多媒体计算机软件系统
多媒体操作系统				
多媒体设备驱动程序				
多媒体计算机硬件系统				

# 1.1.3 多媒体计算机

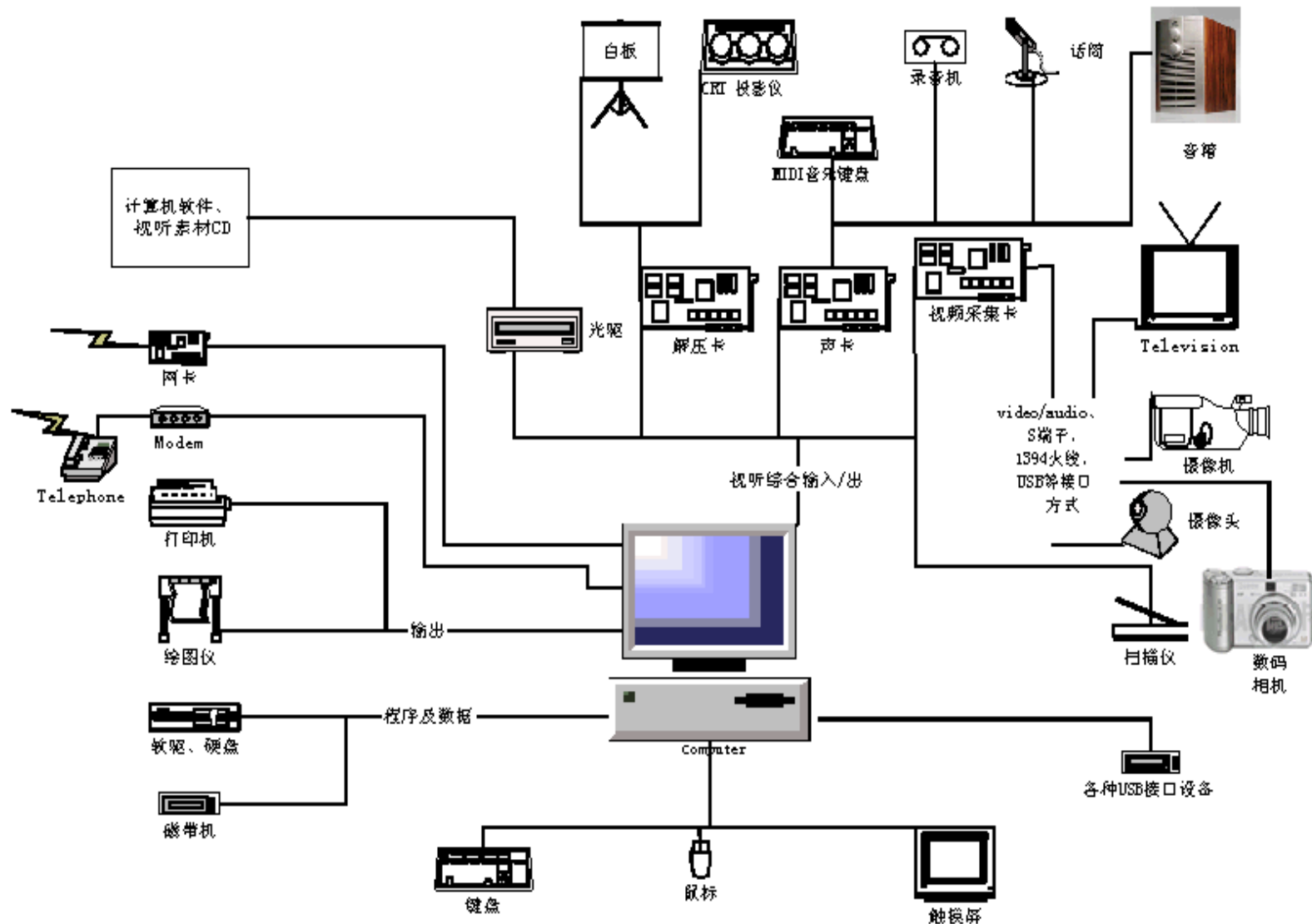
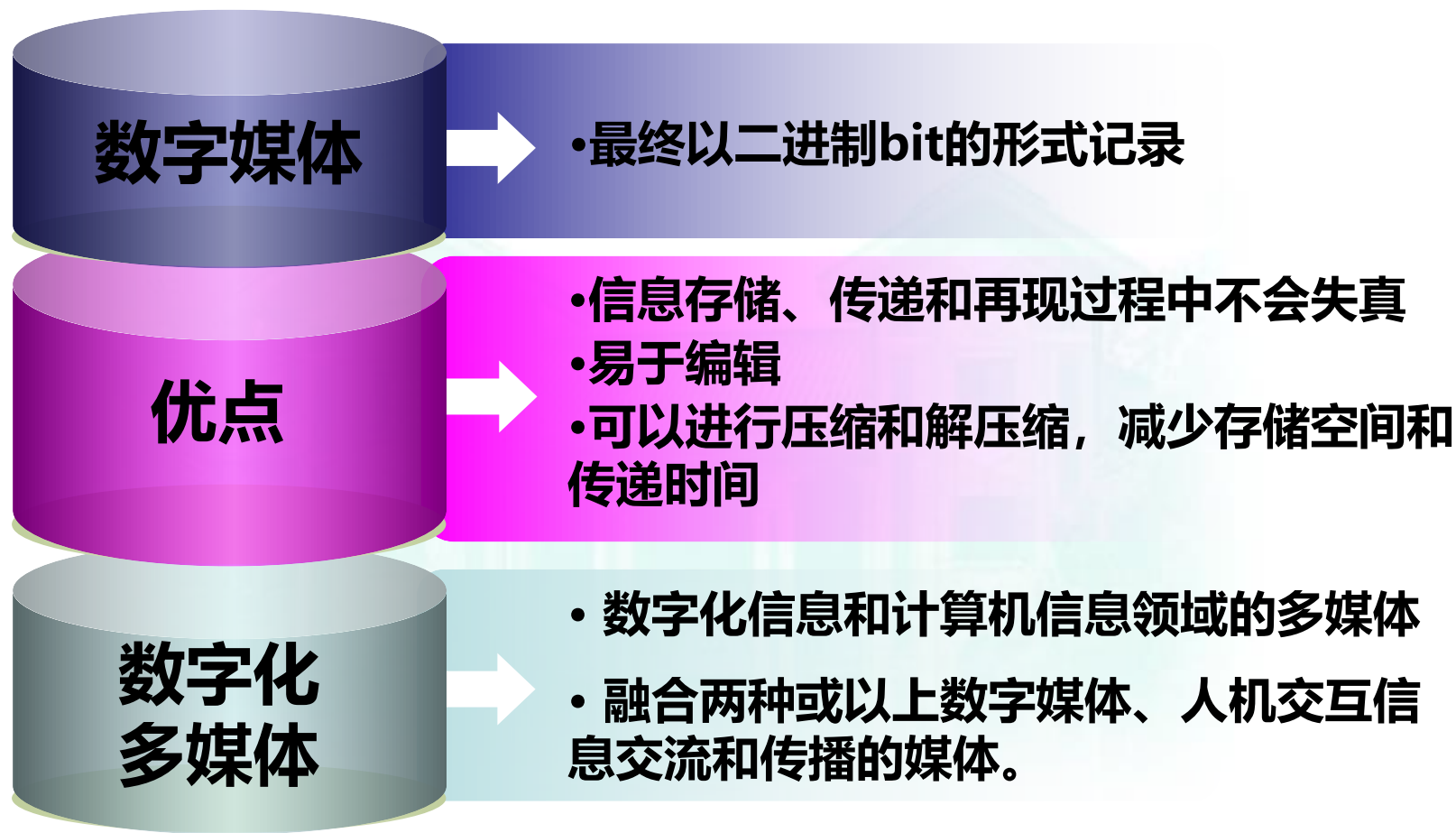


图1.1 多媒体计算机的配置

## 第二部分

### 1.2 数字化多媒体

## 1.2.1 数字媒体与数字化多媒体



# 数字化多媒体技术的主要特征

1.必须由计算机控制;

2.必须能够以数字化的形式存储、记录、变换、传递和再现;

➤数字化

3.集成化的系统;

➤集成化

4.最终用户界面必须是人机交互式;

➤交互性

5.对信息的存取可以是非线性的;

6.出版发行方式可以是非纸张的。

➤传播

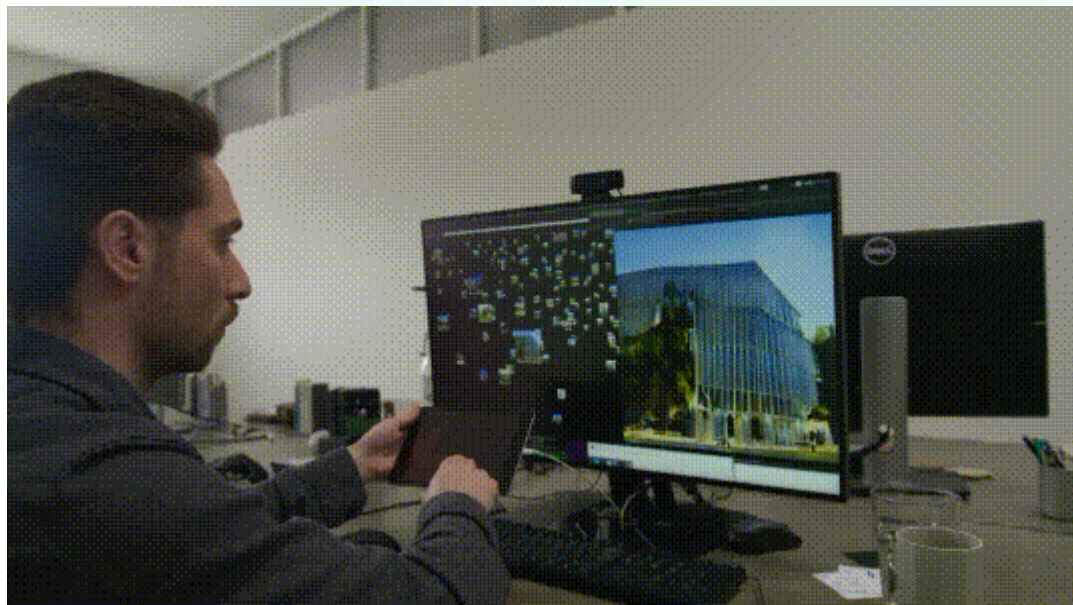
多媒体系统特征



## 1.2.2 数字媒体表示技术

- ❑ 国际电话电报咨询委员会把媒体分为五类：
  - 感觉媒体 (Perception Medium)
  - 表示媒体 (Representation Medium)
  - 表现媒体 (Presentation Medium)
  - 存储媒体 (Storage Medium)
  - 传输媒体 (Transmission Medium)
  
- ❑ **数字媒体 (Digital Media)** 是指基于计算机信息网络技术，以**二进制数**的形式记录、处理、传播和获取的能在全局范围内即时互动传播信息的**信息载体**，这些载体包括数字化的文字、图形、图像、声音、视频影像和动画等感觉媒体，和表示这些感觉媒体的编码等，统称为表示媒体，以及存储、传输、显示逻辑的存储媒体与表现媒体。

- ❑ 在这个数据模型中，AI 通过几千张照片“看到”了自由女神像的不同角度，学习算法产生建筑物的视觉联想，利用360度投影装置创建一系列代表该建筑各个时期不同状态的动态图像，就好像建筑有了生命一样。



# AI+数字媒体

- ❑ 艺术家用投影仪呈现为迪士尼音乐厅外墙上的可视化艺术装置「WDCH Dreams」。此作品也是用生成式对抗网络创作的。



## 多媒体数据表示技术

- 图形等在计算机内表示
- A/V编码,压/解压缩技术
- 语音合成与文语转换
- 海量存储设备
- 多媒体接口设备等硬件

## 多媒体数据库

- 基于内容的图像检索  
根据分析图像的内容,提取颜色、形状纹理及空间关系,建立特征索引。
- 基于内容的音视频检索  
真正成熟的多媒体数据库的出现还需要经过一个漫长的过程。

## 多媒体软件

- 多媒体操作系统
- 多媒体作品构件加工
- 多媒体作品构件制作
- 多媒体作品创作工具
- 超文本与超媒体组织

## 人机交互

- 语音输入
- 手写输入
- 自然语言理解



## 虚拟现实及其应用

- 网络支持环境的开发
- 多媒体系统的集成与开发
- 多媒体通信协议
- 多媒体网络支持
- 科学计算及信息可视化

## 多媒体应用系统开发

- 多媒体节目制作
- 视频会议
- 远程教育系统
- 智能化信息家电



# Example Multimedia Systems



## Real Estate

### House Flip Book

Page 2 of 6

Next Previous First Last Load...

**Cambridge** \$1050000

Architectural gem from the height of the American Renaissance. Historic house on academic side street near Harvard. Beautifully proportioned rooms with stunning

**Cambridge** \$995000

Customized colonial with all-new appointments in exclusive location. The warmth of an older home with the security of newer construction. Prime executive location; perfect

**Cambridge** \$1025000

Vibrant renovation of impressive Victorian home on national historic register. Exquisite gardens and immaculate grounds. Serious inquiries only.

**Cambridge** \$890000

You'll never want to leave home again. 3 soaring chimneys and dormers galore in this spacious hideaway.

### GTE

#### Video Conference

Quit

#### Mortgage Calculator

(values are only approximations)

	A	B
0	Home Mortgage Calculations	
1		
2		
3	House Price	120000.00
4	Down Payment	20000.00
5	Interest Rate	10.00
6	Years of loan	30.00
7		
8	Monthly Payments	877.57
9	Total Payments	335925.77
10	Total Interest	215925.77

This prototype real estate application was developed with the AthenaMuse software at GTE Laboratories. House descriptions and color images are retrieved from a multimedia database and placed in a customized listing booklet. The client and Realtor can discuss candidate homes and financing options via desktop videoconferencing and a shared document facility.

Interface:

Russell Sasnett, at GTE Laboratories.

**Real Estate in MIT's Project Athena**

# Example Multimedia Systems



Photosynth of Microsoft Live Labs.

- ❑ **Multimedia systems involve some basic enabling techniques:**
  - Multimedia data representation and compression.
  - Multimedia data processing and analysis.
  - Transmitting multimedia data through communication networks.
  - Multimedia database, indexing and retrieval.



# Data Compression

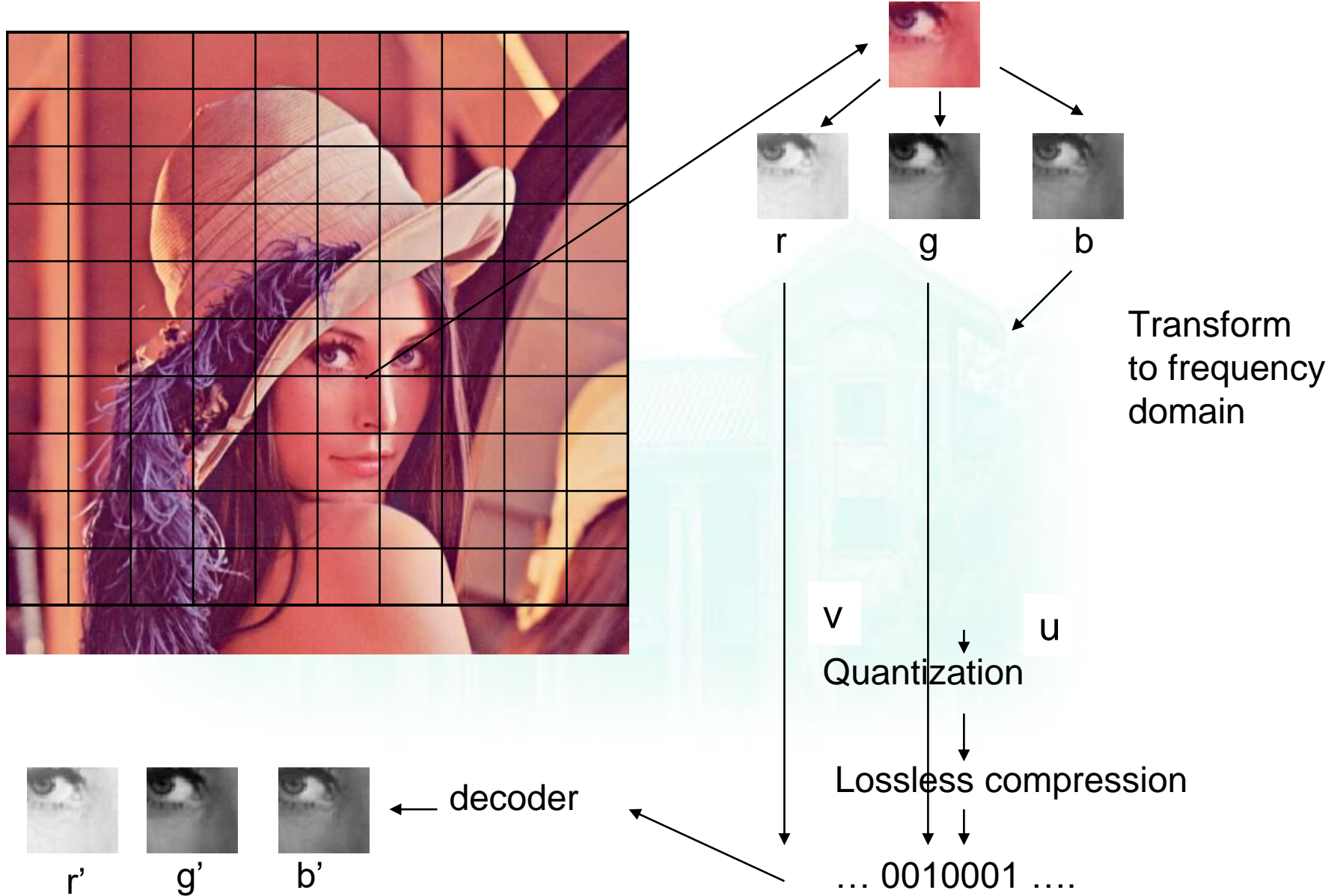


**Raw image takes about 6M bytes  
(without header information)**



**24k bytes with jpeg, Q=50**

# “JPEG” illustration



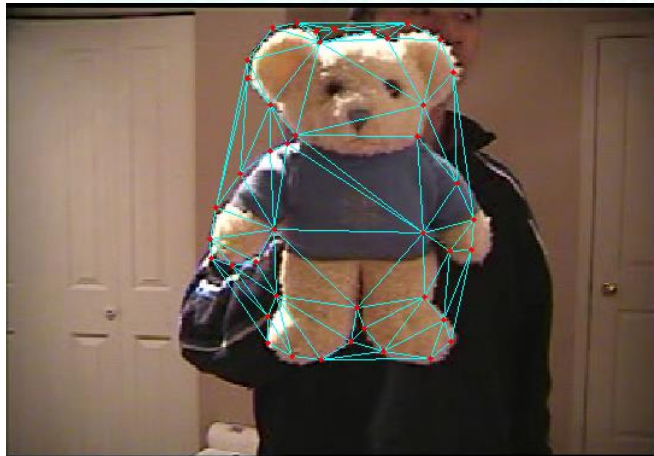
# Media Processing and Analysis

- ❑ In applications such as digital library, automatic data analysis has to be done to extract semantic meanings from audios, images and videos.
- ❑ Based on media processing methods such as:
  - object tracking (face, eyes),
  - object recognition,
  - gesture recognition, etc.,

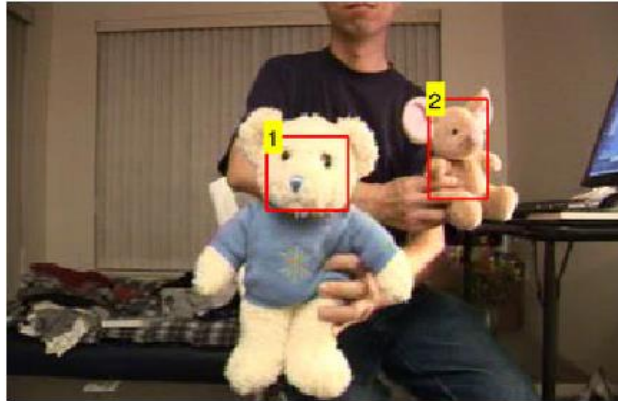
we can build more effective human computer interfaces.



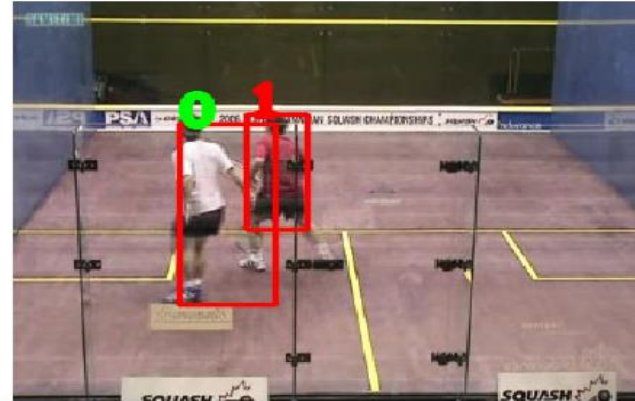
# Tracking Object by Image Matching



# Multiple Object Tracking



Toys



Squash



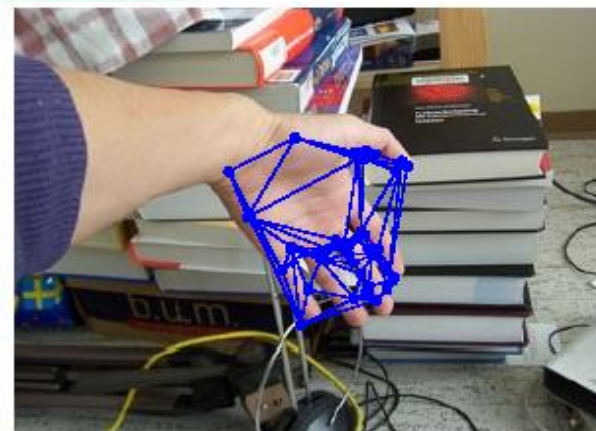
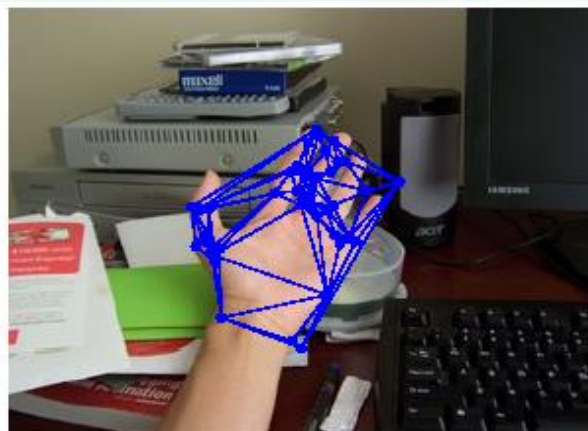
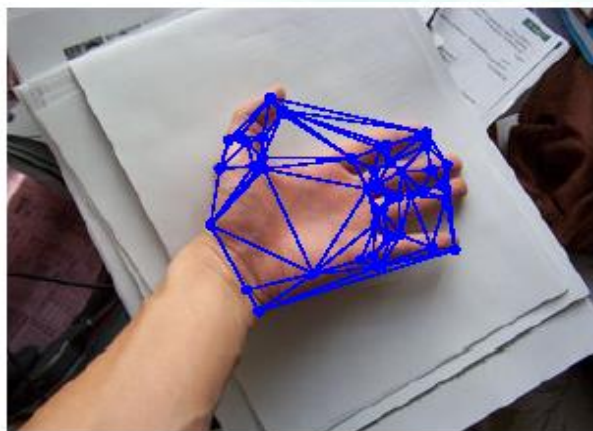
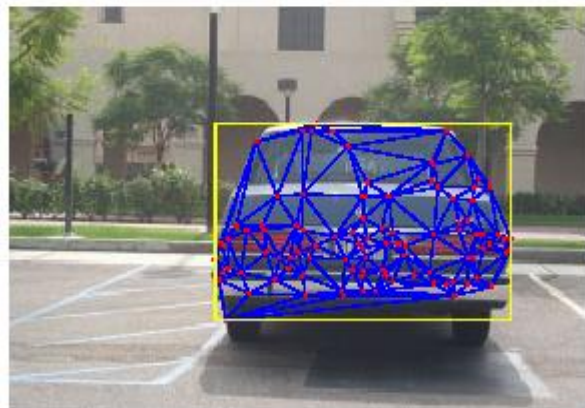
Lab



Double Squash



# Finding Objects in Clutter



# Posture Detection in Images



(a) Top 23 matches for figure skating posture 1. The first image is the exemplar.



(b) Top 23 matches for figure skating posture 2. The first image is the exemplar.

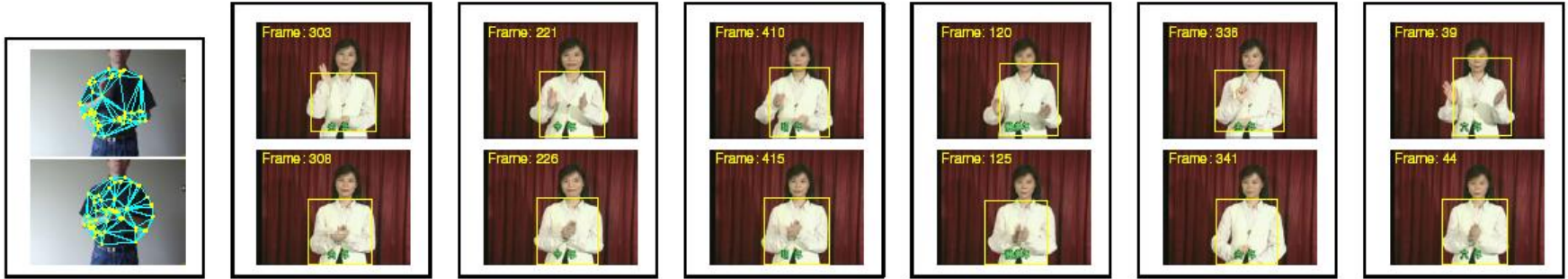


(c) Top 23 matches for figure skating posture 3. The first image is the exemplar.

Template  
Image

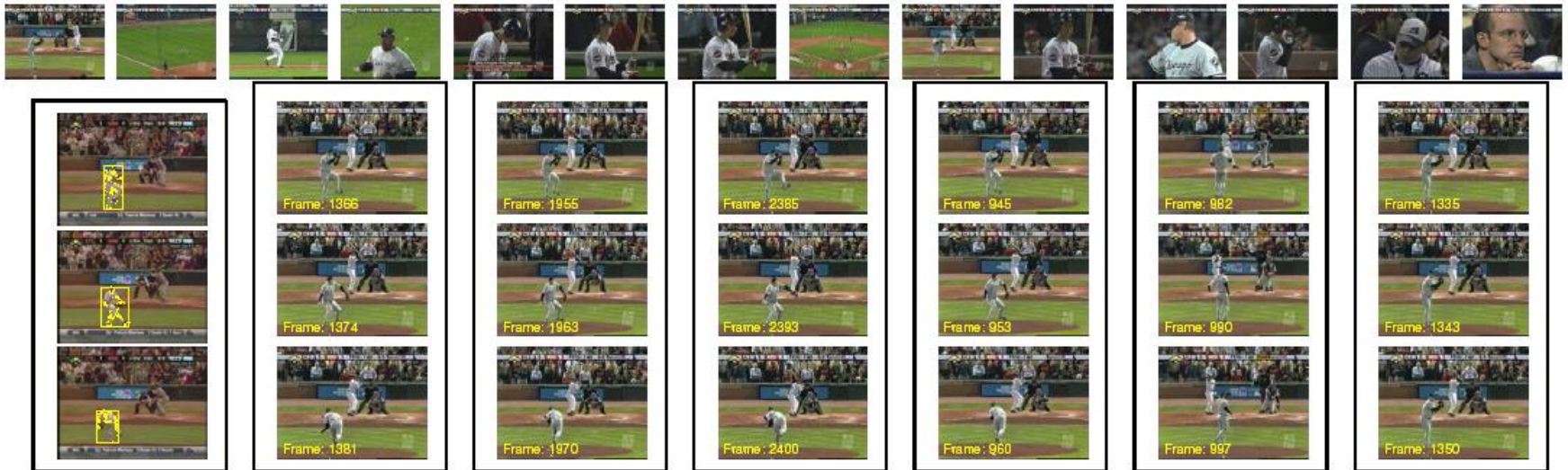


# Action Detection



(a) Templates (b) Cost:43.9 (c) Cost:45.5 (d) Cost:46.0 (e) Cost:46.2 (f) Cost:46.4 (g) Cost:46.5

## Sign language gesture detection



(a) Templates (b) Cost:29.29 (c) Cost:29.68 (d) Cost:29.91 (e) Cost:30.11 (f) Cost:30.17 (g) Cost:30.19

## Detecting actions in baseball sequence

## ❑ Journals

- IEEE Multimedia
- IEEE Transaction on Multimedia
- IEEE Transaction on Image Processing
- IEEE Transaction on Pattern Recognition and Machine Intelligence

## ❑ Conferences

- ACM Multimedia
- International Conference on Multimedia and Expo (**ICME**)
- IEEE Computer Vision and Pattern Recognition (**CVPR**)
- International Conference on Computer Vision (**ICCV**)

# Announcement for all

- ❑ 期末考试（60%）+ 平时成绩（40%）
- ❑ 平时成绩 = 平时作业 + 考勤
- ❑ 杜绝抄袭，按时提交
- ❑ 认定为抄袭者，此次作业0分
- ❑ 不按时提交者，此次作业0分
- ❑ 考勤不到者，按照学校规定处理
- ❑ 期末考试为大作业